

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 10, 2016.

Set in Junicode and Fira Sans with Xe_{La}TeX.

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

List of Figures	ii
List of Tables	iii
Glossing Abbreviations	iv
o Introduction	3
i Phoneme Inventory and Phonotactics	7
1.1 Phoneme Inventory	7
1.1.1 Consonants	7
1.1.2 Vowels	9
1.2 Phonotactics	11
1.2.1 Number of Syllables per Word	12
1.2.2 Phonemic Makeup of Initial Syllables	15
1.2.3 Phonemic Makeup of Medial Syllables	19
1.2.4 Phonemic Makeup of Final Syllables	22
1.2.5 Phonemic Makeup of Single Syllables	25
Bibliography	29

List of Figures

o.1	First design for an Ayeri script	5
o.2	First draft for Tahano Hikamu	6
1.1	Consonant inventory	8
1.2	Vowel inventory	9
1.3	Syllabification of inflected verbs	12

List of Tables

I.1	Frequency of words with different numbers of syllables	12
I.2	Frequency of syllable types per word	14
I.3	Frequency of onsets in initial syllables	16
I.4	Frequency of nuclei in initial syllables	18
I.5	Frequency of codas in initial syllables	19
I.6	Frequency of onsets in medial syllables	20
I.7	Frequency of nuclei in medial syllables	21
I.8	Frequency of codas in medial syllables	22
I.9	Frequency of onsets in final syllables	23
I.10	Frequency of nuclei in final syllables	24
I.11	Frequency of codas in final syllables	25
I.12	Frequency of onsets in single syllables	26
I.13	Frequency of nuclei in single syllables	27
I.14	Frequency of codas in single syllables	27

Glossing Abbreviations

1	First person	NEG	Negative
2	Second person	NMLZ	Nominalizer
3	Third person	NN	Noun
A	Agent	NOM	Nominative
ACC	Accusative	NPST	Near past
AGTZ	Agentizer	P	Patient
AN	Animate	PL	Plural
AT	Agent topic	PST	Past
FUT	Future	PT	Patient topic
GEN	Genitive	PTCP	Participle
HAB	Habitative	REL	Relative
IMP	Imperative	RPST	Remote past
INAN	Inanimate	SBJ	Subject
INDF	Indefinite	SG	Singular
LOC	Locative	TOP	Topic
M	Masculine	VB	Verb
N	Neuter		

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).¹ This is to say that previous attempts of writing a full-fledged grammar of Ayeri have been incomplete due to 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^AT_EX, combined with using GitHub as a version control system, maintaining and editing will become faster, more transparent, and more elegant, since L^AT_EX operates on plain text files, and version control helps in keeping track of changes over time.

Carsten Becker
Marburg, July 4, 2016

¹  *Kutānas-ikan* ‘thanks a lot’ to Bella Boga, Madita Breuning, 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.¹ 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:

¹ A lot of the text here is taken from the blog article, “Happy 10th anniversary, Ayeri!” (Becker 2013).

- (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 layaiyāin mecoyalei ling *pinamea.*
 Ang laya-iy-a-in me-coya-lei ling *pinam-ea
 A.SBJ read-3SG.AN₁-a₁-SBJ INDF.INAN-book-P.INAN top.of bed-LOC
 ‘He reads a book on the bed.’

A word for ‘bed’—*pinam*—was only (re-)introduced on October 24, 2008. 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 familiar really 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 3rd-person referents: originally, every 3rd-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 are 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; Figure 0.1 shows a draft dated February 9, 2004. However, since 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” here 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.

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



[illegible]

Another important date in the history of Ayeri is 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 worked on 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.² This renewed attempt at documentation has been started with the intention to right those wrongs.

² 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 Ayer.

1 Phoneme Inventory and Phonotactics

This chapter will present charts depicting the phoneme inventory of Ayeri, give an analysis of the phonotactics of Ayeri's dictionary entries and also describe stress patterns.

1.1 Phoneme Inventory

1.1.1 Consonants

At 17 consonants, Ayeri has a fairly mid-sized inventory. Figure 1.1 shows the full chart. The sound /w/ only occurs marginally in ᐃᐱᐱᐱ *huākaya* ['wa:kaja] 'frog'. Other instances of it are allophones of /u/ followed by a vowel, for instance in ᐱᐱᐱ *rua-* /rwa/ 'have to, must'. /w/ may also be an allophone of /uj/, as in ᐱᐱᐱ *adauyi* [a'dawi] 'then', ᐱᐱᐱ *edaui* [e'dawi] 'now', or ᐱᐱᐱ *nekuyi* ['nekwi] 'eyebrows'. The negative suffix ᐱᐱ *-oy* is also commonly contracted to [w] before a diphthong:

- (4) ᐱᐱᐱᐱ *mingoyay* → *minguay* [mɪŋ'wai] 'I cannot' (can-NEG-1SG)

Moreover, the affricates /tʃ/ and /dʒ/ are usually allophones of /tj kj/ and /dj gj/, respectively, except if a homorganic nasal /n/ or /ŋ/ is preceding: for instance, ᐱᐱᐱ *ankyū* /'aŋ.kju/ 'really' is realized as [aŋ.kju], not as *[aŋ.tʃu]. The plural marker ᐱᐱ *-ye* is also commonly contracted to [dʒ] when a case suffix beginning with a vowel follows:

- (5) a. ᐱᐱᐱᐱᐱ *nyānyēang* → *nyānjang* ['nja:ndʒaŋ] 'persons' (person-PL-A)
b. ᐱᐱᐱᐱ *netuyēas* → *netujas* [ne'tudʒas] 'brothers' (brother-PL-P)

The plural marker also may contract before the locative marker ᐱᐱ *-ya*, basically for dissimilation:¹

¹ ᐱᐱ *-ea* also occurs as an variant morpheme, so that ᐱᐱ *-ye* + ᐱᐱ *-ea* → ᐱᐱ *-yēa*.

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		(w)		l j ⟨y⟩		

- (6) $\text{နီယဲယ} \text{ nivayeya} \rightarrow \text{nivajya}$ [ni'vadʒja] ‘at the eyes’ (eye-PL-LOC)

Dissimilation of the sequence $\text{ယယ} -yaya$ is attested in Becker (2012: 12), where the relative pronoun $\text{နီယယ} \text{ siyaya}$ appears transcribed as *siyya*:

As far as morphophonology is concerned, the relative pronoun complex *siyya* ‘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.

It is also noted there, however, that the contraction happens “only if both parts are grammatical suffixes” (12).

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

- (7) a. $\text{တပ်သံ} \text{ tavvāṅ}$ [ta'va:ŋ] ‘you get’ (get=2SG.A),
b. $\text{သိပ်သံ} \text{ disyyang}$ [di'sjaŋ] ‘I fasten’ (fasten=1SG.A).

With diphthongs, the sequence $/V_1j/$ is treated as though it were $/Vj.j/$, however, the vowel remains lax in spite of being phonetically in an open position:

- (8) $\text{နီပုယ} \text{ tipuyya}$ [ti'pu.ja] ‘on the grass’ (grass-LOC).

1.1.2 Vowels

Ayeri has a very basic five-to-six-vowel system, shown in Figure 1.2.

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⟩	

The lax vowels $[ɪ \text{ } \epsilon \text{ } ʊ]$ occur as allophones of their tense counterparts $[i \text{ } e \text{ } u]$ in closed syllables, for example:

- (9) a. $\text{မိန့်} \text{ ming}$ [mɪŋ] ‘can, be able’,
b. $\text{မင်း} \text{ enya}$ [ɛnja] ‘everyone’,
c. $\text{မိခင်} \text{ agon}$ [aɡɔn] ‘outer, foreign’, and
d. $\text{ပက်} \text{ pakur}$ [pəkur] ‘ill, sick’.

/ə/ is a marginal phoneme and only occurs in the tense prefixes 𑄢: *kə-* ‘NPST’, 𑄣: *mə-* ‘PST’, 𑄤: *və-* ‘RPST’, as well as in the prefix 𑄥: *mə-* ‘some, whichever’. Otherwise, [ə] occurs as an allophone of /e/ in final unstressed position, e.g. in the word 𑄢𑄣 *mine* [‘minə] ‘affair, matter, issue’.

Ayeri also possesses a number of diphthongs, these are: /aɪ a:ɪ eɪ ɔɪ uɪ au/, spelled ⟨ay⟩, ⟨āy⟩, ⟨ey⟩, ⟨oy⟩, ⟨uy⟩, ⟨au⟩. Furthermore, the vowels [i e a o u] may be long: [i: e: a: o: u:]. Long vowels are lexicalized in a few words, for example:

- (10) a. 𑄢𑄣 *nīsa* ‘wanted’, 𑄢𑄣𑄢 *pasīsa* ‘interesting’;
 b. 𑄢𑄣𑄢 *arēn* ‘anyway, however’, 𑄢𑄣𑄢 *lēra* ‘whore’;
 c. 𑄢𑄣 *lā* ‘tongue’, 𑄢𑄣𑄢 *yāng* ‘he’ (he.A);
 d. 𑄢𑄣 *nōn* ‘wish’; and
 e. 𑄢𑄣𑄢 *babūan* ‘barbarian’.²

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

- (11) 𑄢𑄣𑄢 *aja-* ‘play’ + 𑄢𑄣𑄢 *-an* ‘NMLZ’ → 𑄢𑄣𑄢𑄢 *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, 𑄢𑄣𑄢 *sinā* ‘of which, about which’, as in the following example:

- (12) *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=1SG.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 𑄢𑄣𑄢 *sina* ‘which.GEN’:³

² 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, 𑄢𑄣𑄢 *babūan* ‘barbarian’ and its adjective 𑄢𑄣𑄢 *babū* ‘barbarian (adj.)’ were coined as 𑄢𑄣𑄢𑄢 *prankayē*—things ‘that you put in specifically to make things fit’, another new coining this decision resulted in.

³ A variant which combines the allomorphs of the relativizer and the genitive case marker in the opposite way also exists: 𑄢: *s-* + 𑄢𑄣𑄢 *-ena* → 𑄢𑄣𑄢 *sena*.

- (13) *tamanang ledanena nā sina koronvāng*
 taman-ang ledan-ena nā si-na koron-vāng
 letter-A friend-GEN ISG.GEN REL-GEN know=2SG.A
 ‘the letter of my friend which you know’

As pointed out in (10c), the word ႏ ၵ *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 hiat, a glide /j/ may be inserted, so both of these are possible:

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

With diphthongs, /ɪ/ coalesces with a following /j/ to /j/, but the initial vowel will not become tense, hence:

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

1.2 Phonotactics

For the purpose of this statistical analysis, all of the available translations into Ayerl since 2008 have been used as a text corpus;⁴ 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

⁴ 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).

or suffixes.⁵ Borrowings have been deleted, if they could not 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 translation 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 %

⁵ 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.

Two-syllable words make up the bulk of the sample, which is not surprising since 1,072 (55.43%) of the dictionary subsample are bisyllabic words. Most of Ayeri's roots are bisyllabic; unsurprisingly, most monosyllabic words are function words like the ones cited below. A few examples for each number of syllables per word:

- (16) a. ယှေ *yeng* 'she' (she.A),
 ဂေ *le* (PT),
 ရှိ *rua* 'must';
- b. မိ *datau* 'normal',
 မာရေ *mareng* 'it suffices' (suffice=3SG.INAN.A),
 နာယ *nasay* 'near to';
- c. မာသေ *avanyāng* 'he sinks' (sink=3SG.M.A),
 နာရေ *narānye* 'words' (word-PL),
 တာလေ *tovalay* 'a cloak' (cloak-P.INAN);
- d. မိသေ *binyanveno* (corner.beautiful, a place name),
 မာသေ *mangasatong* 'they used to move' (move-HAB=3PL.N.A),
 မိသေ *mitanena* 'of the palace' (palace-GEN);
- e. မာသေ *haruyamanas* 'beatings' (beat-PTCP-NMLZ-P),
 မိသေ *sirutayena* 'of the night' (night-GEN),
 မိသေ *sungkorankihas* 'geography' (science.map);
- f. မိသေ *kaytomayanena* 'of righteousness' (righteous-NMLZ-GEN),
 မိသေ *koronaryasayan* 'they used to forget' (forget-HAB-3PL.M),
 မိသေ *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 performed 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:

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

Type	Initial		Medial		Final		Single		Total	
CV	2898	67.41%	1974	72.02%	2108	49.03%	578	48.13%	7558	60.27%
CCV	55	1.28%	24	0.88%	47	1.09%	32	2.66%	158	1.26%
CCCV	—	—	—	—	2	0.05%	—	—	2	0.02%
CVC	759	17.66%	610	22.25%	1901	44.22%	298	24.81%	3568	28.45%
CCVC	29	0.67%	10	0.36%	86	2.00%	9	0.75%	134	1.07%
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%

- (17) a. Initial CVCC:
 လှံတံ linktang /lɪŋk.'təŋ/ 'they try' (try=3PL.M.A),⁶
 လှံတံ silvnang /silv.'nəŋ/ 'I see' (see=IPL.A);
- b. Final CCCV:
 မိဂရို migryo /'mi.grjo/ 'flourishes' (flourish-3SG.N),
 ဆဲရို subryo /'su.brjo/ 'ceases' (cease-3SG.N);
- c. Single V:
 အေ ay /aɪ/ 'I' (1SG.TOP).

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 လှံ 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:⁷

⁶ In the dictionary as လှံ linka-, possibly an entry changed at a later point, or simply an error in the text (Sugar Fairies).

⁷ However, `sort` was unable to handle all IPA characters, so `sed 'y/εɪɔʊə:ʃɜŋ/εɪɔʊ@: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 %



- (18) C = (?:tʃ|dʒ|[ptkbgmɲɳvʃhrljw])
 V = (?:[aɛɔʊ]:ɪ|[aɛɔʊ]ɪ|aʊ|[ieaou]:|[ieaou]ɛɔʊə))

As we have seen above (Table 1.2), CCV syllables only make up 1.28% of initial syllables, in so far 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.

Perhaps most striking about the nuclei of initial syllables presented in Table 1.4 is that it is plain vowels which occur most of the time. As mentioned above, lax vowels are counted here as allophones of tense ones as their distribution is complementary, which is why the plain vowels are presented as grouped. Long vowels and diphthongs find themselves below the 5% threshold, and the words with single occurrences are:

- (19) a.  *kuysān* 'comparison',
 b.  *nōn* 'will, intention',
 c.  *nīsa* 'wanted',⁸
 d.  *sēyraya* 'will overcome' (FUT-overcome-3SG.M),
 e.  *sautan* 'cork'.

As [e:ɪ] 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  *-an*, as exemplified in (11) above. Similarly, the 18 cases of /ə/ reported here are mostly from tense prefixes also mentioned above, for instance,  *məkoroṇay* '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  *linktang* 'they try' (try=3PL.M.A), and /lv/ in the word  *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/, and /k/, the latter two only once, /h/ twice:




⁸  *nīsa* and  *nōn* are both related to  *no*- 'want, plan'.

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 %

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 %
h	2	0.05 %
t	1	0.02 %
ŋk	1	0.02 %
lv	1	0.02 %
k	1	0.02 %

- (20) a. မေ့သွားကံ *mehvāng* ‘you are supposed to’ (be.supposed.to=2SG.A),⁹
 ရိသေ့တော *rohtang* ‘they bite’ (bite=3SG.M.A);
- b. ကညာက *patlay* ‘cousin’;
- c. ခိတ်ခိတ် *sik-sik* ‘tits’.

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 for which 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/). Regarding 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

⁹ The dictionary entry is *မှ* *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%

more common here, and /o/ and /u/ switch places. Instead of /e:ɪ/, there is an attestation of /u:/ (see footnote 2), for which there is more reason to be counted as a phoneme than for /e:ɪ/. The sequences /i:/ and /ʊɪ/ also only occur once and twice, respectively, namely in the following words:

- (21) a. ၂ုၼ် *pasīsa* ‘interesting’;
 b. ၂ုၼ် *puluyley* ‘a mirror’ (mirror-P.INAN),
 ၂ုၼ် *tipuyya* ‘on the grass’ (grass-LOC).

The word in (21a), ၂ုၼ် *pasīsa* ‘interesting’, should count as a lexeme in its own right, since it possesses idiomatic meaning. Nonetheless, it rather transparently constitutes a causative derivation of the verb ၂ုၼ် *pasy-* ‘wonder, be curious, be in-

terested’, essentially meaning ‘making one wonder/curious’—the causative suffix :ṣṣ -*isa* can as well be used to derive adjectives with a causative or resultative meaning.

Table 1.8: Frequency of codas in medial syllables (n = 2741)

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 %

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

- (22) a. ကပ်တင် *pangitlan* ‘money change’ and
 b. ṣṣṣṣ *telugtong* ‘they survive’ (survive=3PL.N).¹⁰

As charted in Table 1.2 above, Ayeri very strongly favors CV syllables in medial positions, hence the high count of zero segments here.

1.2.4 Phonemic Makeup of Final Syllables

The onsets of final syllables of polysyllabic words (Table 1.9) show the greatest amount of variety, which is due to Ayeri mostly using suffixes for grammatical purposes. Hence it is no surprise that combinations with /j/ and, indeed, /j/ itself as an onset, are especially common, since /j/ is also what a number of very common suffixes start with, for example the plural marker :ṣṣ -*ye*, the locative marker :ṣṣ -*ya*,

¹⁰ The word for ‘money’ is ကပ် *pangis*, so (22a) is probably a compound, albeit not a fully transparent one. The word for ‘change’ is တင် *tila-*, and there seems to be a nominalizing :ṣṣ -*an*. Ayeri 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 a 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 တင် *tilān* underwent metathesis to *တင် **itlān* to match the rhyme of ကပ် *pangis*. *ကပ် **pangisitlān* then underwent irregular haplology (and shortening of the nominalizing suffix) to ကပ်တင် *pangitlan*.

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

Phoneme	Frequency	Percentage	Phoneme	Frequency	Percentage
Ø	155	3.61%	kj	6	0.14%
j	1101	25.61%	hj	5	0.12%
n	528	12.28%	bj	5	0.12%
r	398	9.26%	tw	4	0.09%
t	268	6.23%	sw	4	0.09%
s	244	5.68%	sj	4	0.09%
l	238	5.54%	kw	3	0.07%
k	199	4.63%	kr	3	0.07%
d	184	4.28%	br	3	0.07%
m	154	3.58%	vr	2	0.05%
v	142	3.30%	rw	2	0.05%
h	128	2.98%	nw	2	0.05%
p	115	2.68%	tv	1	0.02%
g	103	2.40%	tfv	1	0.02%
dʒ	73	1.70%	tfj	1	0.02%
b	73	1.70%	rj	1	0.02%
tf	52	1.21%	nj	1	0.02%
vj	26	0.60%	mw	1	0.02%
pj	22	0.51%	grj	1	0.02%
dʒj	17	0.40%	dv	1	0.02%
tr	10	0.23%	dr	1	0.02%
w	9	0.21%	brj	1	0.02%
pr	7	0.16%			

the dative and participle marker :uɛ -*yam*, as well as third-person animate pronoun agreement suffixes, and the various first-person and third-person animate pronominal clitics. Figure 1.3 above shows exemplarily how verbs resyllabify when suffixes are attached. Even though single-segment onsets are strongly preferred, Cr, Cw, and especially C(C)j seem to be generally permissible.¹¹

¹¹ 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 ꞑ either way, and there is no heeding morpheme breaks either. /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%
i	147	3.42%
i	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%

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 𑖅𑖫 *-an*, which lengthens the vowel of verb stems ending in /a/, as demonstrated in (11). /aɪ/ 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 rule applies here as well, so its status as a phoneme is marginal. All instances of /e:/ in the sample are from the word 𑖅𑖫𑖫𑖫 *arēn* ‘anyway, however’; all evidence for /i:/ is from 𑖅𑖫 *sirī* ‘due to which’ (see section 1.1.2). The only evidence for /u:/ in the sample is from 𑖅𑖫 *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 𑖅𑖫𑖫𑖫 *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.¹²

1.2.5 Phonemic Makeup of Single Syllables

...

...

...

¹² Kroonen (2013: 489–490) identifies PGmc **sūgan-*, **sūkan-* ‘to suck’ as an iterative of PGmc **sukkōn-*, **sugōn* ‘to suck’ and reconstructs PIE **souk-neh₂-*. 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 %

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%

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>.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Edited by Alexander Lubotsky. Leiden Indo-European Etymological Dictionary Series 11. Leiden and Boston: Brill.