

# A Grammar of Ayeri



# A Grammar of Ayeri

DOCUMENTING A FICTIONAL LANGUAGE

*by Carsten Becker*

---

*Benung. The Ayeri Language Resource*

© ⓘ ⓘ Carsten Becker, 2016.  
Published under CC-BY-SA 4.0 license.  
Last edited: July 27, 2016.

Ayeri is a fictional language spoken by fictional people in a fictional setting, and as such is not related to any naturally existing languages. It is thus not to be confused with *Azeri*, a Turkic language spoken in Azerbaijan and its surrounding countries. Ayeri's vocabulary is entirely a priori, this means, no real-world languages have been used specifically as sources of vocabulary. Due to the language's sound and spelling aesthetic being inspired by Austronesian languages, it is not surprising if overlaps with existing words in those languages happen accidentally.

🌐 <http://benung.nfshost.com>  
⚙️ <https://github.com/carbeck/ayerigrammar/>  
⚖️ <https://creativecommons.org/licenses/by-sa/4.0/>

# Contents

---

<b>List of Figures</b>	<b>iii</b>
<b>List of Tables</b>	<b>iv</b>
<b>Glossing Abbreviations</b>	<b>v</b>
<b>o Introduction</b>	<b>3</b>
<b>1 Phonology</b>	<b>9</b>
1.1 Phoneme inventory . . . . .	9
1.1.1 Consonants . . . . .	9
1.1.2 Vowels . . . . .	12
1.2 Phonotactics . . . . .	14
1.2.1 Number of syllables per word . . . . .	15
1.2.2 Phonemic makeup of initial syllables . . . . .	18
1.2.3 Phonemic makeup of medial syllables . . . . .	22
1.2.4 Phonemic makeup of final syllables . . . . .	25
1.2.5 Phonemic makeup of single syllables . . . . .	28
1.2.6 Cross-syllable consonant clusters . . . . .	32
1.3 Notes on prosody . . . . .	33
1.3.1 Stress . . . . .	33
1.3.2 Intonation . . . . .	42
<b>2 Writing system</b>	<b>47</b>
2.1 Consonants . . . . .	49
2.2 Vowels . . . . .	50
2.3 Diacritics . . . . .	52
2.3.1 Subscript diacritics . . . . .	54
2.3.2 Prepended diacritics . . . . .	54
2.3.3 Superscript diacritics . . . . .	58

2.4	Numerals . . . . .	59
2.5	Punctuation and abbreviations . . . . .	59
2.6	Styles . . . . .	61
<b>3</b>	<b>Morphological typology</b>	<b>65</b>
3.1	Typology . . . . .	65
3.2	Morphological processes . . . . .	68
3.2.1	Prefixation . . . . .	68
3.2.2	Suffixation . . . . .	71
3.2.3	Reduplication . . . . .	72
3.2.4	Suprasegmental modification . . . . .	72
3.2.5	Clitics . . . . .	72
3.3	Marking strategies . . . . .	73
	<b>Bibliography</b>	<b>75</b>
	<b>Index</b>	<b>78</b>

## List of Figures

---

o.1	First design for an Ayeri script . . . . .	6
o.2	First draft for Tahano Hikamu . . . . .	7
1.1	Consonant inventory . . . . .	10
1.2	Vowel inventory . . . . .	12
1.3	Syllabification of inflected verbs . . . . .	15
1.4	Declension paradigm for Ayeri ຈໍ <i>niva</i> ‘eye’ . . . . .	34
1.5	Types of heavy syllables . . . . .	35
1.6	Stress patterns for [+ HEAVY, – LONG] in trisyllabic words . . . . .	39
1.7	Stress patterns for [± HEAVY, + LONG] in trisyllabic words . . . . .	39
1.8	Examples of stress patterns in compounds . . . . .	41
2.1	Box script and Hikamu . . . . .	48
2.2	The consonant graphemes of Tahano Hikamu . . . . .	49
2.3	Additional consonant graphemes of Tahano Hikamu . . . . .	50
2.4	Primary vowel graphemes of Tahano Hikamu . . . . .	51
2.5	Secondary vowel graphemes of Tahano Hikamu . . . . .	51
2.6	Bottom-attaching diacritics of Tahano Hikamu . . . . .	53
2.7	Obligatorily prepended diacritics of Tahano Hikamu . . . . .	55
2.8	Allographically prepended diacritics of Tahano Hikamu . . . . .	55
2.9	Superscript diacritics of Tahano Hikamu . . . . .	58
2.10	The numerals of Tahano Hikamu . . . . .	59
2.11	Common punctuation marks of Tahano Hikamu . . . . .	60
2.12	Less common punctuation marks of Tahano Hikamu . . . . .	60
2.13	Tahano Hikamu, ‘book style’ . . . . .	61
2.14	Tahano Hikamu, ‘angular style’ . . . . .	62
2.15	Tahano Hikamu, ‘hand style’ . . . . .	62
2.16	Tahano Hikamu, ‘blackletter style’ . . . . .	63

# List of Tables

---

1.1	Frequency of words with different numbers of syllables . . . . .	15
1.2	Frequency of syllable types per word . . . . .	17
1.3	Frequency of onsets in initial syllables . . . . .	19
1.4	Frequency of nuclei in initial syllables . . . . .	20
1.5	Frequency of codas in initial syllables . . . . .	22
1.6	Frequency of onsets in medial syllables . . . . .	23
1.7	Frequency of nuclei in medial syllables . . . . .	24
1.8	Frequency of codas in medial syllables . . . . .	25
1.9	Frequency of onsets in final syllables . . . . .	26
1.10	Frequency of nuclei in final syllables . . . . .	27
1.11	Frequency of codas in final syllables . . . . .	28
1.12	Frequency of onsets in single syllables . . . . .	29
1.13	Frequency of nuclei in single syllables . . . . .	31
1.14	Frequency of codas in single syllables . . . . .	31
1.15	Frequency of cross-syllable consonant clusters . . . . .	32
3.1	Cliticood tests according to Klavans (1985) and Zwicky (1985) . . . . .	74



# Glossing Abbreviations

---

I	First person	M	Masculine
2	Second person	N	Neuter
3	Third person	NEG	Negative
A	Agent	NFUT	Near future
ACC	Accusative	NMLZ	Nominalizer
AGTZ	Agentizer	NN	Noun
AN	Animate	NOM	Nominative
AT	Agent topic	NPST	Near past
CAUS	Causative	P	Patient
DAT	Dative	PL	Plural
DYN	Dynamic	PROG	Progressive
F	Feminine	PST	Past
FUT	Future	PT	Patient topic
GEN	Genitive	PTCP	Participle
HAB	Habitative	REL	Relative
IMP	Imperative	RFUT	Remote future
INAN	Inanimate	RPST	Remote past
INDF	Indefinite	SBJ	Subject
INS	Instrumental	SG	Singular
IRR	Irrealis	SUPL	Superlative
LOC	Locative	TOP	Topic
LOCT	Locative topic	VB	Verb



# Preface

---

This is my latest attempt to write a grammar of Ayeri, a fictional language which I have been developing since December 2003. Getting to work on grammar writing again was triggered by a growing dissatisfaction with not having a central place of documentation, when the first thing people look for on my website is often the grammar, incomplete as well as partially inaccurate and outdated as it may be. In addition to that, there was a seminar on fictional languages at the University of Tübingen, Germany, in the summer semester of 2016 (Buch 2016). Ayeri was one of the languages that was chosen for students to explore and evaluate.

The student group who worked on Ayeri came to the conclusion that its documentation is severely lacking in the description of basic elements and assumptions, since whole chapters of the grammar had been missing to date (Boga et al. 2016: 12).<sup>1</sup> This is to say that previous attempts of writing a full-fledged grammar of Ayeri have been incomplete due to loss of enthusiasm and creeping neglect.

Although the *Ayeri Grammar* has so far been lying dormant for five years, I have written a whole number of blog articles detailing various grammatical issues (Becker 2016: Blog). These articles have been taken into consideration here. This grammar writing attempt is thus not only a transferral to a different typesetting system, but constitutes an extension to previous formal documentation as well.

I hope that by transferring my previous grammar writing from LibreOffice to L<sup>A</sup>T<sub>E</sub>X, combined with using GitHub as a version control system, maintaining and editing will become faster, more transparent, and more elegant, since L<sup>A</sup>T<sub>E</sub>X operates on plain text files, and version control helps in keeping track of changes over time.

Carsten Becker  
Marburg, July 18, 2016

<sup>1</sup>  *Kutānas-ikan* ‘thanks a lot’ to Bella Boga, Madita Breuninger, Thora Daneyko, and Martina Stama-Kirr for their hard work on making sense of my published materials in spite of information being scattered all over the place, as well as their providing me with the presentation concluding their group work.



## o Introduction

---

In December 2003, the idea for a new fictional language was born, an idea that turned out to stick with me for over 10 years now.<sup>1</sup> At that time, my seventeen years old self was still fairly new to this whole making-up languages business, read things about linguistics here and there, and was not shy to ask questions about terminology (and, looking at old mails, a little impertinently teenager-like so), for example on *Conlang-L* and the *Zompist Bulletin Board*. One thing seemed to catch my interest especially: syntactic alignments other than the NOM/ACC of the few languages I was familiar with, that is, German, English, and French. Apparently this curiosity was big enough for me to grow bored with my second fictional language, Daléian (declared ‘quite complete’ after maybe half a year of work or so), and to start something new from scratch in order to put newly acquired knowledge to test.

I had read about “trigger languages” on *Conlang-L* and wanted to try my hands on making my own. I cannot remember how long it took me to come up with a first draft of an Ayeri grammar, however, I do remember having been told that a good language cannot be made in a summer. Of course, I still did not really know what I was doing then, even though I thought I had understood things and authoritatively declared “this is how it works” in my first grammar draft when things sometimes really do not work that way. But at least an interest had been whetted.

In order to illustrate the various stages from the beginnings to current Ayeri, I went through some old backups contemporary with the very early days. Here is a sentence from the oldest existing document related to it, titled “Draft of & Ideas for my 3rd Conlang”—the file’s last-changed date is December 14, 2003, though I remember having started work on Ayeri in early December. I added glossing for convenience and according to what I could reconstruct from the notes. This uses vocabulary and grammatical markers just made up on the spot and for illustrative purposes; little of it actually managed to make it into actual work on Ayeri:

<sup>1</sup> Most of the text here is taken from the blog article, “Happy 10th anniversary, Ayeri!” (Becker 2013) with some slight rephrasings and extensions.

- (1) *Ayevhoi agiaemaesim coyaielieðamavir vhaieloyaŋaiye.*  
 Ay-evhoi agia-ema-esim coyai-el-i-eðam-avir vhai-el-o-yaŋa-iyē  
 3SG.AN-SBJ read-VB-SBJ.AN book-NN-AN-INDF-P bed-NN-INAN-ON-LOC  
 ‘He reads a book on the bed.’

According to the grammar draft of September 5, 2004, this would have already changed to:

- (2) *Ang layaiyain mecoyalei ling \*pinamea.*  
 Ang laya-iy-a-in me-coya-lei ling \*pinam-ea  
 A.SBJ read-3SG.AN<sub>1</sub>-a<sub>1</sub>-SBJ INDF.INAN-book-P.INAN top.of bed-LOC  
 ‘He reads a book on the bed.’

A word for ‘bed’—𐌸𐌶𐌵 *pinam*—was only (re-)introduced on October 24, 2008, however. In the current state of Ayeri, I would translate the sentence as follows:

- (3) *Ang layaya koyaley ling pinamya.*  
 Ang laya=ya.Ø koya-ley ling pinam-ya  
 AT read=3SG.M.TOP book-P.INAN top.of bed-LOC  
 ‘He reads a book on a/the bed.’

As you can see, quite a bit of morphology got lost already early on, especially the overt part-of-speech marking (!) and animacy marking on nouns. Also, prepositions were just incorporated into a noun complex as suffixes apparently. Gender was originally only divided into animate and inanimate, but I changed that at some point because only being really familiar with European languages, it felt awkward to me not to be able to explicitly distinguish ‘he’, ‘she’, and ‘it’.

A feature that also got lost is the assignment of thematic vowels in personal pronouns to third-person referents: originally, every third-person referent newly introduced into discourse would be assigned one of /a e i o u/ to disambiguate, and there was even a morpheme to mark that the speaker wanted to dissolve the association. Constituent order was theoretically variable at first, but I preferred SVO/AVP due to familiarity with that. Later on, however, I settled on VSO/VAP. Also, I had no idea about what was called “trigger morphology” on *Conlang-L* for the longest time—essentially, this referred to the Austronesian, or Philippine, alignment. I am not claiming that I know all about it now, just that due to reading up on the topic, I have a slightly more informed understanding now. Orthography changed as well over the years, so ⟨c⟩ in the early examples encodes the /k/ sound, not /tʃ/ as it does today; diphthongs were spelled as ⟨Vi⟩ instead of modern ⟨Vy⟩.

What was definitely beneficial for the development of Ayeri was the ever increasing amount of linguistics materials available online and my entering university (to study literature) in 2009, where I learned how to do research and also had a lot of interesting books available at the library.

One of the things people regularly compliment me on is Ayeri's script—note, however, that Tahano Hikamu was not the first one I came up with for Ayeri. Apparently, I had already been fascinated with the look of Javanese/Balinese writing early on;<sup>2</sup> Figure 0.1 shows a draft dated February 9, 2004. However, the letter shapes in this draft looked so confusingly alike that I could never memorize them. About a year later, I came up with the draft in Figure 0.2. What is titled “Another Experimental Script” there is what would later turn into Tahano Hikamu, Ayeri's ‘native’ script. According to the notes in my fictional language ring binder, the script looked much the same as today about a year from then, but things have only been mostly stable since about 2008.


An important date in the history of Ayeri was when I decided to set up an improved website for Ayeri that would include a blog. The idea was that this way, I could more freely write on whatever detail I currently interested me in Ayeri, outside of the constraints of the Grammar. Thus, *Benung. The Ayeri Language Resource* launched on March 1, 2011. Being able to write short articles, however, probably also led to neglecting work on the actual formal reference grammar, which had been lying dormant from January 2011 on. This was always on the premise that I would eventually include the information from blog articles in the grammar. However, juggling such a big document had always felt daunting, so I let laziness take the better part of me eventually as enthusiasm gradually subsided.<sup>3</sup> This renewed attempt at documentation has been started with the intention to right those wrongs.

I hope that by now it should be clear which kind of a fictional language Ayeri is: a personal, artistic language—or *artlang* in community parlance. Thus, my goal in creating Ayeri is not to propose yet another international auxiliary language, like Esperanto. It is also not my goal to make it as logical as possible, like Lojban. Neither is it my goal to engineer it towards certain underlying premises, for example, to reach a maximal amount of information density, like Ithkuil, or to get by on as few different words as possible, like Toki Pona. It is also not a ‘what-if’ language in the sense of “What could the modern language of Old Irish speakers transplanted to Australia look like?” or “Latin piped through Athabascan sound changes.”


<sup>2</sup> Compare, for instance, the charts in Kuipers and McDermott (1996). The Wikipedia articles on either script contain a number of images depicting the scripts in use, both current and historic.

<sup>3</sup> Let me add to my defense, however, that I also worked on my B.A. thesis in 2013 and my M.A. thesis in 2016, which required several months of preparation each and thus left me largely unable to work much on Ayeri.

Figure 0.1: First design for an Ayeri script (February 9, 2004)

ba	𐀀	ma	𐀁	va	𐀂	harden consonant:  e.g. "ca": 𐀃 e.g. "pa": 𐀄
da	𐀅	na	𐀆	sa	𐀇	
ga	𐀈	nga	𐀉	ha	𐀊	
ra	𐀋	la	𐀌	ya	𐀍	

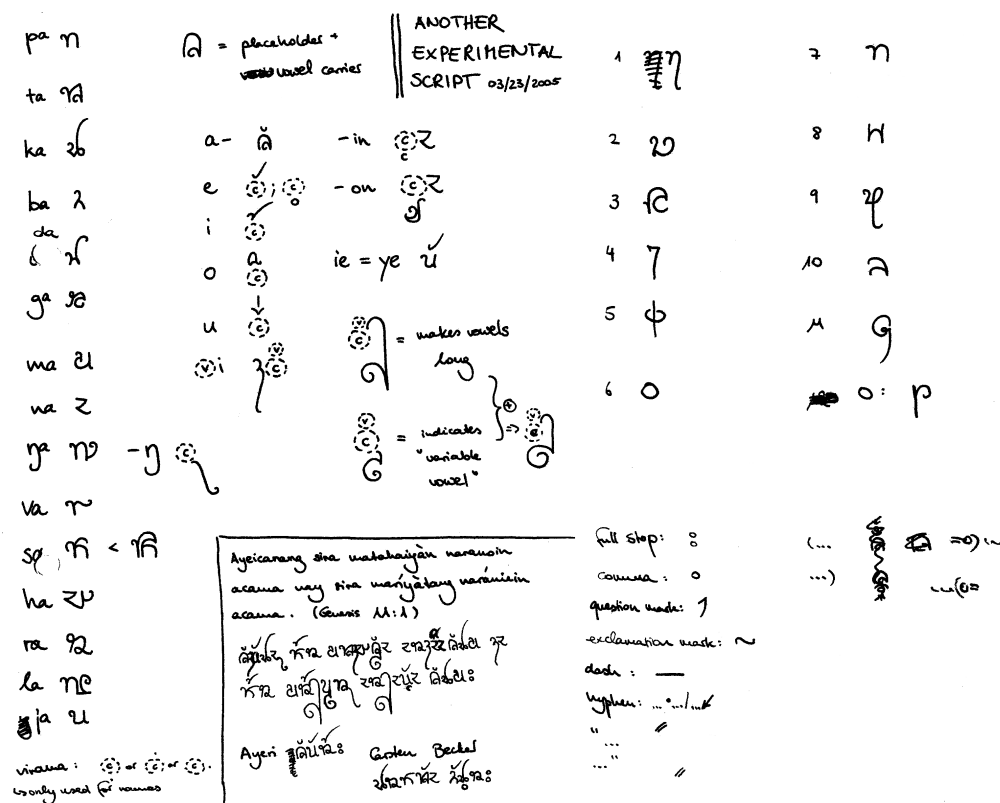
a-	𐀎	𐀏 𐀐 𐀑 𐀒 𐀓	𐀔 𐀕 𐀖 𐀗 𐀘	𐀙 𐀚 𐀛 𐀜 𐀝	var A 𐀞 var E 𐀟 var I 𐀠 var O 𐀡 var U 𐀢	mute -a of a consonant: 
e-	𐀎					
i-	𐀎					
-e	2					
-i	𐀑					
-o	𐀒					
-u	𐀓					

Ayeri is rather an attempt to create an artificial language for personal enjoyment and intellectual stimulation by creating a feedback loop between reading up on linguistics and actively devising rules for a fictional language accordingly, to see how things work within the frame I created, or to try and see whether certain ideas work together at all when combined, and to better understand why they do or do not. Ayeri will only ever be as perfect as miniature models of things can be, since it has not grown organically from millenia of human interaction, and I cannot and will never know about each and every aspect of language myself, in spite of continued curiosity about these matters. Nor will it be possible for me to replicate all the fascinating twists and irregularities that natural languages normally entail. The ultimate goal in my work on Ayeri is, I suppose, to make it emulate natural languages to at least some degree of depth and complexity.

In writing this grammar, I hope that I will find a good balance between applying



Figure 0.2: First draft for Tahano Hikamu (March 23, 2005)



linguistic theory to already existing materials and ideas, and going forth to create rules for aspects of the language that have so far been neglected, often due to my not being aware of them. In my opinion, the split between being able to apply methods of linguistics to what has grown over the course of more than a decade on the one hand, and discovering and developing new aspects of the language on the other is what makes Ayer an interesting piece of “informed nonsense,” as a colleague of mine once put it.

Errors are all mine. If my English is not always fully idiomatic, you find that I got my terminology wrong, or not all aspects of the language or its description are equally well worked-out—which are all very likely events—I ask you to bear with me. For one, English is not my native language, and second, I put up the grammar on GitHub in the hope of making it easier to fix and extend things through patches.<sup>4</sup> This also means that you can file an Issue there in case of issues if you feel inclined to do so. Criticism is always welcome as long as it is constructive.

<sup>4</sup> See <https://github.com/carbeck/ayerigrammar/>.



# 1 Phonology

---

This chapter will present charts depicting the phoneme inventory of Ayeri and describe the various commonly encountered allophones of both consonants and vowels. Following this, a detailed statistical analysis of the words found in a number of translated texts from 2008 to 2016 as well as dictionary entries up to July 2016 will produce insights into Ayeri’s phonotactics. Some notes on stress patterns and intonation will close the chapter.

## 1.1 Phoneme inventory

### 1.1.1 Consonants

At 17 consonants, Ayeri has a “moderately small” inventory, according to Maddieson (2013a). Figure 1.1 shows the full chart of consonant phonemes.

Regarding allophony, /tj kj/ and /dj gj/ are usually realized as [tʃ] and [dʒ], respectively, except if a homorganic nasal /n/ or /ŋ/ is preceding: for instance, ႁႏႱ *ankyū* /ʼaŋkju/ ‘really’ is realized as [ʼaŋkju], not as \*[ʼaŋtʃu] or \*[ʼantʃu]. It is important to note, however, that besides this synchronic palatalization process leading to [tʃ] and [dʒ] as *allophones*, there is also a diachronic one in parallel here—or the diachronic process is still ongoing. For example, there is no way to predict whether ႁႏႱ *cuna* ‘original, initial’, ႏႏႱ *panca* ‘finally, eventually’, and ႏႏႱ *vac-* ‘like’, or ႏႏႱ *jaraṇ* ‘pilgrimage’, ႏႏႱ *aja-* ‘play’, and ႏႏႱ *nui-* ‘pour’ have /tj/ or /kj/, /dj/ or /gj/, respectively, unless we consider the clues given by the conservative native spellings of the respective words.<sup>1</sup> We can rather assume two sound changes, (1) tj, kj → tʃ, and (2) dj, gj → dʒ, leading to the *phonemes* /tʃ/ and /dʒ/ in the present-day language.

<sup>1</sup> Actual scribes would typically err in cases where the merger is complete, so this strategy would, in fact, be of limited use in the real world.

Figure 1.1: Consonant inventory (divergent orthography in pointed brackets)

	Bilabials	Labiodentals	Alveolars	Palatals	Velars	Glottals
Plosives	p      b		t      d		k      g	
Affricates			tʃ ⟨c⟩      dʒ ⟨j⟩			
Nasals		m		n	ŋ ⟨ng⟩	
Fricatives		v	s			h
Taps/Flaps				r		
Approximants			l	j ⟨y⟩		

The plural marker ၵ -ye is commonly contracted to [dʒ] when a case suffix beginning with a vowel follows:<sup>2</sup>

- (1) a. ၵၵၵၵၵၵ nyānyeang → nyānjang [ˈnja:ndʒaŋ] ‘persons’ (person-PL-A);  
 b. ၵၵၵၵ netuyas → netujas [neˈtudʒas] ‘brothers’ (brother-PL-P).

The plural marker may also contract before the locative marker ၵ -ya and the dative marker ၵ -yam, basically for dissimilation:<sup>3</sup>

- (2) a. ၵၵၵ nivyeyya → nivajya [niˈvadʒja] ‘at the eyes’ (eye-PL-LOC);  
 b. ၵၵၵ maviyeyam → mavijyam [maˈvidʒjam] ‘to the sheep’ (sheep-PL-DAT).

Dissimilation of the sequence ၵ -yaya is attested in my translation of Kafka’s short story “Eine kaiserliche Botschaft,” where the relative pronoun ၵၵၵ *siyaya* appears transcribed as *sijya*:

As far as morphophonology is concerned, the relative pronoun complex *sijya* ‘in/at/on which.LOC’ is interesting in so far as it is a contraction of \**siyaya* ‘REL-LOC-LOC’ that I introduced here [...] Since this feature does not occur in previous texts, let’s assume it’s an acceptable variant. (Becker 2012: 12)

The contraction happens “only if both parts are grammatical suffixes” (12), however, so the environments this contraction may appear in are effectively limited to relative pronouns combining locative and locative, or locative and dative marking.

The word ၵၵၵ *lajāy* ‘student’ is special in that it is the only word with ၵ [dʒa] so far. Presumably it is derived from the verb ၵၵၵ *laya-* ‘read’ with the agentive suffix ၵ -maya, except the shortening of the suffix—with or without compensatory lengthening of the final vowel of the modified word stem—was applied irregularly, possibly via \**layāya*. The regular form ၵၵၵ *layamaya* means ‘reader’.

Lastly, /h/ may assimilate to its phonemic environment and is realized as [ç] before front vowels, and as [x] before back vowels in this case:

- (3) a. ၵ tabi [ˈtaçi] ‘favorable’;  
 b. ၵ babo [ˈbaxo] ‘loud’.

While vowels become long when two identical vowels come into succession, consonants do not geminate but are treated like a single consonant:

<sup>2</sup> The customary romanization uses ⟨c⟩ and ⟨j⟩ for allophonic cases of [tʃ] and [dʒ] as well.

<sup>3</sup> ၵ -ea also occurs as an allomorph, so that ၵ -ye + ၵ -ea → ၵ -yēa.

- (4) a.  $\text{တပ်တပ်}$  *tavvāng* [ta'va:ŋ] 'you get' (get=2SG.A),  
 b.  $\text{ညှပ်တပ်}$  *disyyang* [di'sjaŋ] 'I fasten' (fasten=1SG.A).

With diphthongs, the sequence /V<sub>1</sub>.j/ is treated as though it were /Vj.j/, so the double /j/ simplifies to just a single /j/; however, the vowel remains lax in spite of being phonetically in an open position now:

- (5)  $\text{ဆုံပျံ$  *tipuyya* [ti'pu.ja] 'on the grass' (grass-LOC).

### 1.1.2 Vowels

Figure 1.2: Vowel inventory (divergent orthography in pointed brackets>)

	Front	Center	Back
High	i, i: ⟨ī⟩		u, u: ⟨ū⟩
Mid	e, e: ⟨ē⟩	ə ⟨ə, e⟩	o, o: ⟨ō⟩
Back		a, a: ⟨ā⟩	

Ayeri's vowel system distinguishes five qualities, as shown in Figure 1.2; Maddieson (2013c) classifies this as "average." Length, however, is also a factor, and there are five diphthongs as well, as we will see below. At  $\frac{17}{5}$ , the consonant–vowel ratio is 4.25, which Maddieson (2013b) again classifies as "average," although Ayeri finds itself at the upper end of the tier.

The lax vowels [ɪ ɛ ɔ ʊ] occur as allophones of their tense counterparts /i e o u/ in closed syllables, for example:

- (6) a.  $\text{မိန့်}$  *ming* [mɪŋ] 'can, be able',  
 b.  $\text{မိန့်}$  *enya* [ɛn.ja] 'everyone',  
 c.  $\text{မိန့်}$  *agon* [a.gɔŋ] 'outer, foreign', and  
 d.  $\text{မိန့်}$  *pakur* [pa.kʊr] 'ill, sick'.

/ə/ occurs marginally in the tense prefixes  $\text{န}$ : *kə*- 'NPST',  $\text{မ}$ : *mə*- 'PST',  $\text{ပ}$ : *və*- 'RPST', as well as in the prefix  $\text{မိန့်}$ : *mə*- 'some, whichever'. Otherwise, [ə] acts as an allophone of /e/ in final unstressed position, for instance, in the word  $\text{မိန့်}$  *mine* [mi.nə] 'affair, matter, issue'.

Ayeri also possesses a number of diphthongs, these are: /ai ei ɔi ʊi au/, spelled ⟨ay⟩, ⟨ey⟩, ⟨oy⟩, ⟨uy⟩, and ⟨au⟩. Furthermore, there are long equivalents of the short vowels: /i: e: a: o: u:/; in romanization, long vowels are marked with a macron ⟨̄⟩ over the letter. Long vowels are lexicalized in a few words, for example:

- (7) a.  $nīsa$  ‘wanted’,  $pasīsa$  ‘interesting’;  
 b.  $arēn$  ‘anyway, however’,  $lēra$  ‘whore’;  
 c.  $lā$  ‘tongue’,  $yāng$  ‘he’ (he.A);  
 d.  $nōn$  ‘will, intention’; and  
 e.  $babūan$  ‘barbarian’.<sup>4</sup>

Otherwise, long vowels result from two same vowels next to each other, for instance:

- (8)  $aja-$  ‘play’ +  $-an$  ‘NMLZ’  $\rightarrow$   $ajān$  ‘game, play’.

Morphophonologically, long vowels also occur in double-marked relative pronouns where the agreement marker for the relative clause’s head has been omitted, for instance,  $sinā$  ‘of which, about which’, as in the following example:

- (9)  $Le$   $turayāng$   $taman$   $sinā$   $ang$   $ningay$   
 Le tura-yāng [taman-Ø]<sub>1</sub> si-Ø<sub>1</sub>-na ang ning=ay.Ø  
 PT.INAN send=3SG.M.A letter-TOP REL-PT.INAN-GEN AT tell=ISG.TOP  
 $tamala$   $vās$ .  
 tamala vās  
 yesterday 2SG.P

‘The letter which I told you about yesterday, he sent it.’

This is to disambiguate it from the plain genitive-marked relative pronoun  $sinā$  ‘which.GEN’:<sup>5</sup>

- (10)  $tamanreng$   $ledanena$   $nā$   $sina$   $koronvāng$   
 taman-reng [ledan-ena  $nā$ ]<sub>1</sub> si-na<sub>1</sub> koron-vāng  
 letter-A.INAN friend-GEN ISG.GEN REL-GEN know=2SG.A  
 ‘the letter of my friend which you know’

As pointed out in (7c), the word  $lā$  ‘tongue’ ends in a long vowel, so the question is what happens when a case suffix beginning with a vowel is appended. To avoid a hiatus, a glide /j/ may be inserted, so both of the following renditions are possible:

<sup>4</sup> I have gone years without /u:/, but it has always seemed slightly odd to me to lack a vowel in that position when all other vowels can be long. Therefore,  $babūan$  ‘barbarian’ and its adjective  $babū$  ‘barbarian (adj.)’ were coined as  $prankayē$ —things ‘that you put in specifically to make things fit’, another new coining this decision resulted in.

<sup>5</sup> A variant which combines the allomorphs of the relativizer and the genitive case marker in the opposite way also exists:  $s-$  +  $-ena$   $\rightarrow$   $sena$ .

- (11) a. *Aku lāas!*  
           Aka-u lā-as  
           swallow-IMP tongue-P  
           ‘Shut up!’  
       b. *Aku lāyas!*  
           (idem)

With diphthongs—as described above—, /ɪ/ coalesces with a following /j/ to /j/, but the initial vowel will not become tense, thus:

- (12) ၼ်ပုၼ် *tipuyya* [ti'pu.ja] ‘on the grass’ (grass-LOC).

Moreover, /u/ is commonly realized as [w] when followed by a vowel, for example in ပုၼ် *huākaya* ['wa:kaja] ‘frog’ or ရှိ *rua-* [rwa] ‘have to, must’. [w] may also be an allophone of /uj/, as in ၼ် *adauyi* [a'dawi] ‘then’, ၼ် *edauyi* [e'dawi] ‘now’, or ၼ် *nekuyi* ['nekwi] ‘eyebrows’. The negative suffix ၼ် *-oy* is also commonly contracted to [w] before a diphthong:

- (13) ၼ် *mingoyay* → *minguay* [mɪŋ'wai] ‘I cannot’ (can-NEG=ISG.TOP).

## 1.2 Phonotactics

For the purpose of this statistical analysis, most of the available translations into Ayeri from late 2008 to July 2016 have been used as a text corpus;<sup>6</sup> example sentences from various blog articles have also been added, as well as dictionary entries for all nouns, adjectives, adverbs, pronouns, adpositions, conjunctions, and numerals if they were not prefixes or suffixes.<sup>7</sup> Borrowings have been deleted if they could not

<sup>6</sup> These texts are: A Medieval Neighborhood Dispute (2015), A Message from the Emperor (2012), Article 1 of the Universal Declaration of Human Rights (2011), The Beginning of Tolstoy’s *Anna Karenina* (2014), Conlang Christmas Card Exchange 2008/09 (2009), Conlang Holiday Card Exchange 2010/11 (2011), Conlang Relay 15 (2008), Conlang Relay 17 (2010), Conlang Relay 18 (2011), The First Two Chapters from Saint-Exupéry’s *Le Petit Prince* (2013), The Four Candles (2010), Honey Everlasting (2014), LCC4 Relay (2011), The Lord’s Prayer (2015), The North Wind and the Sun (2016), The Origin of the Wind (2009), Ozymandias (2011), Please Call Stella ... (2008), Psalm 23 (2013), The Scientific Method (2014), The Sheep and the Horses (2012), Sugar Fairies (2011), The Upside-Down Ice Skater (2009). The texts can be accessed from Becker (2016: Examples).

<sup>7</sup> This section updates and extends a previous analysis of the phonological makeup of dictionary entries (Becker 2010). The previous study had its focus on gathering frequency statistics for word generation, however, we want to know about words generally here.



reasonably be words in Ayeri. Altogether, the corpus comprises 5,500 words, which is a very small figure for such a study, but there are only so many texts available unfortunately. Words may occur more than once.

Among the dictionary entries, verbs have notably been ignored, since verb stems alone do not constitute independent words—they are always inflected in some way, so that they may end in consonants or consonant clusters that independent words cannot end in. This also has repercussions on syllabification and stress, which depend on the inflection of the verb stem:

Figure 1.3: Syllabification of inflected verbs

Suffix	<i>ca-</i> ‘love’	<i>gum-</i> ‘work’	<i>babr-</i> ‘mumble’
- <i>ay</i> (1SG)	cáy	gu.máy	ba.bráy
- <i>va</i> (2SG)	cá.va	gúm.va	ba.brá.va
- <i>yam</i> (PTCP)	cá.yam	gúm.yam	bá.bryam

For the purpose of gathering statistics on phonemes, the words from translated texts were converted to IPA first. Fortunately, this is rather easy as Ayeri’s romanization is very straightforward. Syllable breaks have also been inserted semi-automatically.

### 1.2.1 Number of syllables per word

First, let us see how many syllables words commonly have (see Table 1.1). The higher the syllable count, the more likely it is for them to be compounds or inflected words.

Table 1.1: Frequency of words with different numbers of syllables (n = 5500)

Segment	Count	Percentage
2 syllables	2277	41.40 %
3 syllables	1393	25.33 %
1 syllable	1201	21.84 %
4 syllables	547	9.95 %
5 syllables	74	1.35 %
6 syllables	8	0.15 %

Two-syllable words make up the bulk of the sample, which is not surprising since 1,072 entries (55.43 %) in the dictionary subsample are disyllabic: most of Ayeri’s roots are disyllabic. Unsurprisingly, most monosyllabic words are function words like the ones cited below. In the following, I will quote a few examples for each number of syllables per word:

- (14) a. ၼံ ၼ *ang* (AT),  
           န *nay* ‘and’,  
           ရ *rua* ‘must’;  
       b. ပါ *datau* ‘normal’,  
           န *nasay* ‘near to’;  
       c. ၼံ *avanyāng* ‘he sinks’ (sink=3SG.M.A),  
           တ *tovale* ‘a cloak’ (cloak-P.INAN);  
       d. ၼ *binyanveno* (corner.beautiful, a place name),  
           မ *mitanena* ‘of the palace’ (palace-GEN);  
       e. ၼ *haruyamanas* ‘a beating’ (beat-PTCP-NMLZ-P),  
           န *sungkorankihas* ‘geography’ (science.map);  
       f. န *kaytomayanena* ‘of righteousness’ (righteous-NMLZ-GEN),  
           န *nasimayajang-ben* ‘all followers’ (follow-AGTZ-PL-A=all).

Table 1.2 shows the frequencies of syllable types by position in a word. It is important to note here that phonemes which consist of more than one segment—affricates, diphthongs, and long vowels—have been counted as only one of C (consonant) or V (vowel), respectively. The following subsections will elaborate on which sounds the Cs and Vs correspond to. Moreover, it is important to note that medial syllables have not been further distinguished by position in the word for the sake of this analysis, so anything between the second and the fifth medial syllable is treated the same. It would furthermore be possible to calculate the frequencies of one syllable type following the other, however, no such calculations have been carried out here.

In all positions, CV is the most common syllable type, followed by CVC. With a very big margin, V is the next most common syllable type, which is also most common in initial syllables and least common in monosyllabic words. The cases with only a few attestations are the following:

- (15) a. Initial CVCC:  
           န *linktang* /lɪŋk.'taŋ/ ‘they try’ (try=3PL.M.A),<sup>8</sup>  
           န *silvnang* /silv.'naŋ/ ‘we see’ (see=IPL.A);  
       b. Final CCCV:  
           မ *migryo* /'mi.grjo/ ‘flourishes’ (flourish-3SG.N),  
           န *subryo* /'su.brjo/ ‘ceases’ (cease-3SG.N);  
       c. Single V:  
           န *ay* /a/ ‘I’ (1SG.TOP).

Table 1.2: Frequency of syllable types per word (n = 5500)

Type	Initial	Medial	Final	Single	Total
CV	2896 67.36%	1974 72.02%	2109 49.06%	578 48.13%	7557 60.26%
CCV	55 1.28%	24 0.88%	46 1.07%	32 2.66%	157 1.25%
CCCV	—	—	2 0.05%	—	2 0.02%
CVC	761 17.70%	610 22.25%	1902 44.24%	298 24.81%	3571 28.48%
CCVC	29 0.67%	10 0.36%	85 1.98%	9 0.75%	133 1.06%
CVCC	2 0.05%	—	—	—	2 0.02%
V	488 11.35%	95 3.47%	67 1.56%	2 0.17%	652 5.20%
VC	68 1.58%	28 1.02%	88 2.05%	282 23.48%	466 3.72%
Total	4299 100.00%	2741 100.00%	4299 100.00%	1201 100.00%	12540 100.00%

The medial and final VC cases may seem like an oddity, but they are mostly due to the previous syllable ending in /ŋ/, with that syllable also containing a lax vowel, which means that this syllable must be closed. An alternative explanation would be to assume that /ŋ/ is ambisyllabic, or actually /n.g ~ ŋ.g/, but realized as [ŋ]. The high number of single-syllable VC is due to ၵၵ ang ‘AT’, which alone appears 255 times in the sample (4.63% of all words, 21.23% of monosyllabic words, 90.43% of monosyllabic VC words).

### 1.2.2 Phonemic makeup of initial syllables

The statistics in the following sections have been gathered from the IPA conversions of translated texts and dictionary entries mentioned above. The transcribed words have been split into syllables and then the collected contents of each position group were written into separate plain text files, one each for:

- all initial syllables of polysyllabic words,
- all medial syllables of polysyllabic words,
- all final syllables of polysyllabic words, and
- all monosyllabic words.

Monosyllabic words are both initial and final syllables at the same time; they have been counted separately for the purpose of this analysis. Onsets, nuclei and codas have been matched by regular expressions; the command line tools `grep`, `sort`, and `uniq` were used to aggregate all occurring variants for each syllable segment as well as their absolute frequencies:<sup>9</sup>

(16) C = (? : tʃ | dʒ | [ptkbgdmnŋvshrʌjw])  
 V = (? : [ae] : ?ɪ | əʊ | [ieaou] : ? | [ɪɛɔʊə])

As we have seen above (Table 1.2), CCV syllables only make up 1.28% of initial syllables, insofar it is no surprise that consonant clusters all appear at the bottom of Table 1.3. There also seem to be combination patterns in that initial clusters exist for all plosives plus /r/, and almost all bilabials plus /j/, with the exception of /bj/, however, /nj/ is added to the group instead. Combinations with /w/ only occur for /b/, /r/, and /s/, which do not share an obvious connection. Syllables without a consonant filling the onset position are marked with ‘Ø’; these numbers correspond to the VC and VCC rows in Table 1.2.

<sup>8</sup> The verb stem is found in the dictionary as ၵၵ linka-, with a final -a, and thus is possibly an entry changed at a later point, or the example from the text (Sugar Fairies) chosen here contains an error.

<sup>9</sup> However, `sort` was unable to handle all IPA characters, so `sed 'y/ɪɛɔʊə:ʃʒŋ/EI0U@:SZN/'` had to be used to compensate by transcribing everything into X-SAMPA.

Table 1.3: Frequency of onsets in initial syllables (n = 4299)

Phoneme	Frequency	Percentage
Ø	556	12.93 %
s	488	11.35 %
t	432	10.05 %
m	418	9.72 %
k	380	8.84 %
n	375	8.72 %
p	334	7.77 %
b	231	5.37 %
d	172	4.00 %
v	164	3.81 %
l	159	3.70 %
r	134	3.12 %
j	126	2.93 %
g	111	2.58 %
h	99	2.30 %
tʃ	30	0.70 %
pr	27	0.63 %
nj	27	0.63 %
kr	8	0.19 %
br	8	0.19 %
tr	6	0.14 %
dʒ	4	0.09 %
gr	3	0.07 %
w	2	0.05 %
sw	1	0.02 %
rw	1	0.02 %
pj	1	0.02 %
mj	1	0.02 %
bw	1	0.02 %

Table 1.4: Frequency of nuclei in initial syllables (n = 4299)

Phoneme	Frequency	Percentage
a	1847	42.96 %
i	1011	23.52 %
<i>i</i>	802	18.66 %
<i>ɪ</i>	209	4.86 %
e	705	16.40 %
<i>e</i>	523	12.17 %
<i>ɛ</i>	164	3.81 %
<i>ə</i>	18	0.42 %
u	260	6.05 %
<i>u</i>	228	5.30 %
<i>ʊ</i>	32	0.74 %
o	227	5.28 %
<i>o</i>	188	4.37 %
<i>ɔ</i>	39	0.91 %
a:	109	2.54 %
aɪ	88	2.05 %
eɪ	40	0.93 %
e:	4	0.09 %
ɔɪ	3	0.07 %
ʊɪ	1	0.02 %
o:	1	0.02 %
i:	1	0.02 %
e:ɪ	1	0.02 %
aʊ	1	0.02 %

Perhaps most striking about the nuclei of initial syllables presented in Table 1.4 is that plain vowels occur most frequently. As mentioned above, lax vowels are counted here as allophones of tense ones as their distribution is complementary and are listed here for the sake of completeness. This is the reason why the plain vowels are presented as grouped with their allophones in this table as well as in subsequent ones. Long vowels and diphthongs find themselves below the 5% threshold, and the words with single occurrences are:

- (17) a.  $\text{kuysān}$  ‘comparison’,  
 b.  $\text{nōn}$  ‘will, intention’,  
 c.  $\text{nīsa}$  ‘wanted’,<sup>10</sup>  
 d.  $\text{sēyraya}$  ‘will overcome’ (FUT-overcome-3SG.M),  
 e.  $\text{sautan}$  ‘cork’.

As the diphthong [e:i] only occurs due to allophony, it should not be counted as a phoneme for the purposes of this analysis. On the other hand, the same could be said for a lot of cases of [a:] included here—this caveat applies to all nouns derived from verbs ending in *-a* with the very common nominalizing suffix  $\text{-an}$ , as exemplified in (8) above. Similarly, the 18 instances of /ə/ reported here are mostly from tense prefixes also mentioned above, for instance,  $\text{mākoronay}$  ‘I knew’ (PST-know=1SG.TOP).

Initial-syllable codas (Table 1.5) are far less diverse than consonant onsets: there are only 10 attested segments in comparison to 28 for onsets (not counting empty codas of C(C)V syllables, which constitute the majority by a large margin), and the only two cluster attested are /ŋk/ in the word  $\text{linktang}$  ‘they try’ (try=3PL.M.A), and /lv/ in the word  $\text{silvnang}$  ‘I see’ (see=1PL.A). There only being two incidences of a CC cluster is very probably an effect of the small sample size. Furthermore, the only unvoiced single coda consonants attested are /s/, /h/, /t/, /tʃ/ and /k/, the latter two only once, /h/ twice:

- (18) a.  $\text{mehvāng}$  ‘you are supposed to’ (be.supposed.to=2SG.A),<sup>11</sup>  
 $\text{robtang}$  ‘they bite’ (bite=3SG.M.A);  
 b.  $\text{mutva}$  ‘you rub’ (rub=2SG.TOP),  
 $\text{patlay}$  ‘cousin’;  
 c.  $\text{sik-sik}$  ‘tits’;  
 d.  $\text{vacvāng}$  ‘you like’ (like=2SG.A).

<sup>10</sup>  $\text{nīsa}$  and  $\text{nōn}$  are both related to  $\text{no-}$  ‘want, plan’.

Table 1.5: Frequency of codas in initial syllables (n = 4299)

Phoneme	Frequency	Percentage
Ø	3441	80.04 %
n	298	6.93 %
ŋ	243	5.65 %
r	129	3.00 %
l	88	2.05 %
m	74	1.72 %
s	20	0.47 %
t	2	0.05 %
h	2	0.05 %
tʃ	1	0.02 %
ŋk	1	0.02 %
lv	1	0.02 %
k	1	0.02 %

### 1.2.3 Phonemic makeup of medial syllables

The onsets of medial syllables (Table 1.6) show properties very similar to those of initial syllables. The order of most common consonants may differ here—for example, the most common onset is /r/, not Ø or /s/—, but there are no restrictions on consonants to appear in this position, with the exception of /ŋ/ for reasons stated above (see section 1.2.1). Regarding initial clusters, there are further attestations for plosive plus /r/ (except for /kr/). As for clusters with /j/, the only one with a bilabial is /bj/, but the set is extended to /sj/ and /kj/. For clusters with /w/, only /sw/ and /kw/ occur here, while attestations for /bw/ and /rw/ as in initial-syllable onsets are lacking. This does not mean that those combinations are not principally possible in this position, however.

As with onset consonants, vowel nuclei of medial syllables (Table 1.7) do not show significant differences compared to those of initial syllables either. /a/ is more common here, and /o/ and /u/ switch places. Instead of /e:ɪ/, there is an attestation of /u:/ (see footnote 4), for which there is more reason to be counted as a phoneme than for /e:ɪ/. The sequences /i:/ and /uɪ/ also only occur once and twice, respectively, namely in the following words:

<sup>11</sup> The dictionary entry for the verb is *ḡ mya-*, so this may be an instance of my changing a word in the dictionary with the old one staying in the text (The Four Candles).



Table 1.6: Frequency of onsets in medial syllables (n = 2741)

Phoneme	Frequency	Percentage
Ø	123	4.49%
r	343	12.51%
n	260	9.49%
j	233	8.50%
t	222	8.10%
d	213	7.77%
k	189	6.90%
s	170	6.20%
m	169	6.17%
l	149	5.44%
v	148	5.40%
h	147	5.36%
p	119	4.34%
g	92	3.36%
b	89	3.25%
tʃ	20	0.73%
dʒ	15	0.55%
tr	11	0.40%
dr	8	0.29%
pr	7	0.26%
w	6	0.22%
sj	2	0.07%
br	2	0.07%
sw	1	0.04%
kw	1	0.04%
kj	1	0.04%
bj	1	0.04%

Table 1.7: Frequency of nuclei in medial syllables (n = 2741)

Phoneme	Frequency	Percentage
a	1480	53.99%
i	480	17.51%
<i>i</i>	387	14.12%
<i>ɪ</i>	93	3.39%
e	254	9.26%
<i>e</i>	206	7.52%
<i>ɛ</i>	48	1.75%
o	194	7.08%
<i>o</i>	119	4.34%
<i>ɔ</i>	75	2.74%
u	120	4.38%
<i>u</i>	101	3.68%
<i>ʊ</i>	19	0.69%
a:	110	4.01%
aɪ	51	1.86%
ɔɪ	33	1.20%
eɪ	5	0.18%
e:	5	0.18%
aʊ	5	0.18%
ʊɪ	2	0.07%
u:	1	0.04%
i:	1	0.04%

- (19) a. *pasisa* ‘interesting’;  
       b. *puluyley* ‘a mirror’ (mirror-P.INAN),  
           *tipuyya* ‘on the grass’ (grass-LOC).

The word in (19a), *pasisa* ‘interesting’, is rather transparently constitutes a causative derivation of the verb *pasy-* ‘wonder, be curious, be interested’, essentially meaning ‘making one wonder/curious’—the causative suffix *-isa* can as well be used to derive adjectives with a causative or resultative meaning. Nonetheless it should count as a lexeme in its own right, since it possesses idiomatic meaning.

With medial-syllable codas (Table 1.8) again, sonorants and /s/ make up the largest number of consonants in this position; /t/ and /g/ only occur once each in

Phoneme	Frequency	Percentage
Ø	2093	76.36 %
n	313	11.42 %
ŋ	193	7.04 %
r	48	1.75 %
m	39	1.42 %
s	32	1.17 %
l	21	0.77 %
t	1	0.04 %
g	1	0.04 %

<sup>12</sup> The word for ‘money’ is ကံနု *pangis*, so (20a) is probably a compound, albeit not a fully transparent one. The word for ‘change’ is တီး *tila-*, and there seems to be a nominalizing -*an*. Ayerī allows noun–verb compounds to have a nominalized verb in the second position in spite of it being the head—noun–noun compounds mostly come in head-initial order—probably due to an avoidance of placing a derivative suffix in the middle of a word. Possibly, what happened after all is that တီးလံ *tilān* underwent metathesis to \*တီးလံ *\*itlān* to match the rhyme of ကံနု *pangis*. \*ကံနုတီးလံ *\*pangisitlān* then underwent irregular haplology (and shortening of the nominalizing suffix) to ကံနုတီး *pangitlan*.

Table 1.9: Frequency of onsets in final syllables (n = 4299)

Phoneme	Frequency	Percentage	Phoneme	Frequency	Percentage
Ø	155	3.61%	pr	7	0.16%
j	1101	25.61%	kj	6	0.14%
n	528	12.28%	hj	5	0.12%
r	398	9.26%	bj	5	0.12%
t	268	6.23%	tw	4	0.09%
s	244	5.68%	sw	4	0.09%
l	238	5.54%	sj	4	0.09%
k	199	4.63%	kw	3	0.07%
d	184	4.28%	kr	3	0.07%
m	154	3.58%	br	3	0.07%
v	144	3.35%	vr	2	0.05%
h	128	2.98%	rw	2	0.05%
p	115	2.68%	nw	2	0.05%
g	103	2.40%	tʃ	1	0.02%
dʒ	73	1.70%	rj	1	0.02%
b	73	1.70%	nj	1	0.02%
tʃ	52	1.21%	mw	1	0.02%
vj	26	0.60%	grj	1	0.02%
pj	22	0.51%	dv	1	0.02%
dʒj	17	0.40%	dr	1	0.02%
tr	10	0.23%	brj	1	0.02%
w	9	0.21%			

are attached. Even though single-segment onsets are strongly preferred, Cr, Cw, and especially C(C)j seem to be generally permissible.<sup>13</sup>

Nuclei of final syllables (Table 1.10) do not bear striking differences to nuclei in other positions. /a:/ comes out second here due to the common nominalizer ʔan, which lengthens the vowel of verb stems ending in /a/, as demonstrated in (8). /a/ is also fairly common here as it is the topic-marked first-person pronoun/pronominal clitic; for the same reason, /a:ɪ/ occurs a number of times—the vowel-lengthening

<sup>13</sup> The sequence /sj/ poses difficulty here as there are examples for /Vs.jV/ as well as for /V.sjV/, and I cannot tell for sure if there is a strict rule in operation. It seems that /V.sjV/ is more likely to occur when the second syllable is stressed, whereas /Vs.jV/ is more likely to occur when the first syllable is stressed. Ayeri's own Tahano Hikamu orthography would not show the difference either, since /sja/ is spelled ʒ either way, and there is no heeding morpheme breaks in placing the diacritic. /Cs.jV/ will be /C.sjV/ in any case, since Ayeri avoids final consonant clusters if possible, see Table 1.2.

Table 1.10: Frequency of nuclei in final syllables (n = 4299)

Phoneme	Frequency	Percentage
a	2408	56.01%
a:	316	7.35%
o	411	9.56%
o	298	6.93%
ɔ	113	2.63%
i	289	6.42%
ɪ	147	3.42%
ɨ	142	3.30%
aɪ	254	5.91%
u	207	4.82%
u	155	3.61%
ʊ	52	1.21%
e	209	4.85%
ɛ	127	2.95%
ə	81	1.88%
ɐ	1	0.02%
eɪ	103	2.40%
ɔɪ	42	0.98%
a:ɪ	23	0.54%
ʊɪ	14	0.33%
aʊ	14	0.33%
e:	5	0.12%
i:	3	0.07%
u:	1	0.02%

rule applies here as well, so its status as a phoneme is marginal. All instances of /e:/ in the sample are from the word ၵရ် *arēn* ‘anyway, however’; all evidence for /i:/ is from ၵရ် *sirī* ‘due to which’ (see section 1.1.2). The only evidence for /u:/ in the sample is from ၵရ် *babū* ‘barbarian (adj.)’.

Table 1.11: Frequency of codas in final syllables (n = 4299)

Phoneme	Frequency	Percentage
Ø	2224	51.73 %
n	899	20.91 %
ŋ	651	15.14 %
s	244	5.68 %
m	225	5.23 %
l	34	0.79 %
r	21	0.49 %
k	1	0.02 %

The list of coda consonants in final syllables (Table 1.11) is very slightly more restrictive than even that of coda consonants in medial syllables (see Table 1.8), since the only non-sonorant attested is /k/, which only occurs in ၵရ် *sik-sik* ‘tits’ again, which—besides being a vulgar term, thus maybe slightly more dispositioned to allow for deviating phonotactics—looks quite like onomatopoeia, possibly for the sound of sucking.<sup>14</sup>

### 1.2.5 Phonemic makeup of single syllables

Onsets of single syllables (Table 1.12) appear to be the least varied category. Still, none of the basic set of consonant morphemes (see Figure 1.1) is missing—the frequency order is just completely different from the other onsets surveyed, not merely a mix of initial and final syllables. Consonant clusters with /j/, /w/ and /r/ exist here as well. Combinations with /j/ are only present for /m/ and /n/, while /r/ again combines with plosives; /w/ combines with /n/ and /r/ at least, which we have already seen in final-syllable onsets (see Table 1.9). Whereas /mj/ has only occurred once in initial-syllable onsets so far (see Table 1.3), it occurs a few more times here, all in the word ၵရ် *mya* ‘be supposed to’, which is very commonly used as an unconjugatable modal particle.

<sup>14</sup> Kroonen (2013: 489–490) identifies PGmc *\*sūgan-*, *\*sūkan-* ‘to suck’ as an iterative of PGmc *\*sukkōn-*, *\*sugōn-* ‘to suck’ and reconstructs PIE *\*souk-neh<sub>2</sub>-*. However, he does not say anything about the Germanic word being onomatopoeic in origin.

Table 1.12: Frequency of onsets in single syllables (n = 1201)

Phoneme	Frequency	Percentage
Ø	284	23.65%
n	231	19.23%
s	147	12.24%
j	144	11.99%
k	51	4.25%
v	48	4.00%
m	46	3.83%
l	44	3.66%
t	41	3.41%
d	33	2.75%
r	26	2.16%
h	23	1.92%
mj	16	1.33%
p	13	1.08%
tʃ	9	0.75%
g	9	0.75%
nj	8	0.67%
rw	7	0.58%
b	7	0.58%
pr	5	0.42%
dʒ	3	0.25%
tr	2	0.17%
nw	1	0.08%
ŋ	1	0.08%
kr	1	0.08%
br	1	0.08%

A consonant onset that can only be found in monosyllables is /ŋ/,<sup>15</sup> in ၵာ်န -*ngas* ‘almost’, a quantifier suffix that has managed to sneak in due to being marked as an adverb in the dictionary, since it can modify a verb:

- (21) *Apayeng-ngas.*  
 Apa-yeng-ngas  
 laugh=3SG.F.A=almost  
 ‘She almost laughed.’

Here, ၵာ်န -*ngas* modifies the verb complex like any other adverb:

- (22) *Apayeng baho.*  
 Apa-yeng baho  
 laugh=3SG.F.A loudly  
 ‘She laughs loudly.’

However, whereas ခဵု *baho* ‘loud’ is treated as a separate unit in terms of intonation, ၵာ်န -*ngas* is unstressed and binds to what it follows:

- (23) a. ၵာ်နၵာ်န ၵာ်န။ *Apayeng-ngas*. [apa'jɛŋas];  
 b. ၵာ်နၵာ်နခဵု။ *Apayeng baho*. [apa'jɛŋ 'baxo].

As with onset consonants of monosyllabic words, nuclei of this syllable type are the least diverse group again (Table 1.13). One segment that is notably absent is /au/, and the marginally phonemic /e:/ is not present either. By having /a/, /aɪ/, /a:/ at the top, monosyllabic words behave similar to final syllables of polysyllabic words (see Table 1.10), however, the order of the most common vowels bears more similarities to that of initial and medial syllables (see Tables 1.4 and 1.7). The very uncommon /o:/ features twice in this group, namely in two instances of the word ခဵု *nōn* ‘will, intention’.<sup>16</sup>

Like the other syllable segments of monosyllabic words, coda consonants (Table 1.14) as well show the lowest degree of variety among all the coda consonants of the various syllable classes discussed so far. The order is basically the same as that of final-syllable codas (see Table 1.11), though /ŋ/ supersedes /n/ and there is some attestation of final /h/. As noted above, the prevalence of /ŋ/ is due to the

<sup>15</sup> At least according to the analysis chosen here, see section 1.2.1 for an explanation.

<sup>16</sup> Ayeri used to have ခဵု -*on* as a nominalizer beside ခဵု -*an*, however, it was not very productive and has long fallen out of use. ခဵု *nōn* is thus, in fact, originally a nominalization of ခဵု *no-* ‘want, plan’.



Table 1.13: Frequency of nuclei in single syllables (n = 1201)

Phoneme	Frequency	Percentage
a	568	47.29%
aɪ	171	14.24%
a:	140	11.66%
i	113	9.41%
<i>i</i>	65	5.41%
<i>ɪ</i>	48	4.00%
e	104	8.66%
<i>ɛ</i>	65	5.41%
<i>ɐ</i>	34	2.83%
<i>ə</i>	5	0.42%
o	45	3.75%
<i>ɔ</i>	30	2.50%
<i>o</i>	15	1.25%
u	20	1.67%
a:ɪ	14	1.17%
ɔɪ	10	0.83%
i:	6	0.50%
eɪ	5	0.42%
ʊɪ	3	0.25%
o:	2	0.17%

Table 1.14: Frequency of codas in single syllables (n = 1201)

Phoneme	Frequency	Percentage
Ø	612	50.96%
ŋ	377	31.39%
n	105	8.74%
s	58	4.83%
m	36	3.00%
l	6	0.50%
h	4	0.33%
r	3	0.25%

agent-topic marker ႁႃ *ang* (see section 1.2.1). /h/ only occurs in the interjections ႁႃ *ah!* and ႁႃ *āh!*, so its status as an actual phoneme in this position is marginal at best.

### 1.2.6 Cross-syllable consonant clusters

Table 1.15: Frequency of cross-syllable consonant clusters (n = 1270)

Interval [%]	Consonant cluster
0.00 ... 0.09	g.t, h.t, h.v, k.s, l.n, lv.n, m.bj, m.d, m.dʒ, m.l, m.n, m.pr, m.r, n.dv, n.g, n.h, n.w, ၵ.dʒ, ၵ.kw, ၵ.m, ၵ.n, ၵ.rj, ၵ.t, ၵk.t, r.b, r.dʒ, r.g, r.l, r.m, r.sj, r.tʃ, r.v, s.dʒ, s.h, s.l, s.n, s.p, s.v, t.v, tʃ.v (0.08%).
0.10 ... 0.24	l.bj, m.br, m.t, n.s, ၵ.b, ၵ.h, ၵ.p, ၵ.w, r.dʒ, r.pj, s.dʒ, s.m, t.l (0.16%); l.dʒ, l.p, m.k, n.sj, ၵ.dʒ, ၵ.g, ၵ.s, r.pr (0.24%).
0.25 ... 0.49	m.v, r.s, s.r (0.31%); n.r, s.t (0.39%); m.pj, n.dʒ, r.d (0.47%).
0.50 ... 0.74	ၵ.kj, ၵ.v, r.k, r.n (0.55%); l.b, l.t, ၵ.r (0.71%).
0.75 ... 1.00	r.p, r.t (0.87%); l.vj (0.94%).
1.0 ... 2.4	m.j (1.18%); ၵ.l (1.34%); n.tʃ (1.50%); n.dʒ (2.13%); n.v (2.28%); l.j (2.36%).
2.5 ... 4.9	m.p (2.52%); s.j (2.60%); n.l (2.91%); l.v (3.15%); m.b (3.23%); ၵ.k (3.78%).
5 ... 9	n.t (5.28%); n.d (6.85%); ၵ.j (7.32%); r.j (8.98%).
10+	n.j (25.35%).

Since a table detailing every combination with its absolute and relative frequency would be too large here, Table 1.15 gives the attested combinations ordered by brackets. As can be expected, bilabials cluster mostly with bilabials (83/112 purely bilabial CC combinations = 74.11%), alveolars with alveolars (317/948 = 33.44%), and velars with velars (59/207 = 28.51%). However, at least for alveolars and velars, the score is even higher with /j/: 52.64% and 44.93%, respectively. /j/ is also the most common second consonant overall, at 47.8% of all consonant clusters; /n.j/ is the most common cluster at a total of 25.35%. Alveolars provide the highest variety of both first and second consonants, with 6 different phonemes making up 74.65% of C<sub>1</sub>, and 8 different phonemes making up 28.74% of C<sub>2</sub>.

Labiodentals and glottals occur least frequently, on the other hand: There is only one cluster with /v/ as a first consonant, namely, /lv.n/ (0.08%). For /h/, there are two, which are /h.v/ and /h.t/ (0.16%). Altogether, however, there are 97 combinations in /v/ (7.64%)—most commonly /l.v/ (3.15%) and /n.v/ (2.28%)—while there are only 4 in /h/ (0.31%): /n.h/, /s.h/, and twice /ၵ.h/.

At 924 first consonants (72.76%), the nasals /m/, /n/, and /ŋ/ make up the largest group going by manner of articulation, followed by the tap /r/, which appears 175 times (13.78%) as the first consonant. For second consonants, approximants constitute the largest group at 669 combinations (52.68%), followed by 387 pairs with plosives second (30.47%).

## 1.3 Notes on prosody

### 1.3.1 Stress

Ayeri uses dynamic accent, that is, stress is based on differences in the loudness of syllables, among others.<sup>17</sup> Which syllable is stressed depends on a mix of which position in a word a syllable occupies and the phonemic shape of it. In fact, English, which also has phonemic stress in pairs such as *record* /'rekərd/ (noun) and /rɪ'kɔrd/ (verb) that Ayeri lacks, does a similar thing (examples adapted from Halle 1998: 552):

- (24) *admire* /æd'maɪər/ — *admirable* /'ædmərəbl/ [English]  
*carnivore* /'karnɪvɔr/ — *carnivorous* /kar'nɪvərəs/  
*ignore* /ɪg'nɔr/ — *ignorant* /'ɪgnərənt/

Stress does not stay at fixed intervals in these words and they even change their sound structure a little, but there are a number of variables which can nonetheless be formally described and applied here (564–565).<sup>18</sup> To demonstrate how word stress moves around in Ayeri, the complete declension paradigm for *ṣ̣ niva* ‘eye’ is presented in Figure 1.4.

It may appear that in the table above, stress is always on the penultimate syllable, which is indeed the case for most forms quoted there, but compare the superficially unmarked form *ṣ̣ niva*, which is disyllabic with stress on the first (= penultimate) syllable, to the agent and patient singular forms, *ṣ̣ nīvāṅ* and *ṣ̣ nīvās*, respectively. These are also disyllabic, however, they are stressed on the second (= ultimate) syllable. Similarly, compare the agent and patient plural forms to each other: the agent plural form *ṣ̣ nivajang* is trisyllabic and has its main

<sup>17</sup> For a discussion of terms, see Kager (2007), for instance.

<sup>18</sup> Halle (1998) takes a generativist approach rather than a more modern Optimality-Theory based one like Kager (2007) does, who only deals with fixed-stress systems in this introductory article, though Halle’s article is still informative. Simplifying a lot, English essentially tries to construct trochaic feet from the right edge of the word. If the last syllable’s vowel is not light, it is skipped and stress moves to the antepenultimate syllable; this process is recursive for words with multiple feet, although some suffixes introduce irregularities in rule application.

<sup>19</sup> Final-syllable stress is possible as well, also in the plural.

Figure 1.4: Declension paradigm for Ayeri ຂໍ້ *niva* ‘eye’

	Singular		Plural	
TOP	<i>ní.va</i>	‘the eye’	<i>ni.vá.ye</i>	‘the eyes’
A	<i>ni.vǎng</i>	‘eye’	<i>ni.va.jǎng</i>	‘eyes’
P	<i>ni.vǎs</i>	‘eye’	<i>ni.vá.jas</i>	‘eyes’
DAT	<i>ni.vá.yam</i> <sup>19</sup>	‘to the eye’	<i>ni.vá.jyam</i>	‘to the eyes’
GEN	<i>ni.vá.na</i>	‘of the eye’	<i>ni.va.yé.na</i>	‘of the eyes’
LOC	<i>ni.vá.ya</i>	‘at the eye’	<i>ni.vá.jya</i>	‘at the eyes’
CAUS	<i>ni.va.í.sa</i>	‘due to the eye’	<i>ni.va.jí.sa</i>	‘due to the eyes’
INS	<i>ni.vá.ri</i>	‘with the eye’	<i>ni.va.yé.ri</i>	‘with the eyes’

stress on the third (= ultimate) syllable, while the equally trisyllabic patient plural form ຂໍ້ນີວາຈິນ *nivajas* is stressed on the second (= penultimate) syllable again.

It should have become clear that even though the basic form ຂໍ້ *niva* has first-syllable stress, *ni* will not necessarily carry stress across the whole paradigm. It should also have become clear that the basic algorithm to determine stressed syllables in Ayeri has something to do with counting syllables from the right edge of a word, although some complications need to be factored in. The following sections will try to describe these formally.

#### Analysis of stress patterns in disyllabic words

The basic foot in Ayeri is a trochee, and for the most part it does not matter whether the syllable is open or closed, or whether there are complex onsets or codas, or no onsets or codas at all:<sup>20</sup>

- (25) a.     $\acute{x}$          $\times$     ||  
           *ba - ri*     ‘pithy, striking’
- b.     $\acute{x}$          $\times$     ||  
           *sa - yan*    ‘hole, cave’  
           *sem - ba*    ‘comb’

<sup>20</sup> In the following, a syllable will be marked by  $\langle x \rangle$  and receives an acute accent  $\langle \acute{} \rangle$  when carrying primary stress, a grave accent  $\langle \grave{} \rangle$  when carrying secondary stress, and no accent when unstressed. Feet are marked by horizontal lines  $\langle | \rangle$ ; the end of a word is marked by two horizontal lines  $\langle || \rangle$ .

- c.     $\acute{\times}$         $\times$        ||  
       *bri* - *ha*       ‘grace’  
       *ba* - *brya*     ‘(he) mumbles’  
       *a* - *gu*       ‘chicken’

It can be deduced from words with more than two syllables that stress assignment is trochaic. Stress assignment furthermore runs from right to left, so that in a word with more than two syllables, the last two syllables form a full foot:

- (26) a.     $\times$     |     $\acute{\times}$         $\times$        ||  
           *ba* - *ha* - *lan*     ‘target, goal’  
           *jar* - *ma* - *ya*     ‘pilgrim’
- b.     $\acute{\times}$         $\times$     |     $\acute{\times}$         $\times$        ||  
       *ho* - *ra* - *ma* - *ya*     ‘sinner’  
       *ya* - *ma* - *na* - *ti*     ‘causer’

In the case of (26b), the stressed syllables of the first foot bear secondary stress while those of the second foot bear primary stress. Complications, then, come in the form of syllables ending in /ŋ/, containing a long vowel, or containing a diphthong, or a combination of those features. Ayeri does not have syllables that contain a diphthong and also end in /ŋ/, though, since consonant codas after a diphthong are largely avoided.<sup>21</sup> Since the presence or the absence of a certain element that is suspected to have an effect on stress assignment is a yes–no decision, we can make a matrix of binary features:

Figure 1.5: Types of heavy syllables

	[+ DIPH, – ŋ]	[– DIPH, + ŋ]	[– DIPH, – ŋ]
[+ LONG]	++	++	++
[– LONG]	+	+	–

The feature matrix above (Figure 1.5) shows the various kinds of syllable types that we will now see have a manipulative effect on trochaic stress assignment. These syllable types can be considered ‘heavy’ in that they attract stress and thus modify the regular assignment of stress to every other syllable from the right edge of a word. For the time being, we will only test their effects on disyllabic words as the most common type. As will be shown in the following example, heavy syllables in ultimate positions attract stress while quasi-regular results are produced when they are in penultimate position and the ultimate syllable is not heavy:

<sup>21</sup> It might thus be possible to alternatively analyze diphthongs in /ɪ/ as /Vj/ sequences, essentially.

- (27) a.  $\times$        $\acute{\times}$       ||  
           *ma* - *tay*      ‘summer, wet season’  
           *pa* - *dang*      ‘mind; heart, mood’  
           *ka* - *nāy*      ‘I marry’ (marry=ISG.TOP)  
           *bras* - *yāng*      ‘he bathes’ (bathe=3SG.M.A)  
           *na* - *rān*      ‘word; speech’
- b.  $\acute{\times}$        $\times$       ||  
           *kār* - *yo*      ‘strong’  
           *key* - *nam*      ‘humans, people’  
           *kan* - *ka*      ‘mind; heart, mood’

Unfortunately, there are no disyllabic examples for the feature sets [+LONG, -DIPH, +η] and [+LONG, +DIPH, -η] in the first syllable (syllables of the type /C(C(j))V:η/ or /C(C(j))Vη/). If there were, they would group with (27b).

So far, we have only looked at heavy syllables combined with regular/light ones. In the following case, however, another property of heavy syllables will become apparent: long syllables outweigh those containing a diphthong or ending in /η/. They are essentially superheavy, which is why some of the fields in Figure 1.5 are marked with two plus signs. The following examples show what happens when heavy syllables are combined with other heavy syllables. Let us start by examining the various combinations possible between [-LONG, +DIPH, -η] and the elements from the [+LONG] row (28a), and the possible combinations between [-LONG, -DIPH, +η] and the [+LONG] row (28b).

- (28) a.  $\times$        $\acute{\times}$       ||  
           *bay* - *hāy*      ‘I govern’ (govern=ISG.TOP)  
           *say* - *lyāng*      ‘he sails’ (sail=3SG.M.A)  
           *kay* - *vān*      ‘container’
- b.  $\times$        $\acute{\times}$       ||  
           *kong* - *āyn*      ‘we enter’ (enter=IPL.TOP)  
           *keng* - *vāng*      ‘you notice’ (notice=2SG.A)  
           *lang* - *-vā*      ‘in the most tiresome way’ (tiresome=SUPL)

We can see here that these words have primary stress invariably on the last/long syllable in spite of a heavy syllable preceding in the examples in (28b). The question then is, however, what happens if we invert this order. This is more problematic than it sounds, however, as initial [+LONG, +DIPH, -η] and [+LONG, -DIPH, +η], as well as final [-LONG, +DIPH, +η] do not occur, insofar there will only be one possible combination here—the reverse pattern of  $\eta\acute{\times}\times$  or *lang-vā* ‘in the most tiresome way’ from (28b) above, also compare with (27):

- (29)     $\acute{x}$          $\times$         ||  
          *cā* - *nang*     ‘love’ (love-A)

There is only one one pattern possible here, which is very little to make a point, however, other words following this syllable pattern, like  $\text{ᠨᠠᠷᠢᠭ}$  *nāreng* ‘rather’, for example, behave in the same way. A long syllable has precedence over other kinds of heavy syllables, so  $\text{ᠨᠠᠭ}$  -*nang* does not take away stress from  $\text{ᠴᠠ}$  *cā*- as one might expect from the examples in (27a). Another question is what happens if we pit elements from the [ $\pm$  LONG] rows against another feature combination of the same row. As above, we will start with the [- LONG] row:

- (30) a.     $\times$          $\acute{x}$         ||  
               *bay* - *tang*     ‘blood’  
       b.     $\times$          $\acute{x}$         ||  
               *pang* - *lay*     ‘goddess’

In the case of examples for [+ LONG] pattern combinations, we need to keep in mind again that initial [+ LONG, + DIPH, - ɳ] and [+ LONG, - DIPH, + ɳ] are not attested, so again, there will only be one possible combination of two syllables with a long vowel:

- (31)     $\acute{x}$          $\acute{x}$         ||  
          *mā* - *sāy*     ‘I traveled’ (PST-travel=ISG.TOP)

Combining two long syllables with each other will result in both being stressed, which is otherwise avoided in Ayeri, as we will see later. Moreover, the following patterns emerge if we combine each pattern with itself; the combinatorial restrictions mentioned above apply again, of course:

- (32) a.     $\times$          $\acute{x}$         ||  
               *kay* - *vay*     ‘without’  
               *dang* - *reng*   ‘bell’ (bell-A.INAN)  
       b.     $\acute{x}$          $\acute{x}$         ||  
               *bā* - *mā*     ‘parents, mom-and-dad’

As demonstrated in (28), the last heavy syllable will receive primary stress, except if two long syllables collide, in which case the first long syllable will receive secondary stress.

To summarize the above findings:

1. Ayeri assigns trochaic stress from the right edge of a word. A foot thus consists of two syllables, of which the first is stressed.
2. Syllables ending in /ŋ/ or ones containing a diphthong are considered heavy. They attract stress and take it away from a preceding stressed syllable if the following syllable is not stressed already.
3. Syllables containing a long vowel are considered superheavy and override both light and heavy syllables in attracting stress, since long vowels cannot be unstressed.
4. Primary stress is assigned to the last stressable syllable, or otherwise the last heavy syllable. In the rare case of two long/superheavy syllables after another, the first syllable receives secondary stress and reduces in duration.
5. Secondary stress is assigned to syllables that are eligible for word stress but which are not in the final foot.

#### *Analysis of stress patterns in trisyllabic words*

So far, we have only considered all the possible combinations of two heavy and light syllables. Doing the same for all combinations of three and more syllables would be possible, though the list of examples were to become even longer. Since the feature pair  $[\pm \text{DIPH}, \pm \eta]$  behaves the same way throughout and both features are in complementary distribution, we need not test iterations of them separately, but can subsume them under the label  $[\pm \text{HEAVY}]$ . The parameters that need testing, then, are  $[\pm \text{HEAVY}]$  in combination with  $[\pm \text{LONG}]$ . There are 4 possible outcomes for these two features, which in the case of three syllables leads us to  $(2 \times 2)^3 = 64$  theoretically possible combinations. For this reason, I want to point out just a few cases, since the general rules sketched out above still apply.

First, let us look at  $[+ \text{HEAVY}, - \text{LONG}]$  combined with  $[- \text{HEAVY}, - \text{LONG}]$  in all positions (Figure 1.6). Finding words that fit the respective permutations is not too much of a problem, especially in cases where there is only one heavy syllable.

It becomes clear from Figure 1.6 that the rules stated at the end of the previous section (p. 38) also hold in the case of trisyllabic words whose syllables alternate short syllables based on the  $[\pm \text{HEAVY}]$  feature: ကုဉ်ဉ်း prantanley, ရာဉ်း ralanghay, နီဉ်း kaybunay, and ပာဉ်း panglay-kay receive final-syllable stress since this is their last heavy syllable. The first syllables of နီဉ်း taykondam and မာဉ်း maykongas, on the other hand, lose the secondary stress they would normally be assigned as two stressed syllables after another are normally avoided; the requirement of long syllables to not be unstressed does not come into effect here. နီဉ်း taykondam is



Figure 1.6: Stress patterns for [+ HEAVY, – LONG] in trisyllabic words

–H –H +H	<i>prantanley</i>	x   x ˘	‘question’ (question-P.INAN)
–H +H –H	<i>sarayya</i>	x   ˘ x	‘(he) bows’ (bow-3SG.M)
+H –H –H	<i>taykondam</i>	x   ˘ x	‘break (n.)’
–H +H +H	<i>ralangbay</i>	x   x ˘	‘thumbnail’
+H –H +H	<i>kaybunay</i>	˘   x ˘	‘by the way’
+H +H –H	<i>maykongas</i>	x   ˘ x	‘harbor’ (harbor-P)
+H +H +H	<i>panglay-kay</i>	˘   x ˘	‘a few goddesses’ (goddess=few)

also an example for the rule that even if a syllable is not heavy, the last syllable that can be assigned stress will receive primary stress.

Carrying out the same analysis as above and moving the feature [+ LONG] through the various positions, we receive the results depicted in Figure 1.7.<sup>22</sup> Since long syllables override stress of both light and heavy syllables as pointed out above (p. 38), the example words in this chart contain both of these syllable types. It was not too hard finding examples for all the slots in this case either, except that words with two long syllables in succession are rather rare. Still, only the case of three long syllables must remain unattested.<sup>23</sup>

Figure 1.7: Stress patterns for [± HEAVY, + LONG] in trisyllabic words

–L –L +L	<i>peraysān</i>	x   x ˝	‘paste’
–L +L –L	<i>raypānya</i>	x   ˝ x	‘at the stop’ (stop-LOC)
+L –L –L	<i>nōneri</i>	˝   x x	‘deliberate, intentional’
–L +L +L	<i>mə-cān-cān</i>	x   ˝ ˝	‘whatever fling’ (whatever=fling)
+L –L +L	<i>sānisān</i>	˝   x ˝	‘copula; clutch (n.)’
+L +L –L	<i>lēāyon</i>	˝   ˝ x	‘manwhore’
+L +L +L	—	—	—

Again, we can see that long syllables attract stress, in that the final syllables of ပါရယ်ဆံ *peraysān* and ဆံနီဆံ *sānisān* are stressed even though the penultimate syllable

<sup>22</sup> For more precision, modifications will be made to the symbols given in footnote 20: let a double acute (˝) denote superheavy syllables with primary stress, and a double grave (˘) denote superheavy syllables with secondary stress.

<sup>23</sup> It would be possible to construct a word with three long syllables if the habitative suffix -asa did not delete the vowel at the end of the verb stem if there is one. မာဆီ *māsāy* ‘I traveled’ (PST-travel=1SG.TOP) would then become \*မာဆီဆီ *\*māsāsāy* ‘I used to travel’ (PST-travel-HAB=1SG.TOP) instead of the actual form မာဆီဆီ *māsāsāy*; the verb stem is ဆီ *asa-* ‘travel’.

is heavy on the virtue of containing a diphthong. As it is in an unstressed position and there is no requirement for the syllable to be stressed, the first syllable of  $\text{ၵၵၵၵၵၵ}$  *raypānya* loses stress adjacent to the stressed long penultimate syllable.  $\text{ၵၵၵၵၵၵ}$  *nōneri* and  $\text{ၵၵၵၵၵၵ}$  *lērāyon* display a secondary-stressed and a primary-stressed syllable next to each other, in the first case due to the rule that long syllables must not be unstressed and  $\text{ၵၵၵၵၵၵ}$  *-eri* forming a valid disyllabic foot that receives regular trochaic stress, and in the second case due to two long syllables next to each other, of which the first—again—must not be unstressed.  $\text{ၵၵၵၵၵၵ}$  *mā-cān-cān* is interesting insofar as the long-syllable stress rules operate on the second foot regularly, while the marker  $\text{ၵၵၵၵၵၵ}$  *mā-* ‘whatever’ is by default unstressed, so even if it is added to stressed monosyllabic words like  $\text{ၵၵၵၵၵၵ}$  *bin* ‘box’, the foot of  $\text{ၵၵၵၵၵၵ}$  *mā-bin* ‘which box so ever’ would appear iambic, even though the syllable  $\text{ၵၵၵၵၵၵ}$  *bin* itself is not heavy.

A further exception is formed by monosyllabic quantifying clitics like  $\text{ၵၵၵၵၵၵ}$  *-ben* ‘all, every’ or  $\text{ၵၵၵၵၵၵ}$  *-ngas* ‘almost’. These are unstressed also when following an unstressed syllable of their host if they contain a short vowel.

#### Stress in compounds

Ayeri has a number of lexicalized compound nouns that are treated as one word morphologically (33a). This is in contrast to compounds that are not as established terms, or formed ad hoc (33b):

- (33) a. *Ang bengay kardangirayya ya Litareng.*  
           *Ang beng-ay.Ø kardang.iray-ya ya Litareng*  
           AT attend=ISG.TOP school.high-LOC LOC= Litareng  
           ‘I attend university in Litareng.’
- b. *Ang pasyye Pila sungkoranyam kibas.*  
           *Ang pasy-ye Ø=Pila sungkoran-yam kihal*  
           AT be.interested.in-3SG.F TOP=Pila science-DAT map  
           ‘Pila is interested in geography.’

For purposes of surveying stress patterns, we will only deal with the kind in (33a), though it may be noted that when not being overtly inflected, the second kind of compound will be treated as a word as well:  $\text{ၵၵၵၵၵၵ}$  *sungkorankihal* ‘geography’. Another kind of indivisible compound as in (33a) is one formed from reduplication, for instance,  $\text{ၵၵၵၵၵၵ}$  *kusang-kusang* ‘model’, from  $\text{ၵၵၵၵၵၵ}$  *kusang* ‘double’. Figure 1.8 gives several examples along with their stress patterns. As a reference for the various rules in operation, consider the list above (p. 38).

The first word,  $\text{ၵၵၵၵၵၵ}$  *apan-apan*, is not very noteworthy but I included it nonetheless as a reference for regular stress assignment to light syllables. The word

Figure 1.8: Examples of stress patterns in compounds

Word	Pattern	Translation	Word	Pattern	Translation	Word	Pattern	Translation
<i>apan-apan</i>	à x   à x	‘extensive’	<i>apan</i>	à x	‘wide’	<i>apan</i>	à x	‘wide’
<i>depangcāti</i>	à x   à x	‘cuckold’	<i>depang</i>	x à	‘fool’	<i>cāti</i>	à x	‘lover’
<i>kusang-kusang</i>	x à   x à	‘model’	<i>kusang</i>	x à	‘double’	<i>kusang</i>	x à	‘double’
<i>latunkema</i>	à x   à x	‘tiger’	<i>latun</i>	à x	‘lion’	<i>kema</i>	à x	‘stripe’
<i>malingkaron</i>	x à   x à	‘coast, seashore’	<i>maling</i>	x à	‘shore’	<i>karon</i>	à x	‘water’
<i>māvaṅganeng</i>	à x   x à	‘mother’s siblings’	<i>māva</i>	à x	‘mother’	<i>ṅanengan</i>	x   x x	‘siblings’
<i>pikunanding</i>	x à   x à	‘mustache’	<i>piku</i>	à x	‘beard’	<i>nanding</i>	x à	‘lip’
<i>sapayyila</i>	x à   x à	‘limbs’	<i>sapay</i>	x à	‘hand’	<i>yila</i>	à x	‘foot’

မူဟာဝဇ္ဇာ *depangcāti* is noteworthy since it follows the same stress pattern as လံင်-လံင် *apan-apan* in spite of consisting of one foot with a heavy second syllable (မူဟာ *de-pang*) and another with a superheavy first syllable (ဝဇ္ဇာ *cāti*). To avoid a clash, stress is not shifted to the heavy syllable in မူဟာ *depang*, since it is not strictly necessary for it to be stressed:

In the case of *maṅgkaṛon* again the rule operates that prohibits two stressed non-long syllables after another. Thus, even if the first component *maṅ* contains a heavy syllable, stress will not move there. In *latunkema*, the syllable /tun/ is assimilated to [tʉŋ] before the /k/ onset of the next syllable. For one, however, this does not make it heavy, and second, even if it did, stress would stay on the first syllable of the word for the same reason as in *maṅgkaṛon*. The same rule of stress hiatus avoidance operates in *sapayyila*.

### 1.3.2 Intonation

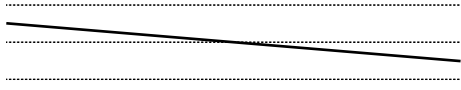
Since intonation contours are notoriously difficult to display in print, I will give very approximate graphs of pitch in the respective examples for each surveyed pattern. Certainly there will be other patterns as well which would require more

<sup>24</sup> Whenever this happens, the fallback is likely to be a mix of German and English, since those are the languages I am most familiar with.

detailed gradations, but for the time being, I will only try to briefly describe those that are most prominent.

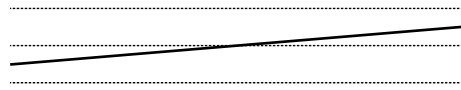
#### Declarative statements

Declarative statements have a gradually falling pitch contour based around an average pitch height, not deviating considerably on both ends:

- (35) 
- Ang gihayo Pintemis minganeri-hen yona.*  
 Ang giha-yo Ø=Pintemis mingan-eri=hen yona  
 AT blow-3SG.N TOP=North Wind ability-INS=all 3SG.N.GEN.  
 ‘The North Wind blew with all of his might.’

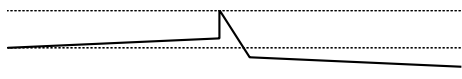
#### Yes-no questions

Since Ayeri does not use a particle or word order to mark closed questions as such, intonation is used to mark the difference from a declarative statement. To achieve a strong contrast, questions exhibit gradually rising intonation:

- (36) 
- Ang gihayo Pintemis minganeri-hen yona?*  
 Ang giha-yo Ø=Pintemis mingan-eri=hen yona  
 AT blow-3SG.N TOP=North Wind ability-INS=all 3SG.N.GEN.  
 ‘Did the North Wind blow with all of his might?’

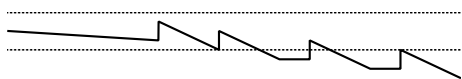
#### ‘Wh-’ questions

Unlike English, Ayeri marks open questions with an in-situ question word. Open questions are thus marked by the question word causing a sharp rise and fall in the overall contour of the question. The first half of the clause has the rising contour of a question, the second half has gradually falling pitch.

- (37) 
- |                  |             |              |             |             |             |
|------------------|-------------|--------------|-------------|-------------|-------------|
| <i>Ang engyo</i> | <i>mico</i> | <i>sinya</i> | <i>luga</i> | <i>toya</i> | <i>sam?</i> |
| Ang eng-yo       | mico        | sinya-Ø      | luga        | toya        | sam         |
| AT be.more-3SG.N | strong      | who-TOP      | among       | 3PL.N.LOC   | two         |
- ‘Who was the stronger of the two?’

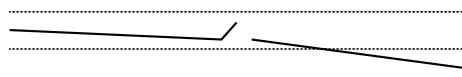
### Lists

List statements have the general gradual downward slope of declarative statements, but the individual items can nonetheless be marked by a pitch rise on the primary accent of each item.

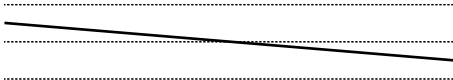
- (38) 
- |                      |               |              |              |            |               |
|----------------------|---------------|--------------|--------------|------------|---------------|
| <i>Le vacyeng</i>    | <i>seygo,</i> | <i>disu,</i> | <i>betay</i> | <i>nay</i> | <i>vasra.</i> |
| Le vac-yeng          | seygo-Ø,      | disu-Ø,      | betay-Ø      | nay        | vasra-Ø       |
| PT.INAN like=3SG.F.A | apple-TOP,    | banana-Ø,    | berry-Ø      | and        | nut-Ø         |
- ‘She likes apples, bananas, berries and nuts.’

### Complement and relative clauses

Complement clauses are characterized by the short spike at the end of the preceding main clause followed by a short break which together signal the beginning of a new syntactic unit within the context of the current sentence, which is broadly similar to list statements. Otherwise, statements with complement clauses as well bear the overall downward-sloping contour of declarative statements if included in such.

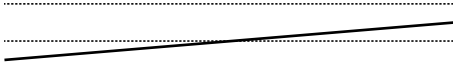
- (39) 
- |                           |               |             |                 |
|---------------------------|---------------|-------------|-----------------|
| <i>Ang manga rantong,</i> | <i>engyo</i>  | <i>mico</i> | <i>sinyāng.</i> |
| Ang manga ran-tong,       | eng-yo        | mico        | sinya-ang       |
| AT PROG argue=3PL.N.A,    | be.more-3SG.N | strong      | who-A           |
- ‘They were arguing who is stronger.’

Relative clauses, on the other hand, do not receive special prosodic marking, but are treated the same as other basic sentence types. They display a continuous downward slope if part of a declarative statement, or a continuous upward slope if part of a question:

(40) a. 

*Lugaya asāyāng si sitang-naykonyāng kong tovaya.*  
 Luga-ya asāya-ang si sitang-naykon-yāng kong tova-ya  
 pass-3SG.M traveler-A REL self=wrap=3SG.M.A inside cloak-LOC

‘A traveler passed who had wrapped himself into a cloak.’

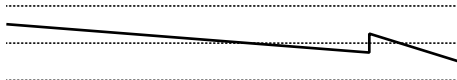
b. 

*Adareng asāyās si le ninyāng tova?*  
 Ada-reng asāya-as si le nin-yāng tova-Ø  
 that-A.INAN traveler-P REL PT.INAN wear=3SG.M.A coat-TOP

‘Is that the traveler who wore the coat?’

#### Contrast

Ayeri uses a kind of topic system for highlighting constituents in a clause by morphosyntactic means, but this is still different from emphasis on semantic grounds, for example when the speaker wants to highlight a semantic difference in the same syntactic position, as in the following example, which presents a possible answer to the question posed in (40b):

(41) 

*Adareng asāyās si le nin-yāng kegan.*  
 Ada-reng asāya-as si le nin-yāng kegan-Ø  
 that-A.INAN traveler-P REL PT.INAN wear=3SG.M.A hat-TOP

‘It is the traveler who wore the *hat*.’

We can see here a spike towards the end of the utterance where the word *kegan* ‘hat’ is placed. This word receives extra stress for contrast with *tova* ‘coat’, which is what the other person had asked about.





## 2 Writing system

In the previous chapter, example words were given in Ayerī's script,  *Tabano Hikamu*, wherever possible. Thus, it seems advisable to include a description of Ayerī's native writing system here as well. Literally,  *Tabano Hikamu* means 'Round Script' (script round), which is an old formation based on the word  *taban-* 'write' that stuck. The current word for 'script' is  *tabanan* 'writing'. Tahano Hikamu was originally named thus because of an earlier draft for a script that never made it very far beyond the drawing board and which was a lot more angular and boxy, see Figure 2.1—Tahano Hikamu was a lot more bubbly in comparison, especially early on (Figure 0.2).<sup>1</sup>

As we have seen in the previous chapter, Ayeri's prosody strongly emphasizes the syllable as a unit. Thus, it is not a surprise that Ayeri's native script, Tahano Hikamu, is an alphasyllabary similar to the Brāhmī alphabets of India and Southeast Asia (Salomon 1996; Court 1996). Scripts like these are

based on the unit of the graphic “syllable” [...], which by definition always ends with a vowel (type V, CV, CCV, etc.). Syllables consisting of a vowel only (usually at the beginning of a word or sentence) are written with the *full* or *initial vowel signs* [...]. But when, as is much more frequently the case, the syllable consists of a consonant followed by a vowel, the vowel is indicated by a diacritic sign attached to the basic sign for the consonant [...]. (Salomon 1996: 376)

For Tahano Hikamu the definition that a syllable consisting only of a vowel is written with an initial vowel sign is only true under certain circumstances, as we will see below. Moreover, Brāhmī scripts are often characterized by conjuncts of clustered consonants which may become quite large and sometimes behave in an idiosyncratic way. Consonant conjuncts like Devanāgarī त्व ⟨tva⟩ from त ⟨ta⟩ + व ⟨va⟩ or idiosyncratic conjuncts like क्ष ⟨kṣa⟩ for क ⟨ka⟩ + ष ⟨ṣa⟩ are not known in Tahano Hikamu, however. Tahano Hikamu also does not know subscript notation for consonant clusters and special diacritics marking coda consonants like in Javanese (Kuipers and McDermott 1996: 478–479). This does not mean, however,

<sup>1</sup> Unfortunately, there is no documentation of the Box script surviving that I know of.

Figure 2.1: Box script and Hikamu



(a) Old and aborted draft: Box script



(b) Ayeri's native script: Tahano Hikamu

that final consonants are simply omitted in writing, since closed syllables are reasonably common enough in Ayeri to warrant indicating them. Thus, there is “a special mark to eliminate the vowel of the previous syllable, thereby leaving a consonant in a syllable-final position.” (Kuipers and McDermott 1996: 476) That is, a diacritic exists which marks the absence of an inherent vowel, rendering the syllable consonant-only.

Another difference from Brāhmī-family scripts is that vowel length and diphthongs in [i] are indicated by dedicated diacritics, so the long vowels are not doubled versions of their short counterparts. Like in Kharoṣṭhī—another historically important ancient script of India—initial vowels are not represented by unique graphemes but they are all written like post-consonantal vowel diacritics (Salomon 1996: 377), though in Tahano Hikamu with a character without an inherent sound value. For this reason, the character is indicated in the table below as  $\emptyset$ ; its native name is  $\text{ṛṇṇ}$  *ranyan* ‘nothing’.<sup>2</sup> Similar to a number of Brāhmī scripts, Tahano Hikamu puts diacritics not only below or above consonant bases, but also before them. This, however, is not limited to vowel graphemes as in Devanāgarī ि ⟨i⟩ or Javanese ꦺ ⟨e, é/è⟩ (Kuipers and McDermott 1996: 478).<sup>3</sup>

<sup>2</sup> I will give the native names of graphemes here, but will refer to them by their English names for clarity in the running text.

<sup>3</sup> Kuipers and McDermott (1996) do not say, but it looks likely to me that both are related, since they are both functionally the only prepended vowel diacritics and both represent a high front sound.

## 2.1 Consonants

Tahano Hikamu is mainly based on consonant bases that are modified by diacritics. Since the vowel /a/ is so highly frequent in Ayeri, it is also the vowel that is *inherent* to every consonant grapheme if not further modified by vowel diacritics. Consonant letters are simply referred to as *pa*, *ta*, *ka*, ... Figure 1.1 displays all the main consonants. The customary collation is—similar to the IPA table—roughly grouping the letters according to their sound value by anteriority (front → back) and sonority (low → high). The script is monocameral, that is, there is no distinction between capital letters and minuscule letters as in the Latin, Greek, Cyrillic, Georgian, and Armenian alphabet. It is also written in lines from left to right.

Figure 2.2: The consonant graphemes of Tahano Hikamu

/pa/	/ta/	/ka/	/ba/	/da/	/ga/
ᠠ	ᠲ	ᠬ	ᠪ	ᠳ	ᠭ
/ma/	/na/	/ŋa/	/va/	/sa/	/ha/
ᠮ	ᠨ	ᠩ	ᠪ	ᠰ	ᠬ
/ra/	/la/	/ja/	/Ø/		
ᠷ	ᠯ	ᠵ	ᠯ		

ᠯ, which in Ayeri has no sound value but is used as a base for initial vowels, may also serve as the character for /ʔa/. What is, moreover, interesting about ᠩ <nga> is that even though before, /ŋ/ was treated strictly as a coda consonant in the previous chapter, it is in fact treated as an onset consonant in writing if a vowel is following:

- (1) ᠠ + ᠩᠰᠢᠰ  
 /pa/     /ŋis/  
 ᠠᠩᠰᠢᠰ *pangis* /paŋ.is/ ‘money’

Tahano Hikamu knows a few ligatures. First of all, when two ᠨ <na> are in succession within a word, they will form a ligature ᠨᠨ <nana>:

- (2) ᠨ + ᠨ → ᠨᠨ  
 /na/     /na/     /nana/

This is distinct from conjuncts like in Devanāgarī et al., though, since the unmodified sound value will still be /nana/, not \*/nna/, so the inherent vowel of each ᠨ

(3) a.  $\mathfrak{K} \langle k\mathfrak{w}a \rangle \leftarrow \mathfrak{K} \langle k\mathfrak{a} \rangle + \mathfrak{r} \langle \mathfrak{v}a \rangle$ ,  
 b.  $\mathfrak{K} \langle \mathfrak{t}sa \rangle \leftarrow \mathfrak{K} \langle \mathfrak{t}a \rangle + \mathfrak{r} \langle \mathfrak{s}a \rangle$ , and  
 c.  $\mathfrak{K} \langle \mathfrak{k}sa \rangle \leftarrow \mathfrak{K} \langle \mathfrak{k}a \rangle + \mathfrak{r} \langle \mathfrak{s}a \rangle$ .

Figure 2.3: Additional consonant graphemes of Tahano Hikamu

/fa/	/wa/	/tsa/	/za/	/ja/	/za/
ᆞ	ᆡ	ᆢ	ᆣ	ᆤ	ᆥ
/ça/	/ksa/	/kwa/	/xa/	/ya/	
ᆦ	ᆧ	ᆨ	ᆩ	ᆪ	

As mentioned above, vowels are written as diacritics that are added to consonants. In principle, every consonant has two slots for vowels, a primary one atop it, and a secondary one below it. Vowels added to consonants in the primary slot delete their inherent /a/:

Figure 2.4 gives the primary vowel signs. Of the vowel signs given there, only ॐ ⟨ə⟩ is not used in Ayeri. ॐ ⟨au⟩ is the only diphthong for which a dedicated grapheme exists, even though its occurrence is rather limited. The independent vowel graphemes are used at the beginning of words or inside words when there is no other way to spell the vowel, which is occasionally the case for secondary vowels. Secondary vowels are vowels that are not parts of diphthongs (even though another language might use them to spell diphthongs that are not covered by default), but follow the vowel of a syllable directly. They are attached underneath a consonant base, for example:

Figure 2.4: Primary vowel graphemes of Tahano Hikamu

	/i/	/e/	/a/	/o/	/u/	/ɔ/	/au/
Diaritics							
Independent							

$$(5) \quad \text{ɸ} \rightarrow \text{ɸ}^{\text{ˆ}} \rightarrow \text{ɸ}^{\text{ˆ}}_{\text{ˆ}}$$

/ja/      /je/      /jea/

In fact, the principle that every consonant base with its diacritics represents one syllable is slightly violated here, which is also the reason why secondary vowels very occasionally need to be spelled as independent vowels, for example when the secondary vowel is long, as in the word *ruān* ‘duty’:

$$(6) \quad \text{r}^{\text{ˆ}} \rightarrow \text{r}^{\text{ˆ}}_{\text{ˆ}} \quad \left( \text{r}^{\text{ˆ}}_{\text{ˆ}} \right)$$

/ru/      /rwa:/      /ru:a/

Example (6) uses a diacritic, to indicate length. If is put directly under *ru* (the moves down where it is not in the way), the syllable will incorrectly spell /ru:a/ instead of the intended /rua:/. This is because diacritics modify consonants and primary vowels, but there is no way to modify a secondary vowel directly. Figure 2.5 gives a list of secondary vowels corresponding to that of primary vowels above. The vowels as well are just referred to by their sound value; ‘primary’ and ‘secondary’, ‘superscript’ and ‘subscript’ or ‘upper’ and ‘lower’ may be chosen to disambiguate their positions; the native names may use *iray* ‘high’ and *eyra* ‘low’ to disambiguate, so *e iray* denotes the superscript ⟨e⟩ diacritic while *e eyra* denotes its subscript counterpart.

Figure 2.5: Secondary vowel graphemes of Tahano Hikamu

/i/	/e/	/a/	/o/	/u/	/ɔ/	/au/

As a further exception, those consonant bases with an ascender ( ⟨ka⟩, ⟨da⟩, ⟨ɕa/⟩) move the primary vowel to the secondary slot below the consonant by default while indicating the vacancy of the primary slot at the top with a dot. This is done to avoid crossing the ascender of the consonant with a vowel diacritic:

$$(7) \quad \begin{array}{ccccc} \text{𐌵} & \longrightarrow & \text{𐌵} & \longrightarrow & \text{𐌵} \\ /ka/ & & /ka.i/ & & /ki/ \end{array}$$

If the primary vowel slot were not silenced by the  $\circ$  diacritic, it could reasonably be assumed that the consonant is not losing its inherent /a/ and the vowel below the consonant indicates a secondary vowel, spelling /CaV/. If, however, a secondary vowel is *actually* added, primary and secondary vowels will be assigned the regular primary and secondary slots, respectively, again (8a). This condition also holds true for subscript diacritics (8b).

$$(8) \quad \begin{array}{ll} \text{a.} & \begin{array}{ccc} \text{𐌵} & \longrightarrow & \text{𐌵} \\ /ki/ & & /ki.e/ \end{array} \\ \text{b.} & \begin{array}{ccc} \text{𐌵} & \longrightarrow & \text{𐌵} \\ /ki/ & & /ki:/ \end{array} \end{array}$$

The order of secondary vowels and subscript diacritics is iconic insofar as it follows the order of sounds in the syllable. Thus, secondary vowels appear below the consonant-doubling diacritic,  $\text{𐌵}$ , while they appear above the syllable-final homorganic nasal diacritic,  $\text{𐌶}$ :

$$(9) \quad \begin{array}{ll} \text{a.} & \begin{array}{ccc} \text{𐌵} & \longrightarrow & \text{𐌵} \\ /ppa/ & \longrightarrow & /ppea/ \end{array} \\ \text{b.} & \begin{array}{ccc} \text{𐌵} & \longrightarrow & \text{𐌵} \\ /peN/ & \longrightarrow & /peaN/ \end{array} \end{array}$$

## 2.3 Diacritics

We have already encountered a few diacritics, though Tahano Hikamu comes with a lot more, some of which undergo non-trivial positioning and repositioning rules. As vowels are primarily expressed as superscripts, diacritics are primarily realized as subscripts, so in the following I will first describe subscript diacritics; then prepended diacritics, which Ayeri also has a number of, both as graphemes in their own right and as allographs of other subscript diacritics; and then, lastly, superscript diacritics.



### 2.3.1 Subscript diacritics

Figure 2.6 shows the bottom-attaching diacritics. The ‘large diacritics’ cause the secondary slot of consonants to move down below the diacritic. ‘Small diacritics’ can attach in this place as well as secondary vowels, as does the homorganic nasal diacritic ၂ in this diacritic-fraught example:

$$(10) \quad \begin{array}{ccc} \text{တၢ်နၢ်} & + & \text{ပုၤလၢ်} \rightarrow \text{တၢ်နၢ်ပုၤလၢ်} \\ /'tʃa:n/ & /pu'lɔɪ/ & /,tʃa:mpu'lɔɪ/ \\ \text{တၢ်နၢ်ပုၤလၢ်} & \text{cāmpuluy} & \text{'heterosexual'} \end{array}$$

It also needs to be noted that diacritics like ၂ are applied progressively to words as a whole, not stopping at morpheme and syllable boundaries, so even though *toryeng* ‘she sleeps’ may be composed of တၢ် tor- ‘sleep’ + ပုၤ -yeng (=3SG.F.A) and syllabifies as /tor.'jeŋ/, the spelling is not \*တၢ်ပုၤ as one might expect, but တၢ်ပုၤ.

Even though the primary position for small diacritics is underneath consonants, the diacritic deleting the inherent vowel, ၂, very commonly also appears after a consonant letter at the end of words:

$$(11) \quad \begin{array}{cccc} \text{ပ} & \text{နီၤကၢ်} & \text{ကၢ်} & \text{နီၤကၢ်} \\ \text{Ya} & \text{nimreng} & \text{pangan} & \text{narānyena.} \\ \text{Ya} & \text{nim-reng} & \text{pangan-Ø} & \text{narān-ye-na} \\ \text{LOCT} & \text{appear=3SG.INAN.A} & \text{end-TOP} & \text{word-PL-GEN} \end{array}$$

‘It appears at the end of words.’

This strategy is advantageous in that Tahano Hikamu leaves very little space between individual words: ပၤနီၤကၢ်ကၢ်နီၤကၢ်. With the dot after the consonant, word boundaries are more visible.

### 2.3.2 Prepended diacritics

Example (10) leads us directly to the next class of diacritics—ones that are prepended to the consonant letter, either because they are simply placed there or because of allography. Let us first list those diacritics that appear in front of consonants obligatorily (Figure 2.7).

As Figure 2.7 shows, the only obligatorily prepended diacritic that Ayeri uses is the one that marks diphthongs, ၂. It needs to be noted here that ၂ changes into ပ (ya) proper when a vowel follows, but stays ၂ when a ပ (ya) follows:

<sup>4</sup> In a Tahano Hikamu orthography I devised for English once, ၂ was used for /ə/, as in the NURSE vowel in American English: ၂ nurse.



Figure 2.7: Obligatorily prepended diacritics of Tahano Hikamu

Native name	Function	Example
ꠘꠤꠎꠤꠎꠤꠎꠤ <i>lentankusang</i> 'double-sound'	Marks a diphthong with /ɪ/	ꠎ <i>pe</i> → ꠎꠤ <i>pey</i>
ꠇꠤꠎꠤꠎꠤ <i>tilamaya</i> 'changer'	Marks raised vowels (i.e. umlaut; not used in Ayeri)	ꠎ <i>po</i> → ꠎꠤ <i>/pø/</i>
ꠇꠤꠎꠤꠎꠤ <i>hiyamaya</i> 'roller'	Marks retroflex consonants (not used in Ayeri) <sup>4</sup>	ꠎ <i>ta</i> → ꠎꠤ <i>/tʰa/</i>

Figure 2.8: Allographically prepended diacritics of Tahano Hikamu

Native name	Function	Example
ꠇꠤꠎꠤꠎꠤꠎꠤ <i>tupasati marin</i> 'anterior long-maker'	Lengthens the primary vowel of the syllable	ꠎ <i>sya</i> → ꠎꠤ <i>syā</i> , ꠎ <i>na</i> → ꠎꠤ <i>nā</i>
ꠇꠤꠎꠤꠎꠤ <i>ya marin</i> 'anterior ya'	⟨ya⟩ following another consonant, also across syllables.	ꠎ <i>na</i> → ꠎꠤ <i>nya</i>
ꠎꠤꠎꠤꠎꠤꠎꠤ <i>ringaya marin</i> 'anterior raiser'	Also used as an allograph for the palatalization proper diacritic.	ꠎ <i>/sʰa/</i> → ꠎꠤ <i>/sʰj/</i>
ꠇꠤꠎꠤꠎꠤꠎꠤ <i>ulangaya marin</i> 'anterior breather'	(Pre-)Aspiration or frication of a consonant (not used in Ayeri)	ꠎ <i>nga</i> → ꠎꠤ <i>/ŋʰa/</i> ; ꠎ <i>ta</i> → ꠎꠤ <i>/tʰa/</i>

- (12) a.  $\text{ဃာပျံ } \textit{haday}$  ‘hero’  $\rightarrow$   $\text{ဃာပျံဟံ } \textit{hadayang}$  ‘the hero’ (hero-A); \* $\text{ဃာပျံခါး}$ ;  
 b.  $\text{မိနီ } \textit{tipuy}$  ‘grass’  $\rightarrow$   $\text{မိနီယံ } \textit{tipuyya}$  ‘in the grass’ (grass-LOC); \* $\text{မိနီယ}$ .

Besides  $\text{နိ}$ , there are also a number of diacritics that are also obligatorily prepended to consonants, but do so as context-sensitive allographs (Figure 2.8). The selection of the variant diacritics is not random or up to the aesthetic eye of the writer (even though the device itself is certainly a matter of aesthetics), but it is governed by rules. The prepended forms listed in Figure 2.8 are thus triggered

1. when there is no stem or bowl for the regular subscript diacritic to attach to, which is the case for  $\text{န}$  <na>,  $\text{ဂ}$  <nga>,  $\text{ဝ}$  <va>, and  $\text{ဏ}$  <wa>:

- (13) a.  $\text{န} \rightarrow \text{ဝန}$  c.  $\text{ဝ} \rightarrow \text{ဝဝ}$   
           /na/       /na:/                   /va/       /va:/  
 b.  $\text{ဂ} \rightarrow \text{ဝဂ}$  d.  $\text{ဏ} \rightarrow \text{ဝဏ}$   
           /ŋa/       /ŋa:/                   /wa/       /wa:/

2. when a large subscript diacritic would be added after another large subscript diacritic—this position can only be occupied once, so further large subscripts are prepended:

- (14)  $\text{တ} \xrightarrow{+ \text{ဒ}} \text{တဒ} \xrightarrow{+ \text{ဃ}} \text{တဒဃ} \xrightarrow{+ \text{ဝ}} \text{တဒဃဝ} \xrightarrow{+ \text{ခါ}} \text{တဒဃဝခါ}$   
           /ta/       /taʰa/       /taʰja/       /taʰji/       /taʰji:/

The order of diacritics follows the logic of the respective language’s phoneme inventory, so if there are, for example, retroflex consonants and both dental and retroflex consonants can be aspirated, retroflexion would be marked first, then aspiration. If there is a palatalization contrast on top of this, the diacritic would be added after aspiration.

When adding large diacritics to stemless consonants, they are prepended from the beginning, as we saw in (13), and just like in (14), this principle continues:

- (15)  $\text{န} \xrightarrow{+ \text{ဃ}} \text{နဃ} \xrightarrow{+ \text{ခါ}} \text{နဃခါ} \xrightarrow{+ \text{နိ}} \text{နိနဃခါ}$   
           /na/       /ŋja/       /ŋja:/       /ŋja:ɪ/

3. with consonants directly following ခ (na), to avoid a clash with its swash:

$$(16) \quad \begin{array}{ccccc} \text{ခ} & + & \text{ပ} & \rightarrow & \text{ခပ} & (* \text{ခပ}) \\ /na/ & & /pa:/ & & /napa:/ & \end{array}$$

An exception to this exception occurs, however, when the consonant is not directly following. In this case, no reordering happens, only ခ (na) *may* reduce its swash in size to accommodate the following prepended diacritic:<sup>5</sup>

$$(17) \quad \begin{array}{ccccc} \text{ခ} & + & \text{ပာ} & \rightarrow & \text{ခပာ} & (? \text{ခပာ}) \\ /na/ & & /pa:/ & & /napa:/ & \end{array}$$

4. in other cases where a clash of subscript diacritics needs to be avoided:

$$(18) \quad \begin{array}{ccccc} \text{ဒ်} & + & \text{ပ} & \rightarrow & \text{ဒ်ပ} & (* \text{ဒ်ပ}) \\ /di/ & & /pa:/ & & /dipa:/ & \end{array}$$

Alternatively, the following solution is permissible:

$$(19) \quad \begin{array}{ccccc} \text{ဒ်} & + & \text{ပ} & \rightarrow & \text{ဒ်ပ} \\ /di/ & & /pa:/ & & /dipa:/ & \end{array}$$

When two long syllables follow each other, as in *bāmā* ‘mom-and-dad’, one of the length diacritics should definitely be pulled to the front:

$$(20) \quad \begin{array}{ccccc} \text{ပ} & + & \text{မ} & \rightarrow & \text{ပမ} & (? \text{ပမ}) \\ \text{or: } \text{ပ} & + & \text{မ} & \rightarrow & \text{ပမ} \\ /ba:/ & & /ma:/ & & /ba:ma:/ & \end{array}$$

<sup>5</sup> The font I am using here is designed so that the reduced combination looks nicer, but if unreduced, ခ (na)’s swash is not so long as to cross the descender of ပ either in this particular case.

### 2.3.3 Superscript diacritics

Ayeri's standard position for diacritics is below consonants, but sometimes it is nicer to put them on top, especially for the letter ၃ ⟨na⟩ due to its swash, as well as for ၃ ⟨va⟩ since the space below its flag is empty otherwise, thus not providing much of a visual connection. The only diacritic that is normally attaching to the top of consonants is that for the glottal stop—we have already encountered its subscript allograph earlier. Since Ayeri's phoneme inventory does not possess a phonemic glottal stop or glottalization, this diacritic is not used in Ayeri. The list of superscript diacritics is given in Figure 2.9.

Figure 2.9: Superscript diacritics of Tahano Hikamu

Native name	Function	Example
◌် ၵ်းမုၢ်ဂံၢ် <i>godaya</i> <i>ling</i> 'upper extinguisher'	Deletes inherent /a/ of consonant, e.g. in consonant clusters or closed syllables	ၵ်း <i>vara</i> → ၵ်း <i>vra</i>
◌် ၵ်းသီၣ် <i>vināti</i> <i>ling</i> 'upper nasalizer'	Indicates a homorganic nasal or nasalizes the vowel, depending on language/context	ၵ်း <i>naka</i> → ၵ်း <i>nanka</i> /naŋka/ or /nāka/
◌် ၵ်းကံၢ်သီၣ် <i>kusangisāti</i> <i>ling</i> 'upper duplicator'	Indicates a geminated or otherwise double consonant	ၵ်း <i>pana</i> → ၵ်း <i>panna</i>
◌် ၵ်းယု <i>raypāya</i> 'stopper'	Glottal stop coda or glottalization of a consonant (not used in Ayeri)	ၵ်း <i>ta</i> → ၵ်း /taʔ/; ၵ်း <i>sa</i> → ၵ်း /s'a/

At times, it may be necessary to attach both a superscript diacritic and a vowel sign above a consonant. In this case, the consonant-modifying diacritic is placed first and the vowel diacritic on top of it—this is exactly equivalent to the rule exemplified for subscript diacritics in (9).

## 2.4 Numerals

Ayeri uses a duodecimal number system, that is, a system based on the powers of 12, which is a typological rarity.<sup>6</sup> There is a digit for zero, so the system is positional, like the Hindu–Arabic digits used by the Latin alphabet. The numerals for the numbers from 1 to 12 are shown in Figure 2.10.

Figure 2.10: The numerals of Tahano Hikamu

1	2	3	4	5	6
12	𐌒	𐌓	𐌔	𐌕	𐌖
7	8	9	A	B	10
𐌗	𐌘	𐌙	𐌚	𐌛	𐌜

## 2.5 Punctuation and abbreviations

Tahano Hikamu’s system of manipulating the sound of syllables is very sophisticated, so it comes as no surprise that it is also host of a large number of punctuation marks. Figure 2.11 lists the ones commonly encountered, Figure 2.12 the ones not so commonly encountered.

• ⟨.⟩ does not look very much like a dot or a point, but it is derived from a sign that looks like two circles stacked on top of each other, similar to : ⟨-⟩ (see Figure 0.2). There is no mark for a comma as such, so : ⟨-⟩ or – ⟨-⟩ cannot be used in this way. Instead of a comma, a wide word space is used to separate syntactic units. A long dash — ⟨—⟩ is also sometimes found at the end of paragraphs or texts to mark their end. The strong exclamation mark ~ may appear in its exclamatory function at the end of a line, but does not necessarily indicate strong emphatic force in this case, but just an emphatic statement.

Regarding the less common marks, some of these seem like all to bland copies of modern punctuation, especially the brackets and the decimal point. Still, however, they may serve their purpose sometimes, and the brackets 𐌛𐌜 maybe come with the redeeming notion that they push off the text around the inclusion rather than encapsulating the inclusion within it, so the visual effect is slightly different. The name brackets 𐌛𐌜 are interesting and maybe useful insofar as many names in Ayeri are derived from common words, for example, 𐌛𐌜𐌚 *Ajān*, a male name, is literally ‘play, game’, relating to a playful character; 𐌛𐌜𐌚 *Migoray*, a female name, literally means ‘flower’. The name brackets, then, make it unmistakeably clear that a proper

<sup>6</sup> And one possibly overrepresented by fictional languages due to its rarity in natural languages.

Figure 2.11: Common punctuation marks of Tahano Hikamu

	Native name	Function	Example
	၎် dan ‘dot’	Full stop	ဓာယျာ၎် Sarayāng. ‘He left.’
:	၎်၎် dan-dan ‘little dot’	A separator for small things, like clitics and abbreviations; divides the constituents of reduplication	အိ၎်၎် ada-nanga ‘this house’; နာ၎် ၄:pd ‘၄ hrs’; ၎်၎် dan-dan ‘dot-dot, little dot’
—	၎်၎် puntān ‘dash’	General sign for a longer pause, equivalent to a dash, colon, semicolon, brackets	ယ၎်-ဓာရ် Yan – saru! ‘Yan – go!’
၎်	၎်၎်၎် damprantan ‘question point’	Marks questions	ဓာရ်၎် Manisu? ‘Hello?’
၎်	၎်၎်၎် dambahān ‘shouting point’	Marks exclamations; strong exclamations may be marked by the ၎် variant.	ဓာရ်၎် Manisu! ‘Hello!'; ဖိ Yi! ‘Urgh!’

Figure 2.12: Less common punctuation marks of Tahano Hikamu

	Native name	Function	Example
၎်	၎်၎်၎် danarān ‘speaking point’	Quotation marks	ဓာယျာ၎် ‘ဓာရ်၎် Narayāng “Manisu!” ‘He says, “Hello!”’
၎်	၎်၎်၎် dankayvo ‘beside-point’	Bracketing of text	ဓာယျာ၎်၎်း babis (larau) ‘a (nice) day’
၎်	၎်၎်၎် dangaran ‘name-point’	Explicitly marks a name as such. For the closing name parenthesis, ၎် can be found as well.	အိ၎်၎်၎်၎် Ajān Savati; ဂိ၎်၎်၎်၎် Pila Lay Maran
’	၎်၎်၎် dansinday ‘number-point’	Marks (duo)decimal fractions	၎်၎်၎်၎် ၁၇.၄၅၈၈၂ ‘19.37482’
၎်	၎်၎်၎် adrumaya ‘breaker’	Marks line breaks within a phrase	



subscript letters and in Ayeri, and the number of large swirling diacritics is also rather low, so there is still definitely a difference in appearance. The ‘angular’ style is also the one that is comparable in function to our bold face or italic style letters, since it is used in captions or to highlight special text within running text. This letter style (ဆိ၃၃ *binya* ‘angular’) is displayed in Figure 2.14.

Figure 2.14: Tahano Hikamu, ‘angular style’

The image shows two lines of handwritten text in the 'angular style'. The characters are bold and angular, with straight lines and sharp corners. The first line reads: 'က ကံလိၣ် နုၣ်ပုၣ်ဝဲၣ် ဖဲၣ်ကံၣ် နုၣ် နုၣ်ကံၣ် နီၣ် နုၣ် နီၣ်။' The second line reads: 'ဝဲၣ် နီၣ်ကံၣ် နီၣ်ကံၣ် နုၣ် ဝဲၣ်ကံၣ် နုၣ် ဝဲၣ် နုၣ် နီၣ်ကံၣ် နီၣ်ကံၣ်။' The text is written in a dark ink on a light background.

The greatest difference to the ‘book’ style is that many of the main strokes double to become a thick and a parallel thin line and the shape of န <na> changes to have its swirl straightened into a simple descending line. The vowel carrier န changes to a flattened *O*-like circle, and the bottom curl in န <ta> changes to a wedge. While the right side of the န <sa> character in the ‘book style’ consists of two strokes—a flag and a downwards bow, both independently attached to the main stem—they connect here to form an *R*-like shape.

Neatly reproducing the shapes of either the ‘book’ style or the ‘angular’ style by hand goes rather slowly, so I was wondering what daily handwriting could look like. Of course, this presupposes pen and paper again; Salomon (1996: 377) mentions that inscriptions of Brāhmī and related scripts have been found on copper plates and plates made of other metals, besides stone. Metal plates can be inscribed with metal styluses and should allow similar shapes as modern pens. Wax tablets—a staple in European literacy until the use of paper became widespread—should as well allow for relative freedom of stroke direction. Figure 2.15 shows what Tahano Hikamu might look like quickly jotted down by hand.

Figure 2.15: Tahano Hikamu, ‘hand style’

The image shows two lines of handwritten text in the 'hand style'. The characters are simplified and more rounded than in the 'angular style'. The first line reads: 'က ကံလိၣ် နုၣ်ပုၣ်ဝဲၣ် နီၣ်ကံၣ် နုၣ် နုၣ်ကံၣ် နီၣ် နုၣ် နီၣ်။' The second line reads: 'ဝဲၣ် နီၣ်ကံၣ် နီၣ်ကံၣ် နုၣ် ဝဲၣ်ကံၣ် နုၣ် ဝဲၣ် နုၣ် နီၣ်ကံၣ် နီၣ်ကံၣ်။' The text is written in a dark ink on a light background.

Many letter shapes become simplified, specifically န <ba>, န <ga>, န <ka>, န <na>, န <nga>, the vowel carrier န, and the vowel န <i>. Not shown here is the vowel length diacritic, န, which is simplified to a reverse *C* shape. The abbreviation န *na* ‘and’ is used throughout, though in a shape that is more similar to its ‘angular’ form န. န <na> is also taken from the ‘angular’ style န, which opens the possibility that



this is actually the basic shape rather than the ‘book’ style’s ၃, or both are different developments from a shared ancestor.

Most recently, I also wondered what Tahano Hikamu might look like if they were adapted to European blackletter style with its characteristic broken arches. This, of course, constitutes a sharp contrast to Ayeri’s usual look and feel, which made the experiment all the more interesting, though decidedly non-‘canonic’. Figure 2.16 shows what our example passage might have looked like at a time when Gothic book hands flourished.

Figure 2.16: Tahano Hikamu, ‘blackletter style’



The letter shapes from the ‘book’ style stay largely intact here, though all curves are broken up into at least two strokes, and strokes from the bottom left to the top right, which would push a quill in a way that causes ink to splatter, are avoided completely. The characters that differ most are ၂ (ga), ၃ (ra), ၄ (nga), and the vowel carrier ၅. ၆ (na) again appears in the ‘angular’ shape, though without its descender word-internally and in the abbreviation ၇ *nay*. ၈ (ta) comes with a horizontal stroke instead of a curl at the bottom; ၉ (sa) gains a descender, as does ၁၀ (ra). Not shown here either are changes to the ‘large’ diacritics.



## 3 Morphological typology

---

The first chapter dealt with the smallest constituent parts of words—speech sounds, which ones there are, and how they assemble into valid words. Consequentially, the following chapters will be about the next step up from this: morphemes, the atoms of meaning. First we will have a more general look at which kinds of morphemes there are, and then look at them more closely by part of speech. This chapter on morphological typology will first deal with general questions about Ayeri’s degree of synthesis, and then will try to answer questions about the kinds of processes the various morpheme classes carry out in the language.

### 3.1 Typology

For the largest part, Ayeri is an *agglutinative* language since it modifies word roots with affixes for the purposes of inflection and derivation, and these affixes, in the form of suffixes more specifically, can be stacked, especially on verbs; the average number of morphemes per word is thus greater than 1:

- (1) *Le kondasayāng hemaye pruyya nay napayya kayvay.*  
Le kond-asa-yāng hema-ye-Ø pruy-ya nay napay-ya kayvay  
PT.INAN eat-HAB=3SG.M.A egg-PL-TOP salt-LOC and pepper-LOC without  
‘He always eats his eggs without salt and pepper.’

The verb root 𐌕𐌔 *kond-* ‘eat’ is inflected here for a habitual action with the suffix 𐌕𐌔𐌕𐌔 *-asa*, and also carries a person-inflection clitic, 𐌕𐌔𐌕𐌔 *-yāng*, marking a third person singular masculine agent. With the exception of person-inflection clitics, affixes tend to encode a single grammatical function. Verbs are not the only part of speech that can inflect; nouns and the relativizing conjunction can as well:

- (2) a. *Ang mətabanay tamanyeley yeyam.*  
 Ang mə-tahan-ay.Ø taman-ye-ley yeyam.  
 AT PST-write=ISG.TOP letter-PL-P.INAN 3SG.F.DAT  
 ‘I wrote letters to her.’
- b. *Le turayāng taman sinā ang ningay*  
 Le tura-yāng taman-Ø si-Ø-na ang ning=ay.Ø  
 PT.INAN send=3SG.M.A letter-TOP REL-PT.INAN-GEN AT tell=ISG.TOP  
*tamala vās.*  
 tamala vās  
 yesterday 2SG.P

‘The letter which I told you about yesterday, he sent it.’

The principle of not conflating several grammatical functions into a single suffix can be observed in (2a) regarding the word *tamanyeley* ‘letters’, in which the plural marker *-ye* is distinct from the inanimate-patient case marker *-ley*. Strictly speaking, the pronoun *yeyam* ‘to her’ is also composed, namely of the third person feminine base form *ye* and the dative case marker *yam*. Example (2b) is one we have already encountered before (p. 13). Here, the relative pronoun, *sinā* ‘of/about which’ is inflected for genitive case, and stress on the usually unstressed last syllable suprasegmentally marks that this form is contracted from *sileyena* (*si-ley-ena*, REL-P.INAN-GEN).

So far, we have concentrated on suffixes, but there are a number of prefixes as well; (2a) exhibits the past prefix *mə-* (which is actually redundant in this case). There are also deictic prefixes on nouns, however. In the following example, the clitic *eda-* ‘this-’ joins the noun *peham* ‘carpet’ to indicate a specific carpet.

- (3) *Le no intoyyang eda-peham.*  
 Le no int-oy-yang eda-peham-Ø  
 PT.INAN want buy-NEG=ISG.A this=carpet-TOP

‘I do not want to buy this carpet.’

Besides prefixes and suffixes, Ayeri also possesses at least one grammatical morpheme of the kind Zwicky (1977) calls a ‘bound word’. These are cases where morphemes which are

always bound and always unaccented show considerable syntactic freedom, in the sense that they can be associated with words of a variety of morphosyntactic categories. Frequently, such a *bound word* is semantically associated with an entire constituent while being phonologically attached to one word of this constituent, and ordinarily the bound word is located at the very margins of the word, standing outside even inflectional affixes. (6)

This is the case with the marker  $\text{ᄇᆞᆫ}$  *manga*, which is treated as an independent word, but can modify verbs and prepositions—heads of VPs and PPs, respectively—is unstressed and appears at the margin of its modification target:

- (4) a. *Ang manga yavaya ayon bariley.*  
 Ang manga yava-ya ayon-Ø bari-ley  
 AT PROG roast-3SG.M man-TOP meat-P.INAN  
 ‘The man is roasting meat.’
- b. *Ya mətapyyāng maritay misley manga luga bari.*  
 Ya mə-tapy-yāng maritay mis-ley manga luga bari-Ø  
 LOCT PST-put=3SG.M.A before spit-P.INAN DYN between meat-TOP  
 ‘The meat, he had put a spit through it before.’

In (4a),  $\text{ᄇᆞᆫ}$  *manga* modifies the verb  $\text{ᄇᆞᆫ}$  *yava-* ‘roast’ and indicates that this is a temporarily ongoing action, like the English progressive, except not as strongly grammaticalized.<sup>1</sup> In (4b),  $\text{ᄇᆞᆫ}$  *manga* modifies the preposition, on the other hand, to indicate that it is dynamic:  $\text{ᄇᆞᆫ}$  *luga* by itself means ‘among, between’, while its dynamic form  $\text{ᄇᆞᆫ}$  *manga luga* means ‘through; during, while, for’.

As we have seen in the examples above, person suffixes on verbs are single morphemes that encode more than one property, for example  $\text{ᄇᆞᆫ}$  *-yeng* encodes the person features third person, feminine, singular, and agent. Personal pronouns, of which the person clitics on verbs are an instance, are the main case of fusion among agglutination in Ayeri, although some of the forms, like  $\text{ᄇᆞᆫ}$  *yeyam* ‘to her’ above, can be decomposed into root and suffix without problem.<sup>2</sup>

Perpendicular to the axis isolation–agglutination runs the axis analytic–synthetic. On the latter axis, Ayeri scores mostly as *synthetic*, since it prefers compactness over spreading a construction over several words, though it does not incorporate object NPs, so it is not going as far as being polysynthetic. It is nonetheless theoretically possible, due to suffixation being a prominent pattern, to form foot-long words like

<sup>1</sup> I suppose, a better parallel is the so-called *rheinische Verlaufsform* ‘Riparian progressive’ (*sein* ‘be’ + *am/beim* ‘at the’ + infinitive) in German, a construction common in the colloquial language which parallels the English progressive construction and is not yet fully grammaticalized (Eisenberg et al. 2016: 435). Speakers will thus accept both *Er lernt gerade*, literally ‘He studies right now’, and *Er ist am Lernen* ‘He is studying’.

<sup>2</sup> Originally, Ayeri’s personal pronouns were indeed agglutinative as well, so  $\text{ᄇᆞᆫ}$  *yeng* ‘she’ used to be  $\text{ᄇᆞᆫ}$  *iyegang* (*iy-e-ang*, 3SG-F-A). This also gives an explanation to Boga et al. (2016)’s observation that Ayeri’s plural pronouns are formed “[v]ielleicht sogar zu regelmäßg” ([15]; ‘possibly in an even too regular way’).

- (5) *da-mətabasongoyyang-ikan*  
 da=mə-taha-asa-ong-oy=yang=ikan  
 such=PST-have-HAB-IRR-NEG=ISG.A=much  
 ‘I would not much used to have had such’

Cases of analytic morphology are compound prepositions as we have seen one in (4b), but verbs as well show analytic structures not only with the progressive marker, but also with modals:

- (6) *Ming sahayyang dabas.*  
 Ming saha-oy-yang dabas  
 can come-NEG=ISG.A today  
 ‘I can’t come today.’

Most of the information the VP contains in this example is marked on the content verb, *saha-* ‘come’, except for ability, which is expressed by the particle *ming* ‘can’. *ming* is an uninflected form of the verb expressing ability and may be counted as an auxiliary verb in that the full semantic content of the VP is spread out over two verb forms, one major, one minor.<sup>3</sup> Consider also the following example in which *ming* is inflected like a regular verb:

- (7) *Da-mingya ang Diyan.*  
 Da-ming-ya ang Diyan.  
 so=can-3SG.M A=Diyan  
 ‘Diyan can (do it).’

## 3.2 Morphological processes

### 3.2.1 Prefixation

Prefixes in Ayeri apply mainly to verbs, but nouns, pronouns, adjectives and conjunctions can also appear with them, often in the form of proclitics. With verbs, ‘true’ prefixes—that is, morphemes that have been semantically bleached by grammaticalization to the point where they only express grammatical functions (Lehmann 2015: 157ff.) and which subcategorize words (Klavans 1985: 117)—are the tense prefixes marking both three degrees of past tense and three degrees of future tense, for example:

<sup>3</sup> *manga* has, in fact, a verbal counterpart *manga-* ‘move; remove’ as well, which served as the origin of both the progressive and the dynamic marker.

- (8) *Ang səsarāyn ya Makapetang.*  
 Ang sə-sara-ayn.Ø ya=Makapetang  
 AT FUT-go=IPL.TOP LOC=Makapetang  
 ‘We will go to Makapetang.’

Here, the prefix 𐄢: sə- marks future tense on the verb, 𐄢: sara- ‘go’. The other tense prefixes are 𐄢: kə- (NPST), 𐄢: mə- (PST), 𐄢: və- (RPST), and 𐄢: pa- (NFUT) and 𐄢: ni- (RFUT). Besides this set of prefixes, there are also a number of proclitics that can appear with verbs, though not exclusively. These are the anaphora 𐄢: da- ‘thus, so, such’ and the reflexive marker 𐄢: sitang- ‘self’:

- (9) *Da-mingya ang Diyan.*  
 Da-ming-ya ang Diyan.  
 so=can-3SG.M A=Diyan  
 ‘Diyan can (do it).’

- (10) *Sitang-kecāng.*  
 Sitang-ket-yāng  
 self=wash=3SG.M.A  
 ‘He washes *himself*.’

𐄢: sitang- can also be used as a preverb in situations where the agent is also the instrument, so both of the following two sentences are equivalent in meaning:

- (11) a. *Sa apicāng nanga ikan sitang-yari.*  
 Sa apit-yāng nanga ikan sitang-yari  
 PT clean=3SG.A house complete self=3SG.M.INS  
 ‘He cleaned the whole house by himself.’  
 b. *Sa sitang-apicāng nanga ikan.*  
 Sa sitang-apit-yāng nanga ikan  
 PT self=clean=3SG.A house complete  
 (idem)

Example (11a) shows the more common application of 𐄢: sitang-, that is, as a reflexive modifier of pronouns. The prefix 𐄢: da- can as well be used with noun phrases and is part of the demonstrative set of prefixes, 𐄢: da- ‘such’, 𐄢: eda- ‘this’, and 𐄢: ada- ‘that’:

- (12) *eda-ganang*  
 eda-gan-ang  
 this=child-A  
 ‘this child’

The demonstrative prefixes are also used to form the demonstrative pronouns *edanya* ‘this one’, *adanya* ‘that one’ and *danya* ‘such a’. A special case in this regard is the postposition *da-nārya* ‘in spite of, despite’ where *da-* combines with the conjunction *nārya* ‘but, although, except’. There is also a fixed adverbial expression using one of these prefixes, *eda-tadayyam* ‘for the time being, for now’ (this=time-DAT).

Last but not least, the prefix *ku-* ‘like, as though’ can be used with both adjectives and nouns (or, more precisely, phrases containing nominals):

- (13) a. *ku-koyaya*  
 ku-koya-ya  
 like=book-LOC  
 ‘like in a book’  
 b. *ku-prasi*  
 ku=prasi  
 ‘as though (it were) sour’

An example of a set-phrasal adverbial consisting of *ku-* and a verb is *ku-nasya* ‘as follows’, *nasya-* meaning ‘follow’. What is curious here is that this fossilized form is lacking person marking and is just extended with an epenthetic *-a* since *-sy* is not a permissible coda. The expected form would be *\*ku-nasyareng* (like=follow=3SG.INAN.A).

Following Klavans (1985), who suggests that clitics best be defined as “affixation at the phrasal level,” (117) a very common kind of clitical prefix to the verb *phrase* are the topic markers. They are counted as parts of the VP but do not interact with it regarding stress assignment (they are always unstressed) while always being in an initial position, preceding any other preverbal elements:

- (14) a. *Ang tabanya tamanley.*  
 Ang tahan-ya taman-ley  
 AT write-3SG.M letter-P.INAN  
 ‘He writes a letter.’  
 b. *Ang mətabanya tamanley.* ‘He wrote a letter.’



- c. *Ang manga mətabanya tamanley.* ‘He was writing a letter.’  
 d. *Ang manga no mətabanya tamanley.* ‘He was wanting to write a letter.’

### 3.2.2 Suffixation

As a largely agglutinative language, most grammatical marking in Ayeri is done by means of suffixes. These occur mainly with nouns and verbs, however, quantifiers take the shape of suffixes as well. Quantifiers, then, may modify content words almost regardless of their part of speech—noun, verb, adjective or adverb. The most pervasive examples of suffixation are certainly those of case marking on nouns and of person marking on verbs, for example:

- (15) *Sa pəharuyang va manga miday tangya vana suyareri!*  
 Sa pə-haru-yang va manga miday tang-ya vana suyar-eri  
 PT NFUT-beat=ISG.A 2SG.TOP DYN around ears-LOC 2SG.GEN ladle-INS  
 ‘I’ll beat you around your ears with a ladle!’

This example shows marking of 𐌂𐌆𐌇𐌸 *tang* ‘ears’ with the locative case suffix 𐌶𐌵𐌹𐌸 *-ya* and the marking of 𐌸𐌶𐌹𐌸 *suyar* ‘ladle’ with the instrumental case suffix 𐌸𐌶𐌹𐌸𐌸𐌶𐌹𐌸 *-eri*; the previous examples already provide instances of the exceedingly common markers for agent and patient case, 𐌸𐌶𐌹𐌸 *-ang* and 𐌸𐌶𐌹𐌸 *-as*, respectively. Besides case, nouns can also be marked for plural with the suffix 𐌶𐌵𐌹𐌸 *-ye*, and verb roots may be extended by the mood markers 𐌸𐌶𐌹𐌸 *-ong* (IRR), 𐌸𐌶𐌹𐌸 *-asa* (HAB) and 𐌸𐌶𐌹𐌸 *-oy* (NEG), the last of which should be the most frequently occurring. The mood suffixes can also be stacked, leading to the long word in (5) above. Person marking on verbs comes as agreement suffix or as a clitic personal pronoun depending on whether an agent NP proper is present or not for the verb to agree with; in (15), a cliticized agent pronoun 𐌸𐌶𐌹𐌸 *-yāng* ‘he’ (3SG.M.A) appears.

Quantifiers mostly appear as clitic suffixes on content words like in the following examples; note, however, that 𐌸𐌶𐌹𐌸𐌸𐌶𐌹𐌸 *-ikan* ‘much, many, very’ can also occur as a predicative adjective. 𐌸𐌶𐌹𐌸 *kagan* ‘too, overly’ always acts like an adverb.

- (16) a. *No sarayang-ikan*  
 No sara-yang-ikan  
 want go=ISG.A=much  
 ‘I really want to go.’

- b. *Tigalyeng para-ma.*  
 Tigal-yeng para-ma  
 swim=3SG.F.A fast=enough  
 ‘She swims fast enough.’
- c. *Yang valuy-eng, sabavāng.*  
 Yang valuy-eng, saha-vāng.  
 1SG.A glad=rather, come=2SG.A  
 ‘I am rather glad that you come.’
- d. *Adareng babisley mino-ing*  
 Ada-reng bahis-ley mino-ing  
 that=3SG.INAN.A day-P.INAN happy=so  
 ‘It was such a happy day.’
- e. *Ang konjan prikanley-ani*  
 Ang kond-yan.Ø prikan-ley-ani  
 A eat=3SG.M.TOP soup=not.at.all  
 ‘They did not eat any soup at all.’

### 3.2.3 Reduplication

...

### 3.2.4 Suprasegmental modification

...

### 3.2.5 Clitics

I have used the term ‘clitic’ above and claimed that various bound morphemes belong to this category. However, this term is notoriously vague and thus needs some specification and formalization as to why certain morphemes in Ayeri are probably best categorized as clitics rather than words or affixes. After the publication of Zwicky’s preliminary thoughts on the topic (Zwicky 1977), a number of articles were published that tried to provide a formal definition of what constitutes a clitic. Zwicky (1985) and Klavans (1985) provide guidelines in determining whether a given element that is suspected to be a clitic should better be categorized as a ‘particle’ word or as a suffix, or whether it is indeed an in-between thing, and thus a clitic. Table 3.1 lists a catalog of “symptoms” (Zwicky 1985: 285) that is supposed to serve

as a heuristic in determining the status of the various bound morphemes listed at the top and tries to categorize them by keeping score:

- o 'does not apply',
- o.5 'applies under certain circumstances', and
- 1 'applies'.

The table mostly contains '1' in the various fields, insofar most of the given criteria for cliticness are met. Thus, the few places in which 'o' or 'o.5' appear demand further discussion, as do the cases where the difference quotient indicates a narrow decision. A 'o' answer is also interesting as it means that the morpheme is behaving uncharacteristically of clitics for the respective criterion.

...

### 3.3 Marking strategies

...

Table 3.1: Clitichood tests according to Klavans (1985) and Zwicky (1985)

Symptom	topic markers + VP	dynamic marker + V'	V' + pronoun	da-	eda-/ada-	quantifiers	case marker + N'	ku- + NP	mo- + N'
Clitic vs. word (Zwicky 1985: 286-289)									
1.1 Word-internal sandhi applies	0	0	1	1	1	1	0	1	1
1.2 Word accent rules apply	0	0	1	0.5	0.5	0.5	0	0.5	0.5
1.3 Segmental phonology applies	–	–	–	–	–	–	–	–	–
2 Accent dependent	1	1	1	1	1	0	1	1	1
3.1 Bound morpheme	1	1	1	1	1	0	1	1	1
3.2 Closes off word	1	0	0	0	0	1	1	1	0
3.3 Modifies a single constituent	1	1	1	1	1	1	1	1	1
3.4 Strict ordering	1	1	1	1	1	1	1	1	1
3.5 Only with a certain constituent	1	0	1	0	0.5	0	1	1	1
3.6 No internal construction	1	1	0	1	0	1	1	1	1
4.1 Immunity to deletion	1	1	1	1	1	1	1	1	1
4.2 Immunity to replacement	1	1	1	1	1	1	1	1	1
4.3 Unavailable to movement	1	1	1	1	1	1	1	1	1
5 For $z = x + y$ , $y$ not part of syntactic rule on $z$	1	1	1	1	1	1	1	1	1
Clitic vs. affix (Klavans 1985)									
Semantic content	0	0	1	1	1	1	0	1	1
No semantic relationship to host	1	0	1	1	1	1	0	1	1
Subcategorization of phrasal node	1	1	1	1	1	1	1	1	1
Not very selective of host	1	0	0	1	1	1	0	1	0
Yes (clitic)	14	10	14	14.5	14	13.5	12	16.5	14.5
No (word or affix)	3	7	3	2.5	3	3.5	5	0.5	2.5
Difference quotient	0.65	0.18	0.65	0.71	0.65	0.58	0.41	0.94	0.71

## Bibliography

---

- Becker, Carsten. 2010. The syllable structure of Ayeri's stems: A statistical analysis. Draft. April 3. Accessed June 20, 2016. <http://benung.nfsbhost.com/files/frequency.pdf>.
- . 2012. Imperial messages: On translating Franz Kafka's "Eine kaiserliche Botschaft" into Ayeri. August 8. Accessed July 4, 2016. <http://benung.nfsbhost.com/wp-content/uploads/2012/04/2012-04-07-imperial-messages.pdf>.
- . 2013. Happy 10th anniversary, Ayeri! *Benung: The Ayeri Language Resource* (December 1). Accessed July 3, 2016. <http://benung.nfsbhost.com/archives/3817>.
- . 2016. Benung: The Ayeri language resource. Accessed June 9, 2016. <http://benung.nfsbhost.com>.
- Boga, Bella, Madita Breuninger, Thora Daneyko, and Martina Stama-Kirr. 2016. Ayeri. Universität Tübingen. June 30. Accessed July 4, 2016. <http://www.sfs.uni-tuebingen.de/-abuch/16ss/Referate/Ayeri.pdf>.
- Buch, Armin. 2016. Kunst- und Plansprachen – von Esperanto bis Dothraki. Accessed July 2. <http://www.sfs.uni-tuebingen.de/-abuch/16ss/conlang.html>.
- Court, Christopher. 1996. The spread of Brahmi scripts into southeast Asia. In Daniels and Bright 1996, 445–449.
- Daniels, Peter T., and William Bright, eds. 1996. *The world's writing systems*. New York: Oxford University Press.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Accessed July 12, 2016. <http://wals.info>.

- Eisenberg, Peter, Jörg Peters, Peter Gallmann, Cathrine Fabricius-Hansen, Damaris Nübling, Irmhild Barz, Thomas A. Fritz, Reinhard Fiehler, and Mathilde Hennig. 2016. *Die Grammatik: Unentbehrlich für richtiges Deutsch*. 9th ed. Edited by Angelika Wöllstein. Duden 4. Berlin: Dudenverlag.
- Everson, Michael. 2008. *Proposal for encoding the Javanese script in the UCS*. International Organization for Standardization, January 28. Accessed July 24, 2016. <http://www.evertype.com/standards/iso10646/pdf/n3319r2-javanese.pdf>.
- Halle, Morris. 1998. The stress of English words 1968–1998. *Linguistic Inquiry* 29 (4): 539–568.
- Kager, René. 2007. Feet and metrical stress. In *The Cambridge handbook of phonology*, edited by Paul de Lacy, 195–227. Cambridge: Cambridge University Press.
- Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61 (1): 95–120.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Edited by Alexander Lubotsky. Leiden Indo-European Etymological Dictionary Series 11. Leiden: Brill.
- Kuipers, Joel C., and Ray McDermott. 1996. Insular southeast Asian scripts. In Daniels and Bright 1996, 474–484.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*. 3rd ed. Classics in Linguistics 1. Berlin: Language Science Press. Accessed July 18, 2016. <http://langsci-press.org/catalog/book/88>.
- Maddieson, Ian. 2013a. Consonant inventories. In Dryer and Haspelmath 2013.
- . 2013b. Consonant-vowel ratio. In Dryer and Haspelmath 2013.
- . 2013c. Vowel quality inventories. In Dryer and Haspelmath 2013.
- Peterson, David J. 2015. *The art of language invention: From horse-lords to dark elves, the words behind world building*. New York: Penguin.
- Salomon, Richard G. 1996. Brahmi and Kharoshthi. In Daniels and Bright 1996, 373–383.
- United Nations. 1948. Universal declaration of human rights. United Nations. Accessed July 23, 2016. <http://www.un.org/en/universal-declaration-human-rights/>.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington, Ind.: Indiana University Linguistics Club. Accessed July 22, 2016. [https://web.stanford.edu/~zwicky/on\\_clitics.pdf](https://web.stanford.edu/~zwicky/on_clitics.pdf).

- . 1985. Clitics and particles. *Language* 61 (2): 283–305. Accessed July 26, 2016.  
<https://web.stanford.edu/~zwicky/cliticsparticles.pdf>.

# Index

---

- agglutination, 65
- allophony
  - of consonants, 9–12, 42
  - of vowels, 12–14, 21
- bound words, 66
- Box script, 47
- Brāhmī scripts, 47
- clitics, 65, 68
- complement clause, 44
- compounds, 40–42
- consonants, 9–12, 18, 21, 22, 24, 25, 28, 30, 49–50
  - clusters, 18, 25, 32–33
- dative, 11
- diacritics, 52–58
- diphthongs, 12, 21, 35, 54
- goals, 5, 6
- morphophonology
  - of relative pronouns, 11, 13
- nouns, 40
- numerals
  - digits, 59
- plural, 9, 11
- prefixes, 66, 68
- pronouns
  - demonstrative, 70
  - personal, 67
  - reflexive, 69
- punctuation, 59–61
- quantifiers, 40
- questions
  - closed, 43
  - open, 43
- reduplication, 72
- relative clause, 44
- stress, 33–40
- suffixes, 65, 71
- suprasegmental, 66, 72
- syllable
  - final, 25–28
  - heavy, 35–40
  - initial, 18–22
  - medial, 22–25
  - monosyllabic words, 28–32, 40
  - superheavy, 36, 39–40
  - types, 16, 35
- Tahano Hikamu, 5, 47–63
  - angular style (ᲛᲚᲗ), 61
  - blackletter style, 63
  - hand style (ᲛᲗᲚ), 62
  - regular/book style, 61
- tense, 69



- past, 12
- topic, 45
- typology
  - of morphemes, 65–73
- vowels, 12–14, 18, 22, 26, 30, 50–52