A Grammar of Ayeri

A Grammar of Ayeri

Documenting a Fictional Language

by Carsten Becker

Benung. The Ayeri Language Resource

Last edited: July 15, 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

List of Figures

List of Tables

Glossing Abbreviations

1	First person
2	Second person
3	Third person
A	Agent
ACC	Accusative
AGTZ	Agentizer
AN	Animate
AT	Agent topic
CAUS	Causative
DAT	Dative
F	Feminine
FUT	Future
GEN	Genitive
IMP	Imperative
INAN	Inanimate
INDF	Indefinite
INF	Infinitive
INS	Instrumental
LOC	Locative
M	Masculine
N	Neuter
NEG	Negative
NMLZ	Nominalizer
NN	Noun
NOM	Nominative
NPST	Near past
P	Patient
PL	Plural
PST	Past
PT	Patient topic
PTCP	Participle
REL	Relative
RPST	Remote past
SBJ	Subject
SG	Singular
SUPL	Superlative
TOP	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when the first thing people look for on my website is often the grammar, incomplete as well as partially inaccurate and outdated as it may be.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).[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 LaTeX, combined with using GitHub as a version control system, maintaining and editing will become faster, more transparent, and more elegant, since LaTeX operates on plain text files, and version control helps in keeping track of changes over time.

Carsten Becker
Marburg, July 4, 2016

[1] ཀཱུཏཱནཱསཿཨིཀཱན *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.

0 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] Most of the text here is taken from the blog article, "Happy 10th anniversary, Ayeri!" (Becker 2013).

(1)

Ayevhoi	*agiaemaesim*	*coyaieliеðamavir*	*vhaieloyaŋaiye.*
Ay-evhoi	agia-ema-esim	coyai-el-i-eðam-avir	vhai-el-o-yaŋa-iye
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	*layaiyạin*	*mecoyalei*	*ling*	**pinamea.*
Ang	laya-iy-a-in	me-coya-lei	ling	*pinam-ea
A.SBJ	read-3SG.AN$_1$-a$_1$-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, aligment. 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)

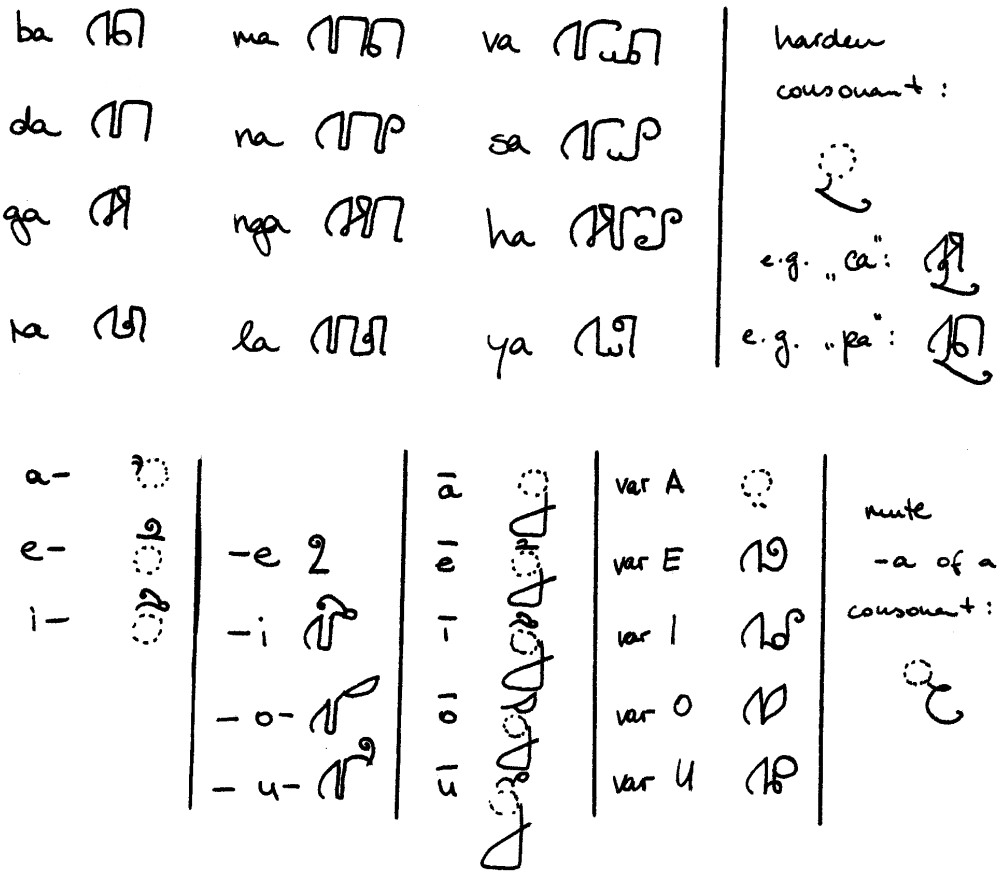
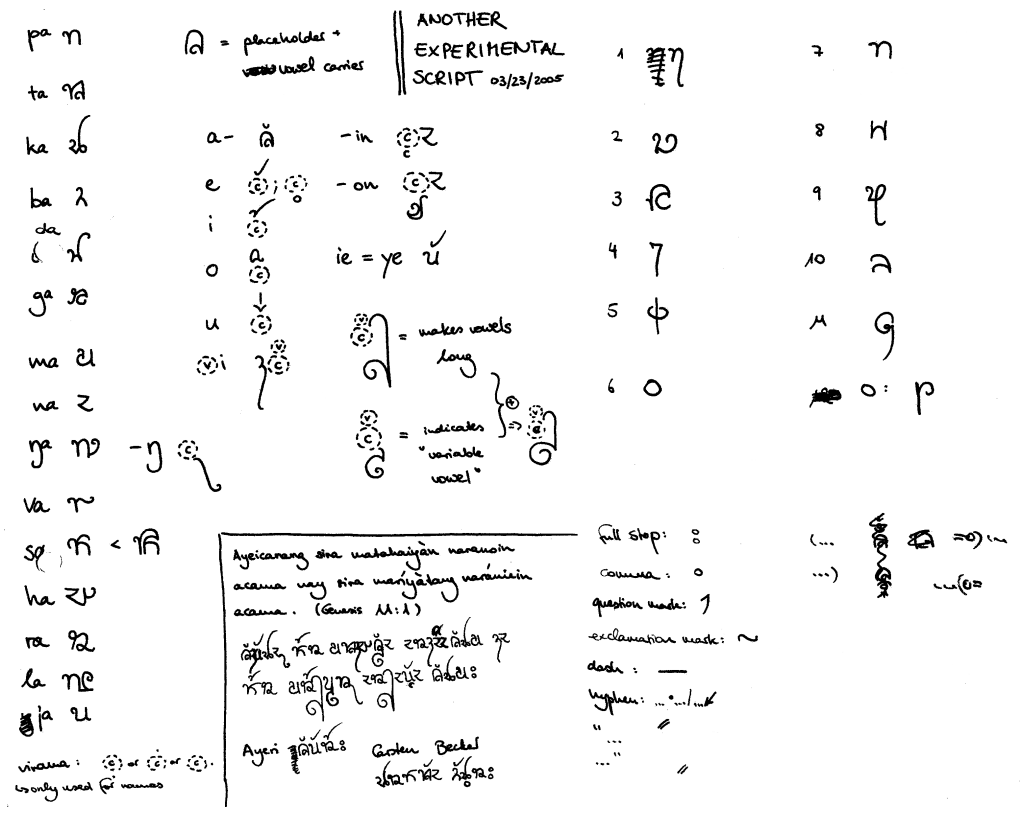


Figure 0.2: First draft for Tahano Hikamu (March 23, 2005)



Another 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 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 form 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.

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

[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 Ayeri.

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."

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 be as perfect as miniature models of things can ever 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 languages will normally entail. The ultimate goal in my work on Ayeri is 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 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 Ayeri an interesting piece of "informed nonsense," as a colleague of mine once put it. 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.[3] Criticism is always welcome as long as it is constructive.

[3] 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], respecitively, except if a homorganic nasal /n/ or /ŋ/ is preceding: for instance, *ankyu* /ˈaŋkju/ 'really' is realized as [ˈaŋkju], not as *[ˈaŋtʃu] or *[ˈantʃu]. 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 instance, there is no way to predict whether *cuna* 'original, initial', *panca* 'finally, eventually', and *vac-* 'like', or *jaran* 'pilgrimage', *aja-* 'play', and *nuj-* 'pour' have /tj/ or /kj/, /dj/ or /gj/, respectively, unless we consider the clues given by the conservative native spellings of the respective words.[1] 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.

[1] 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 *-ye* is commonly contracted to [dʒ] when a case suffix beginning with a vowel follows:[2]

(4) a. *nyān**ye**ang* → *nyān**j**ang* [ˈnjaːndʒaŋ] 'persons' (person-PL-A);

b. *netu**ye**as* → *netu**j**as* [neˈtudʒas] 'brothers' (brother-PL-P).

The plural marker may also contract before the locative marker *-ya* and the dative marker *-yam*, basically for dissimilation:[3]

(5) a. *niva**ye**ya* → *niva**j**ya* [niˈvadʒja] 'at the eyes' (eye-PL-LOC);

b. *mavi**ye**yam* → *mavi**j**yam* [maˈvidʒjam] 'to the sheep' (sheep-PL-DAT).

Dissimilation of the sequence *-yaya* is attested in my translation of Kafka's short story "Eine kaiserliche Botschaft," where the relative pronoun *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.

Lastly, /h/ may be realized as [ç] before front vowels, and as [x] before back vowels:

(6) a. *ta**h**i* [ˈtaçi] 'favorable';

b. *ba**h**o* [ˈbaxo] 'loud'.

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

(7) a. *tavvāng* [taˈvaːŋ] 'you get' (get=2SG.A),

b. *disyyang* [diˈsjaŋ] 'I fasten' (fasten=1SG.A).

With diphthongs, the sequence /Vɪ.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:

(8) *tipu**yy**a* [tiˈpʊ.ja] 'on the grass' (grass-LOC).

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

[3] *-ea* also occurs as an variant morpheme, so that *-ye* + *-ea* → *-yēa*.

1.1.2 Vowels

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

	Front	Center	Back
High	i, iː ⟨ī⟩		u, uː ⟨ū⟩
Mid	e, eː ⟨ē⟩	(ə ⟨ə, e⟩)	o, oː ⟨ō⟩
Back		a, aː ⟨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. The consonant–vowel ratio is 4.25, which Maddieson (2013b) also 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:

(9) a. *m**i**ng* [mɪŋ] 'can, be able',
 b. ***e**nya* [ˈɛnja] 'everyone',
 c. *ag**o**n* [ˈagɔn] 'outer, foreign', and
 d. *pak**u**r* [ˈpakʊr] 'ill, sick'.

/ə/ is a marginal phoneme and only occurs in the tense prefixes *kə-* 'NPST', *mə-* 'PST', *və-* 'RPST', as well as in the prefix *mə-* 'some, whichever'. Otherwise, [ə] occurs as an allophone of /e/ in final unstressed position, for instance, in the word *min**e*** [ˈminə] 'affair, matter, issue'.

Ayeri also possesses a number of diphthongs, these are: /aɪ eɪ ɔɪ ʊɪ aʊ/, spelled ⟨ay⟩, ⟨ey⟩, ⟨oy⟩, ⟨uy⟩, and ⟨au⟩. Furthermore, there are long equivalents of the short vowels: /iː eː aː oː uː/. Long vowels are lexicalized in a few words, for example:

(10) 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'.[4]

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

(11) ᜀᜃᜒ *aja-* 'play' + ᜐᜀᜈ᜔ *-an* 'NMLZ' → ᜀᜃᜒᜐᜀᜈ᜔ *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, ᜐᜒᜈᜏ *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 ᜐᜓᜈ *sina* 'which.GEN':[5]

(13)

tamanreng	*ledanena*	*nā*	*sina*	*koronvāng*
taman-reng	ledan-ena	nā	si-na	koron-vāng
letter-A.INAN	friend-GEN	1SG.GEN	REL-GEN	know=2SG.A

'the letter of my friend which you know'

As pointed out in (10c), the word ᜎᜃ᜔ *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)

[4] 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, ᜊᜓᜊᜓᜀᜈ᜔ *babūan* 'barbarian' and its adjective ᜊᜓᜊᜓ᜔ *babū* 'barbarian (adj.)' were coined as ᜉᜇᜈ᜔ᜑᜌ᜔ *prankayé*—things 'that you put in specifically to make things fit', another new coining this decision resulted in.

[5] A variant which combines the allomorphs of the relativizer and the genitive case marker in the opposite way also exists: ᜐᜒ *s-* + ᜈᜏ᜔ *-ena* → ᜐᜓᜈ *sena*.

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

(15) *tip**uyy**a* [ti'pʊ.ja] 'on the grass' (grass-LOC).

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

(16) *ming**oy**ay* → *ming**u**ay* [mɪŋ'waɪ] 'I cannot' (can-NEG=1SG.TOP).

1.2 Phonotactics

For the purpose of this statistical analysis, all of the available translations into Ayeri from late 2008 to July 2016 have been used as a text corpus;[6] 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.[7] 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:

6 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).

7 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.

Figure 1.3: Syllabification of inflected verbs

Suffix	*ca-* 'love'	*gum-* 'work'	*babr-* 'mumble'
-ay (1SG)	cā́y	gu.máy	ba.bráy
-va (2SG)	cá.va	gúm.va	ba.brá.va
-yam (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%

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. In the following, I will quote a few examples for each number of syllables per word:

(17) a. *yeng* 'she' (she.A),
rua 'must';

b. *datau* 'normal',
nasay 'near to';

c. *avanyāng* 'he sinks' (sink=3SG.M.A),
tovaley 'a cloak' (cloak-P.INAN);

d. *hinyanveno* (corner.beautiful, a place name),
mitanena 'of the palace' (palace-GEN);

e. *haruyamanas* 'a beating' (beat-PTCP-NMLZ-P),
sungkorankihas 'geography' (science.map);

f. *kaytomayanena* 'of righteousness' (righteous-NMLZ-GEN),
nasimayajang-hen '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:

(18) a. Initial CVCC:
linktang /lɪŋk.ˈtaŋ/ 'they try' (try=3PL.M.A),[8]
silvnang /sɪlv.ˈnaŋ/ 'we see' (see=1PL.A);

b. Final CCCV:
migryo /ˈmi.grjo/ 'flourishes' (flourish-3SG.N),
subryo /ˈsu.brjo/ 'ceases' (cease-3SG.N);

c. Single V:
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 *ang* 'AT', which alone appears 255

[8] The verb stem is found in the dictionary as *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.

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%

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:[9]

(19)
```
C = (?:tʃ|dʒ|[ptkbdgmnŋvshrljw])
V = (?:[ae]:?ɪ|aʊ|[ieaou]:?|[ɪɛɔʊə])
```

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:

[9] However, `sort` was unable to handle all IPA characters, so `sed 'y/ɛɪɔʊəːʃʒŋ/EIOU@: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	*802*	*18.66%*
	ɪ	*209*	*4.86%*
e		705	16.40%
	e	*523*	*12.17%*
	ɛ	*164*	*3.81%*
	ə	*18*	*0.42%*
u		260	6.05%
	u	*228*	*5.30%*
	ʊ	*32*	*0.74%*
o		227	5.28%
	o	*188*	*4.37%*
	ɔ	*39*	*0.91%*
aː		109	2.54%
aɪ		88	2.05%
eɪ		40	0.93%
eː		4	0.09%
ɔɪ		3	0.07%
ʊɪ		1	0.02%
oː		1	0.02%
iː		1	0.02%
eːɪ		1	0.02%
aʊ		1	0.02%

(20) a. *kuysān* 'comparison',
b. *nōn* 'will, intention',
c. *nīsa* 'wanted',[10]
d. *sēyraya* 'will overcome' (FUT-overcome-3SG.M),
e. *sautan* 'cork'.

As [eːɪ] 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 [aː] 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, *məkoronay* 'I knew' (PST-know=1SG.TOP).

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%

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/, /tʃ/ and /k/, the latter two only once, /h/ twice:

[10] *nīsa* and *nōn* are both related to *no-* 'want, plan'.

(21) a. *mehvāng* 'you are supposed to' (be.supposed.to=2SG.A),[11]
rohtang 'they bite' (bite=3SG.M.A);

b. *mutva* 'you rub' (rub=2SG.TOP),
patlay 'cousin';

c. *sik-sik* 'tits';

d. *vacvāng* 'you like' (like=2SG.A).

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 different 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 more common here, and /o/ and /u/ switch places. Instead of /eːɪ/, there is an attestation of /uː/ (see footnote 4), for which there is more reason to be counted as a phoneme than for /eːɪ/. The sequences /iː/ and /ʊɪ/ also only occur once and twice, respectively, namely in the following words:

(22) a. *pasīsa* 'interesting';

b. *puluyley* 'a mirror' (mirror-P.INAN),
tipuyya 'on the grass' (grass-LOC).

The word in (22a), *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 *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.

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

[11] The dictionary entry for the verb is *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	*387*	*14.12 %*
	ɪ	*93*	*3.39 %*
e		254	9.26 %
	e	*206*	*7.52 %*
	ɛ	*48*	*1.75 %*
o		194	7.08 %
	o	*119*	*4.34 %*
	ɔ	*75*	*2.74 %*
u		120	4.38 %
	u	*101*	*3.68 %*
	ʊ	*19*	*0.69 %*
aː		110	4.01 %
aɪ		51	1.86 %
ɔɪ		33	1.20 %
eɪ		5	0.18 %
eː		5	0.18 %
aʊ		5	0.18 %
ʊɪ		2	0.07 %
uː		1	0.04 %
iː		1	0.04 %

(23) a. *pangitlan* 'money change' and

b. *telugtong* 'they survive' (survive=3PL.N).[12]

12 The word for 'money' is *pangis*, so (23a) is probably a compound, albeit not a fully transparent one. The word for 'change' is *tila-*, and there seems to be a nominalizing *-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 *tilān* underwent metathesis to **itlān* to match the rhyme of *pangis*. **pangisitlān* then underwent irregular haplology (and shortening of the nominalizing suffix) to *pangitlan*.

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%

As documented 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 ːᥡ *-ye*, the locative marker ːᥙ *-ya*, the dative and participle marker ːᥙᥒ *-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.[13]

Nuclei of final syllables (Table 1.10) do not bear striking differences to nuclei in other positions. /aː/ comes out second here due to the common nominalizer ːᥘᥲ *-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, /aːɪ/ occurs a number of times—the

[13] 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 ᥝ either way, and there is no heeding morpheme breaks either. /CsjV/ will be /C.sjV/ in any case, since Ayeri avoids final consonant clusters if possible, see Table 1.2.

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

Phoneme	Frequency	Percentage
Ø	155	3.61%
j	1101	25.61%
n	528	12.28%
r	398	9.26%
t	268	6.23%
s	244	5.68%
l	238	5.54%
k	199	4.63%
d	184	4.28%
m	154	3.58%
v	144	3.35%
h	128	2.98%
p	115	2.68%
g	103	2.40%
dʒ	73	1.70%
b	73	1.70%
tʃ	52	1.21%
vj	26	0.60%
pj	22	0.51%
dʒj	17	0.40%
tr	10	0.23%
w	9	0.21%

Phoneme	Frequency	Percentage
pr	7	0.16%
kj	6	0.14%
hj	5	0.12%
bj	5	0.12%
tw	4	0.09%
sw	4	0.09%
sj	4	0.09%
kw	3	0.07%
kr	3	0.07%
br	3	0.07%
vr	2	0.05%
rw	2	0.05%
nw	2	0.05%
tʃj	1	0.02%
rj	1	0.02%
nj	1	0.02%
mw	1	0.02%
grj	1	0.02%
dv	1	0.02%
dr	1	0.02%
brj	1	0.02%

vowel-lengthening rule applies here as well, so its status as a phoneme is marginal. All instances of /eː/ in the sample are from the word 𐑨𐑮𐑧́𐑯 *arēn* ‘anyway, however’; all evidence for /iː/ is from 𐑕𐑦𐑮𐑰 *sirī* ‘due to which’ (see section 1.1.2). The only evidence for /uː/ in the sample is from 𐑚𐑩𐑚𐑵 *babū* ‘barbarian (adj.)’.

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 𐑕𐑦𐑒·𐑕𐑦𐑒 *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.[14]

[14] Kroonen (2013: 489–490) identifies PGmc **sūgan-*, **sūkan-* ‘to suck’ as an iterative of PGmc **sukkōn-*, **sugōn-* ‘to suck’ and reconstructs PIE **souḱ-neh₂-*. However, he does not say anything about the Germanic word being onomatopoetic in origin.

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

Phoneme		Frequency	Percentage
a		2408	56.01%
aː		316	7.35%
o		411	9.56%
	o	*298*	*6.93%*
	ɔ	*113*	*2.63%*
i		289	6.42%
	ɪ	*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%*
	e	*1*	*0.02%*
eɪ		103	2.40%
ɔɪ		42	0.98%
aːɪ		23	0.54%
ʊɪ		14	0.33%
aʊ		14	0.33%
eː		5	0.12%
iː		3	0.07%
uː		1	0.02%

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%

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 𐑡 *mya* 'be supposed to', which is very commonly used as an indeclinable modal particle.

A consonant onset that can only be found in monosyllables is /ŋ/,[15] in 𐑓𐑦𐑕 *-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:

(24) *Apayeng-ngas.*
Apa-yeng-ngas
laugh=3SG.F.A=almost

'She almost laughed.'

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

[15] At least according to the analysis chosen here, see section 1.2.1 for an explanation.

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%

(25) *Apayeng* *baho.*
Apa-yeng baho
laugh=3SG.F.A loudly
'She laughs loudly.'

However, whereas ꦧꦲꦺꦴ *baho* 'loud' is treated as a separate unit in terms of intonation, ꦔꦏ꦳ *-ngas* is unstressed and binds to what it follows:

(26) a. ꦄꦥꦪꦺꦁꦔꦏ꦳꧉ *Apayeng-ngas.* [ˌapaˈjɛŋas];
b. ꦄꦥꦪꦺꦁ ꦧꦲꦺꦴ꧉ *Apayeng baho.* [ˌapaˈjɛŋ ˈbaxo].

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

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

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 /aʊ/, and the marginally phonemic /eː/ is not present either. By having /a/, /aɪ/, /aː/ 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 /oː/ features twice in this group, namely in two instances of the word *nōn* 'will, intention'.[16]

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%

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

1.2.6 Cross-Syllable Consonant Clusters

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 (74.11%), alveolars with alveolars (33.44%), and velars with velars (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 25.35%. Alveolars provide the highest

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

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, ŋ.dʒj, ŋ.kw, ŋ.m, ŋ.n, ŋ.rj, ŋ.t, ŋk.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, ŋ.b, ŋ.h, ŋ.p, ŋ.w, r.dʒj, r.pj, s.dʒj, s.m, t.l (0.16%); l.dʒ, l.p, m.k, n.sj, ŋ.dʒ, ŋ.g, ŋ.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.dʒj, r.d (0.47%)
0.50 … 0.74	ŋ.kj, ŋ.v, r.k, r.n (0.55%); l.b, l.t, ŋ.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%); ŋ.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%); ŋ.k (3.78%)
5 … 9	n.t (5.28%); n.d (6.85%); ŋ.j (7.32%); r.j (8.98%)
10+	n.j (25.35%)

variety of both first and second consonants, with 6 different phonemes making up 74.65% of C_1, and 8 different phonemes making up 28.74% of C_2.

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 /ŋ.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 as the first consonant (13.78%). 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 stress, that is, stress is primarily based on differences in the loudness of syllables. Which syllable is stressed depends on a mix of which position in a word a syllable occupies and the phonemic shape of syllables. English, on the

other hand, possesses a system where a certain syllable in a word will stay stressed even if prefixes or suffixes are added to a word:

(27) *establish* /ɪˈstæblɪʃ/ [English]
establishment /ɪˈstæblɪʃmənt/
disestablish /dɪsɪˈstæblɪʃ/

In all cases, stress stays on the second syllable of *establish*, whether a prefix or a suffix is added. Thus, it is not possible to predict the stressed syllable in a given word without knowing something about its morphology—one cannot simply count *n* syllables from the front or the back and reach a valid conclusion. German may be an even more illustrative example than English here, as it is still richer in morphology than English:

(28) *reden* /ˈreːdən/ 'talk' (talk-INF) [German]
redete /ˈreːdətə/ 'talked' (talk-PST-1/3SG)
geredet /gəˈreːdət/ 'talked' (PTCP-talk-PTCP)
überredete /ˌyːbərˈreːdətə/ 'persuaded' (over.talk-PST-1/3SG)

In all these words, stress remains on (the first syllable of) the word stem, as is typical of modern Germanic languages, no matter whether one or several affixes are added.[17] The position of this syllable in a given inflected word is effectively variable, though, so counting syllables is, again, no use. In Ayeri, complications are slightly different. To demonstrate, the complete declension paradigm for རྙི *niva* 'eye' is given in Figure 1.4.

Figure 1.4: Declension paradigm for Ayeri རྙི *niva* 'eye'

	Singular		**Plural**	
TOP	*ní.va*	'the eye'	*ni.vá.ye*	'the eyes'
A	*ni.vắng*	'eye'	*ni.va.jáng*	'eyes'
P	*ni.vắs*	'eye'	*ni.vá.jas*	'eyes'
DAT	*ni.vá.yam*[18]	'to the eye'	*ni.vá.jyam*	'to the eyes'
GEN	*ni.vá.na*	'of the eye'	*ni.va.yé.na*	'of the eyes'
LOC	*ni.vá.ya*	'at the eye'	*ni.vá.jya*	'at the eyes'
CAUS	*ni.va.í.sa*	'due to the eye'	*ni.va.jí.sa*	'due to the eyes'
INS	*ni.vá.ri*	'with the eye'	*ni.va.yé.ri*	'with the eyes'

17 Wiese (2006: 282–???) elaborates that …

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 𑖦𑖨𑖪 *nisa*, which is bisyllabic with stress on the first (= penultimate) syllable, to the agent and patient singular forms, 𑖦𑖨𑖪𑖽 *nivāng* and 𑖦𑖨𑖪𑖾 *nivās*, respectively. These are also bisyllabic, however, they are stressed on the second (= ultimate) syllable. Similarly, compare the agent and patient plural forms to each other: the agent plural form 𑖦𑖨𑖪𑖿𑖽 *nivajang* is trisyllabic and has its main stress on the third (= ultimate) syllable, while the equally trisyllabic patient plural form 𑖦𑖨𑖪𑖿𑖾 *nivajas* is stressed on the second (= penultimate) syllable again.

It should have become clear that even though the basic form 𑖦𑖨𑖪 *nisa* has first-syllable stress, *ni* will not necessarily carry stress across the whole paradigm, as it would be the case in English or German. It should also have become clear that the basic algorithm to determine stressed syllables in Ayeri is based on counting syllables from the right edge of a word, although some complications need to be factored in.

Analysis of Stress Patterns in Disyllabic Words

The basic foot in Ayeri is a trochee, and it does not matter whether the syllable is open or closed, or whether there are complex onsets or codas, or no onsets or codas:

(29) a. ˊ× × ||
ha - ri ‘pithy, striking’

b. ˊ× × ||
sa - yan ‘hole, cave’
sem - ba ‘comb’

c. ˊ× × ||
bri - ha ‘grace’
ba - brya ‘(he) mumbles’
a - gu ‘chicken’

That stress assignment is basically trochaic can be deduced from words with more than two syllables. Metricization furthermore runs from right to left:

18 Final-syllable stress is possible as well, also in the plural.

(30) a. × | ×́ × ||
ba - ha - lan 'target, goal'
jar - ma - ya 'pilgrim'

b. ×̀ × | ×́ × ||
ho - ra - ma - ya 'sinner'
ya - ma - na - ti 'causer'

In the case of (30b), 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.[19] Thus, we can make a feature matrix:

Figure 1.5: Types of heavy syllables

	[+ DIPH, − ŋ]	[− DIPH, + ŋ]	[− DIPH, − ŋ]
[+ LONG]	++	++	++
[− LONG]	+	+	−

As the name suggests, heavy syllables always carry stress:

(31) a. × ×́ ||
ma - tay 'summer, wet season'
pa - dang 'mind; heart, mood'
ka - nāy 'I marry' (marry=1SG.TOP)
bras - yāng 'he bathes' (bathe=3SG.M.A)
na - rān 'word; speech'

b. ×́ × ||
kār - yo 'strong'
key - nam 'humans, people'
kan - ka 'mind; heart, mood'

Unfortunately, there are no bisyllabic examples for the feature sets [+ LONG, − DIPH, + ŋ] and [+ LONG, + DIPH, − ŋ] in the first syllable. If there were, they would group with (31b).

So far, we have only looked at heavy syllables combined with regular ones. In this case, another property of heavy syllables will become apparent: long syllables

[19] This would it possible to alternatively analyze diphthongs in /ɪ/ as /Vj/ sequences, essentially.

outweigh those containing a diphthong or ending in /ŋ/. They are essentially superheavy, which is why the cells 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, − ŋ] and the elements from the [+ LONG] line (32a), and the possible combinations between [− LONG, − DIPH, + ŋ] and the [+ LONG] line (32b).

(32) a. × ˊ× ||
bay - hāy 'I govern' (govern=ISG.TOP)
say - lyang 'I sail' (sail=ISG.A)
kay - vān 'container'

b. × ˊ× ||
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 heavy syllable. The question then is, however, what happens if we invert this order. This is more problematic than it sounds, however, as initial [+ LONG, + DIPH, − ŋ] and [+ LONG, − DIPH, + ŋ], as well as final [− LONG, + DIPH, + ŋ] do not occur, insofar there will only be one possible combination here—the reverse of ꠇꠤꠞꠣꠞ *lang-vā* 'in the most tiresome way' from (32b) above:

(33) ˊ× × ||
cā - nang 'love' (love-A)

There is only one one example here, which is very little to make a point, however, other words following this syllalbe pattern, like ꠘꠤꠞꠞꠣ *nāreng* 'rather', for example, behave the same. A long syllable will attract stress either way, as we have already seen in (31b). Another question is what happens if we pit element from the [± LONG] lines against another feature combination of the same line. As above, we will start with the [− LONG] line:

(34) a. × ˊ× ||
bay - tang 'blood'

b. × ˊ× ||
pang - lay 'goddess'

In the case of examples for [+ LONG] pattern combinations, we need to keep in mind again that initial [+ LONG, + DIPH, − ŋ] and [+ LONG, − DIPH, + ŋ] are not

attested, so again, there will only be one possible combination of two syllables with a long vowel:

(35) ˊ× ˊ× ||
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 when we combine each pattern with itself; the combinatorical restrictions mentioned above apply again, of course:

(36) a. × ˊ× ||
kay - vay 'without'
dang - reng 'bell' (bell-A.INAN)

b. ˋ× ˊ× ||
bā - mā 'parents, mom-and-dad'

As demonstrated in (32), 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 for bisyllabic words:

1. Ayeri assigns trochaic stress from the right edge of a word.
2. Syllables ending in /ŋ/ or ones containing a diphthong are considered heavy. They will attract stress.
3. Syllables containing a long vowel are considered superheavy and will override both light and heavy syllables in attracting stress, as 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, the first syllable will receive secondary stress.

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 becomes even longer. Since [± DIPH, ∓ ŋ] behaves the same way throughout, we need not test these combinations separately,

but can subsume them under the label [± HEAVY]. The parameters that need testing, then, are [± HEAVY] in combination with [± 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 [+ HEAVY, −LONG] combined with [− HEAVY, −LONG] in all positions (Figure 1.6). Finding words that fit the associated iterations is not too much of a problem, especially in cases where there is only one heavy syllable.

Figure 1.6: Stress patterns for [± HEAVY, − LONG] in trisyllabic words

− − +	*prantanley*	× | × ́× ||	'question' (question-P.INAN)
− + −	*sarayya*	× | ́× × ||	'(he) bows' (bow-3SG.M)
+ − −	*taykondam*	× | ́× × ||	'break (n.)'
− + +	*ralanghay*	× | × ́× ||	'thumbnail'
+ − +	*kaybunay*	̀× | × ́× ||	'by the way'
+ + −	*maykongas*	× | ́× × ||	'harbor' (harbor-P)
+ + +	*panglay-kay*	̀× | × ́× ||	'a few goddesses' (goddess=few)

It becomes clear from Figure 1.6 that the rules stated at the end of the previous section (p. 37) also hold in the case of trisyllabic words whose syllables alternate short syllables based on the [± HEAVY] feature: ꦥꦿꦟ꧀ꦠꦤ꧀ꦭꦺ *prantanley*, ꦫꦭꦁꦲꦪ꧀ *ralanghay*, ꦏꦪ꧀ꦧုꦤꦪ꧀ *kaybunay*, and ꦥꦁꦭꦪ꧀ꦏꦪ꧀ *panglay-kay* receive final-syllable stress since this is their last heavy syllable. The first syllables of ꦠꦪ꧀ꦏွꦤ꧀ꦝꦩ꧀ *taykondam* and ꦩꦪ꧀ꦏ꧀ꦔꦺꦴꦱ꧀ *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. ꦠꦪ꧀ꦏွꦤ꧀ꦝꦩ꧀ *taykondam* is 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.

x x LONG po.ran.vaː.ŋ
x LONG x raɪ.paː.n.ja
LONG x x noː.ne.ri
x LONG LONG a.saː.jaː.ŋ
LONG x LONG saː.ni.saː.n
LONG LONG x leː.raː.jon
OTHER tʃaː.m.pu.lʊɪ baɪ.si.kaː.n paː.ta.daɪ

Bibliography

Becker, Carsten. 2010. The syllable structure of Ayeri's stems: A statistical analysis. Draft. April 3. Accessed June 20, 2016. *http://benung.nfshost.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.nfshost.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.nfshost.com/archives/3817*.

———. 2016. Benung: The Ayeri language resource. Accessed June 9, 2016. *http://benung.nfshost.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*.

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*.

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

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.

Wiese, Richard. 2006. *The phonology of German*. Reprint. Oxford: Oxford University Press.

Index