

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: August 22, 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. Ayeri is also not derived from any specific real-world language family by means of sound changes. Due to the language's sound and spelling aesthetic being inspired by Austronesian languages, however, occasional overlaps with words existing in those languages may happen, but only accidentally so.

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





Contents

List of Figures	iii
List of Tables	v
Glossing Abbreviations	vi
Preface	vii
o Introduction	i
1 Phonology	7
1.1 Phoneme inventory	7
1.1.1 Consonants	7
1.1.2 Vowels	10
1.2 Phonotactics	12
1.2.1 Number of syllables per word	13
1.2.2 Phonemic makeup of initial syllables	16
1.2.3 Phonemic makeup of medial syllables	20
1.2.4 Phonemic makeup of final syllables	23
1.2.5 Phonemic makeup of single syllables	26
1.2.6 Cross-syllable consonant clusters	28
1.3 Notes on prosody	31
1.3.1 Stress	31
1.3.2 Intonation	40
2 Writing system	45
2.1 Consonants	47
2.2 Vowels	48
2.3 Diacritics	50
2.3.1 Subscript diacritics	50

2.3.2	Prepended diacritics	52
2.3.3	Superscript diacritics	56
2.4	Numerals	57
2.5	Punctuation and abbreviations	57
2.6	Styles	59
3	Morphological typology	63
3.1	Typology	63
3.2	Morphological processes	66
3.2.1	Prefixation	66
3.2.2	Suffixation	69
3.2.3	Reduplication	69
3.2.4	Suprasegmental modification	70
3.3	Marking strategies	71
4	Grammatical categories	75
4.1	Nouns	75
4.1.1	Gender	76
4.1.2	Number	78
4.1.3	Case	81
4.1.4	Prefixes on nouns	93
4.1.5	Compounding	95
4.1.6	Reduplication	104
4.1.7	Nominalization	106
4.2	Pronouns	111
4.2.1	Personal pronouns	111
	Bibliography	113
	Index	117

List of Figures

o.1	First design for an Ayeri script	4
o.2	First draft for Tahano Hikamu	5
1.1	Consonant inventory	8
1.2	Vowel inventory	10
1.3	Syllabification of inflected verbs	13
1.4	Declension paradigm for Ayeri ເຊັ ນີວາ ‘eye’	31
1.5	Types of heavy syllables	33
1.6	Stress patterns for [+ HEAVY, – LONG] in trisyllabic words	36
1.7	Stress patterns for [± HEAVY, + LONG] in trisyllabic words	37
1.8	Examples of stress patterns in compounds	39
2.1	Box script and Hikamu	46
2.2	The consonant graphemes	47
2.3	Additional consonant graphemes	48
2.4	Primary vowel graphemes	48
2.5	Secondary vowel graphemes	49
2.6	Bottom-attaching diacritics	51
2.7	Obligatorily prepended diacritics	53
2.8	Allographically prepended diacritics	53
2.9	Superscript diacritics	56
2.10	The numerals	57
2.11	Common punctuation marks	58
2.12	Less common punctuation marks	58
2.13	Tahano Hikamu, ‘book style’	59
2.14	Tahano Hikamu, ‘angular style’	60
2.15	Tahano Hikamu, ‘hand style’	60
2.16	Tahano Hikamu, ‘blackletter style’	61

4.1	Declension paradigm for Ayeri  <i>badan</i> ‘father’	75
4.2	Declension paradigm for Ayeri  <i>māva</i> ‘mother’	76
4.3	Declension paradigm for Ayeri  <i>kirin</i> ‘street’	76
4.4	Declension paradigm for Ayeri  <i>pera</i> ‘measure’	77
4.5	Grammatical genders in Ayeri	77

List of Tables

1.1	Frequency of words with different numbers of syllables	13
1.2	Frequency of syllable types per word	15
1.3	Frequency of onsets in initial syllables	17
1.4	Frequency of nuclei in initial syllables	18
1.5	Frequency of codas in initial syllables	19
1.6	Frequency of onsets in medial syllables	21
1.7	Frequency of nuclei in medial syllables	22
1.8	Frequency of codas in medial syllables	23
1.9	Frequency of onsets in final syllables	24
1.10	Frequency of nuclei in final syllables	25
1.11	Frequency of codas in final syllables	26
1.12	Frequency of onsets in single syllables	27
1.13	Frequency of nuclei in single syllables	29
1.14	Frequency of codas in single syllables	29
1.15	Frequency of cross-syllable consonant clusters	30
4.1	Compounds in the Ayeri dictionary	95

Glossing Abbreviations

1	First person	IRR	Irrealis
2	Second person	ITER	Iterative
3	Third person	LNK	Linker
A	Agent	LOC	Locative
ACC	Accusative	LOCT	Locative topic
AGR	Agreement	M	Masculine
AGTZ	Agentizer	N	Neuter
AN	Animate	NEG	Negative
AT	Agent topic	NFUT	Near future
CAUS	Causative	NMLZ	Nominalizer
CAUT	Causative topic	NN	Noun
COMP	Comparative	NOM	Nominative
DAT	Dative	NPST	Near past
DATT	Dative topic	P	Patient
DIM	Diminutive	PL	Plural
DYN	Dynamic	PROG	Progressive
F	Feminine	PST	Past
FUT	Future	PT	Patient topic
GEN	Genitive	PTCP	Participle
GENT	Genitive topic	REL	Relative
HAB	Habitative	RFUT	Remote future
HORT	Hortative	RPST	Remote past
IMP	Imperative	SBJ	Subject
INAN	Inanimate	SG	Singular
INDF	Indefinite	SUPL	Superlative
INS	Instrumental	TOP	Topic
INST	Instrumental 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 loss of enthusiasm and creeping neglect.

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

I hope that by transferring my previous grammar writing from LibreOffice to L^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 18, 2016

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

o Introduction

In December 2003, the idea for a new fictional language was born, an idea that turned out to stick with me for over 10 years now.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) with some slight rephrasings and extensions.

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

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

- (2) *Ang layaiyain mecoyalei ling *pinamea.*
 Ang laya-iy-a-in me-coya-lei ling *pinam-ea
 A.SBJ read-3SG.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, however. In the current state of Ayeri, I would translate the sentence as follows:

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

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

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

What was definitely beneficial for the development of Ayeri was the ever increasing amount of linguistics materials available online and my entering university (to

study literature) in 2009, where I learned how to do research and also had a lot of interesting books available at the library.

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

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

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

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

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

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

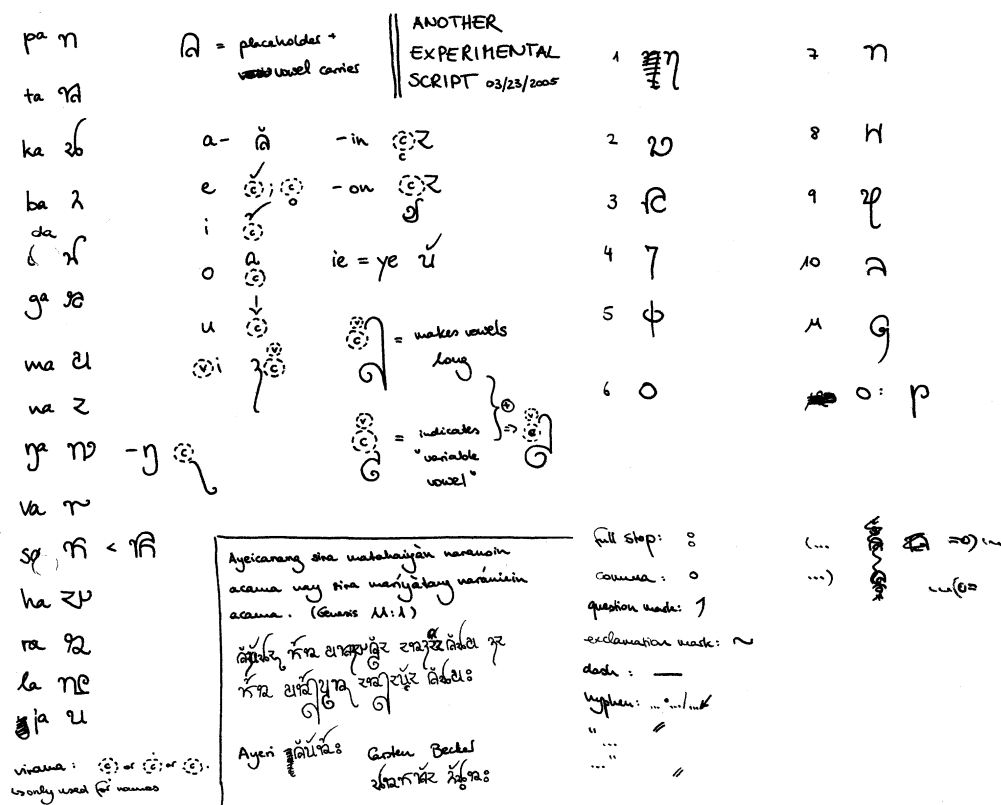
Figure 0.1: First design for an Ayeri script (February 9, 2004)



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

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

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



not being aware of them. In my opinion, the split between being able to apply methods of linguistics to what has grown over the course of more than a decade on the one hand, and discovering and developing new aspects of the language on the other is what makes Ayer an interesting piece of “informed nonsense,” as a colleague of mine once put it.

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

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

1 Phonology

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

1.1 Phoneme inventory

1.1.1 Consonants

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

Regarding allophony, /tj kj/ and /dj gj/ are usually realized as [tʃ] and [dʒ], respectively, except if a homorganic nasal /n/ or /ŋ/ is preceding: for instance, ႁႏႱ *ankyū* /ʼaŋkju/ ‘really’ is realized as [ʼaŋkju], not as *[ʼaŋtʃu] or *[ʼantʃu]. It is important to note, however, that besides this synchronic palatalization process leading to [tʃ] and [dʒ] as *allophones*, there is also a diachronic one in parallel here—or the diachronic process is still ongoing. For example, there is no way to predict whether ႁႏႱ *cuna* ‘original, initial’, ႁႏႱ *panca* ‘finally, eventually’, and ႁႏႱ *vac-* ‘like’, or ႁႏႱ *jaraṇ* ‘pilgrimage’, ႁႏႱ *aja-* ‘play’, and ႁႏႱ *nuy-* ‘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 ၵ -*ye* is commonly contracted to [dʒ] when a case suffix beginning with a vowel follows:²

- (1) a. ဘီယော့ *nyānyeaŋ* → *nyānjāŋ* [ˈnjaːndʒaŋ] ‘persons’ (person-PL-A);
 b. နီယော့ *netuyeaŋ* → *netujāŋ* [neˈtudʒaŋ] ‘brothers’ (brother-PL-P).

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

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

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

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

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

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

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

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

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

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

³ -*ea* also occurs as an allomorph, so that ၵ -*ye* + ၵ -*ea* → ၵ -*yēa*.

- (4) a. ṭavvāṅ [ta'va:ŋ] 'you get' (get=2SG.A),
 b. ḍisyyang [di'sjaŋ] '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:

- (5) ṭipuyya [ti'pu.ja] 'on the grass' (grass-LOC).

1.1.2 Vowels

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

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

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

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

- (6) a. ṭing [mɪŋ] 'can, be able',
 b. ḍenya ['ɛn.ja] 'everyone',
 c. ḍagon ['a.gɔn] 'outer, foreign', and
 d. ṭakur ['pa.kʊr] 'ill, sick'.

/ə/ occurs marginally in the tense prefixes ḱ : $kə$ - 'NPST', ḡ : $mə$ - 'PST', ṛ : $və$ - 'RPST', as well as in the prefix ḡ : $mə$ - 'some, whichever'. Otherwise, [ə] acts as an allophone of /e/ in final unstressed position, for instance, in the word ṭine ['minə] 'affair, matter, issue'.

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

- (7) a. ခဉ်း *nīsa* ‘wanted’, ကုန််း *pasīsa* ‘interesting’;
 b. ခဉ်း *arēn* ‘anyway, however’, ကုန််း *lēra* ‘whore’;
 c. ကုန််း *lā* ‘tongue’, ယုန််း *yāng* ‘he’ (he.A);
 d. ခဉ်း *nōn* ‘will, intention’; and
 e. ခဉ်း *babūan* ‘barbarian’.⁴

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

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

- (9) *Le turayāng taman sinā ang ningay tamala vās.*
Le tura-yāng [taman-Ø]₁ si-Ø₁-na ang ning=ay.Ø tamala vās
 PT.INAN send=3SG.M.A letter-TOP REL-PT.INAN-GEN AT tell=1SG.TOP 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 ခဉ်း *sina* ‘which.GEN’:⁵

- (10) *tamanreng ledanena nā sina koronvāng*
taman-reng [ledan-ena nā]₁ si-na₁ 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 (7c), the word ကုန််း *lā* ‘tongue’ ends in a long vowel, so the question is what happens when a case suffix beginning with a vowel is appended. To avoid a hiatus, a glide /j/ may be inserted, so both of the following renditions are possible:

⁴ I have gone years without dictionary entries for /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, ခဉ်း *babūan* ‘barbarian’ and its adjective ခဉ်း *babū* ‘barbarian (adj.)’ were coined as ကုန််း *prankaye*—things ‘that you put in specifically to make things fit’, another new coining this decision resulted in. Note, however, that it should have always been possible to form words like ခဉ်း *kūbo* ‘as though bitter’, from ခဉ်း *ubo* ‘bitter’ + ခဉ်း *ku-* ‘like, as though’.

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

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

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

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

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

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

1.2 Phonotactics

For the purpose of this statistical analysis, most of the available translations into Ayeri from late 2008 to July 2016 have been used as a text corpus;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

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.

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)	cá.y	gu.máy	ba.bráy
- <i>va</i> (2SG)	cá.va	gúm.va	ba.brá.va
- <i>yam</i> (PTCP)	cá.yam	gúm.yam	bá.bryam

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

1.2.1 Number of syllables per word

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

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

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

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

- (14) a. ၼာ *ang* (AT),
 နှ *nay* ‘and’,
 ရှိ *rua* ‘must’;
- b. ပါဏိ *datau* ‘normal’,
 နာယ *nasay* ‘near to’;
- c. ၼာဝေၼာ *avanyāng* ‘he sinks’ (sink=3SG.M.A),
 ၼာၼာ *tovale* ‘a cloak’ (cloak-P.INAN);
- d. ဃုၼာ *binyanveno* (corner.beautiful, a place name),
 မိတၼာ *mitanena* ‘of the palace’ (palace-GEN);
- e. ဃာယမာ *haruyamanas* ‘a beating’ (beat-PTCP-NMLZ-P),
 နီၼာ *sungkorankihas* ‘geography’ (science.map);
- f. နီၼာ *kaytomayanena* ‘of righteousness’ (righteous-NMLZ-GEN),
 နီၼာ *nasimayajang-ben* ‘all followers’ (follow-AGTZ-PL-A=all).

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

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

- (15) a. Initial CVCC:
 လီၼာ *linktang* /liŋk.'taŋ/ ‘they try’ (try=3PL.M.A),⁸
 လီၼာ *silvnang* /silv.'naŋ/ ‘we see’ (see=1PL.A);
- b. Final CCCV:
 မိဂရီ *migr̥yo* /'mi.gr̥jo/ ‘flourishes’ (flourish-3SG.N),
 မိဂရီ *subr̥yo* /'su.br̥jo/ ‘ceases’ (cease-3SG.N);
- c. Single V:
 အီ *ay* /aɪ/ ‘I’ (1SG.TOP).

The medial and final VC cases may seem like an oddity, but they are mostly due

⁸ The verb stem is found in the dictionary as လီၼာ *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%

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 *ᐱᓂ 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:⁹

```
(16) C = (? : t f | d ʒ | [ptkbgdmnŋvshrljw])
      V = (? : [ae] : ? ɪ | əʊ | [ieaou] : ? | [ɪɛʊə])
```

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

Perhaps most striking about the nuclei of initial syllables presented in Table 1.4 is that plain vowels occur most frequently. As mentioned above, lax vowels are counted here as allophones of tense ones as their distribution is complementary and

⁹ However, *sort* was unable to handle all IPA characters, so *sed 'y/ɛɪʊə:fʒŋ/EIŌU@:SZN/'* had to be used to compensate by transcribing everything into X-SAMPA.

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

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

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

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

are listed here for the sake of completeness. This is the reason why the plain vowels are presented as grouped with their allophones in this table as well as in subsequent ones. Long vowels and diphthongs find themselves below the 5% threshold, and the words with single occurrences are:




- (17) a.  *kuysān* ‘comparison’,
 b.  *nōn* ‘will, intention’,
 c.  *nīsa* ‘wanted’,¹⁰
 d.  *sēyaya* ‘will overcome’ (FUT-overcome-3SG.M),
 e.  *sautan* ‘cork’.

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

¹⁰  *nīsa* and  *nōn* are both related to  *no* ‘want, plan’.

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=IPL.A). There only being two incidences of a CC cluster is very probably an effect of the small sample size. Furthermore, the only unvoiced single coda consonants attested are /s/, /h/, /t/, /tʃ/ and /k/, the latter two only once, /h/ twice:

- (18) a. mehvāng ‘you are supposed to’ (be.supposed.to=2SG.A),¹¹
 robtang ‘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 on consonants to appear in this position, with the exception of /ŋ/ for reasons stated above (see section 1.2.1). Regarding initial clusters, there are further attestations for plosive plus /r/ (except for /kr/). As for clusters with /j/, the only one with a bilabial is /bj/, but the set is extended to /sj/ and /kj/. For clusters with /w/, only /sw/ and /kw/ occur here, while attestations for /bw/ and /rw/ as in initial-syllable onsets are lacking. This does not mean that those combinations are not principally possible in this position, however.

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

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

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>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%

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

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

- (20) a. *పాంగిలాన్* *pangitlan* ‘money change’,¹² and
 b. *తెలుగులు* *telugutong* ‘they survive’ (survive=3PL.N).

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 :ṽ -ye, the locative marker :ṽ -ya, the dative and participle marker :ṽ -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

¹² The word for ‘money’ is ṽ pangis, so (20a) is probably a compound, albeit not a fully transparent one. The word for ‘change’ is ṽ tila-, and there seems to be a nominalizing ṽ -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 head-initial order. Possibly, what happened at the morpheme borders is that ṽ tilān underwent metathesis to *ṽ itlān to match the rhyme of ṽ pangis. *ṽ pangisitlān then underwent irregular haplology (and shortening of the nominalizing suffix) to ṽ pangitlan.

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 ṽ either way, and there is no heeding morpheme breaks

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

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

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 *-an*, which lengthens the vowel of verb stems ending in /a/, as demonstrated in (8). /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 *arēn* ‘anyway, however’; all evidence for /i:/ is from *sirī* ‘due to which’ (see section 1.1.2). The only evidence for /u:/ in the sample is from *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

in placing the diacritic. /CsjV/ will be /C.sjV/ in any case, since Ayeri avoids final consonant clusters if possible, see Table 1.2.

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

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

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 only non-sonorant attested is /k/, which only occurs in *ṣik-ṣik* ‘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

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 mixture 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 *ṣmya* ‘be supposed to’, which is very commonly used as an unconjugatable modal particle.

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

¹⁴ 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.

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%

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

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

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

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

- (23) a. ၵာနၵာန ၵာန *Apayeng-ngas.* [apa'jeŋas];
 b. ၵာနၵာနၵိ ၵာန *Apayeng babo.* [apa'jeŋ 'baxo].

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

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 ၵာ *ang* (see section 1.2.1). /h/ only occurs in the interjections ၵိ *ah!* and ၵိ *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 brack-

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

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

Phoneme	Frequency	Percentage
a	568	47.29%
aɪ	171	14.24%
a:	140	11.66%
i	113	9.41%
i	65	5.41%
ɪ	48	4.00%
e	104	8.66%
ɛ	65	5.41%
ɐ	34	2.83%
ə	5	0.42%
o	45	3.75%
ɔ	30	2.50%
o	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%

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%).

ets. As can be expected, bilabials cluster mostly with bilabials (83/112 purely bilabial CC combinations = 74.11%), alveolars with alveolars (317/948 = 33.44%), and velars with velars (59/207 = 28.51%). However, at least for alveolars and velars, the score is even higher with /j/: 52.64% and 44.93%, respectively. /j/ is also the most common second consonant overall, at 47.8% of all consonant clusters; /n.j/ is the most common cluster at a total of 25.35%. Alveolars provide the highest variety of both first and second consonants, with 6 different phonemes making up 74.65% of C₁, and 8 different phonemes making up 28.74% of C₂.

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 (13.78%) as the first consonant. For second consonants, approximants constitute the largest group at 669 combinations (52.68%), followed by 387 pairs with plosives second (30.47%).

1.3 Notes on prosody

1.3.1 Stress

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

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

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

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

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

¹⁷ For a discussion of terms, see Kager (2007), for instance.

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

¹⁹ Final-syllable stress is possible as well, also in the plural.

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 $\tilde{n}i$ *niva*, which is disyllabic with stress on the first (= penultimate) syllable, to the agent and patient singular forms, $\tilde{n}i$ *nivāng* and $\tilde{n}i$ *nivās*, respectively. These are also disyllabic, however, they are stressed on the second (= ultimate) syllable. Similarly, compare the agent and patient plural forms to each other: the agent plural form $\tilde{n}i$ *nivajang* is trisyllabic and has its main stress on the third (= ultimate) syllable, while the equally trisyllabic patient plural form $\tilde{n}i$ *nivajas* is stressed on the second (= penultimate) syllable again.

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

Analysis of stress patterns in disyllabic words

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

- (25) a. \acute{x} \times ||
 ba - *ri* 'pithy, striking'
- b. \acute{x} \times ||
 sa - *yan* 'hole, cave'
 sem - *ba* 'comb'
- c. \acute{x} \times ||
 bri - *ba* 'grace'
 ba - *brya* '(he) mumbles'
 a - *gu* 'chicken'

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

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

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

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

Figure 1.5: Types of heavy syllables

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

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

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

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

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

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

- (28) a. × × ||
 bay – *bāy* ‘I govern’ (govern=1SG.TOP)
 say – *lyāng* ‘he sails’ (sail=3SG.M.A)
 kay – *vān* ‘container’
- b. × × ||
 kong – *āyn* ‘we enter’ (enter=IPL.TOP)
 keng – *vāng* ‘you notice’ (notice=2SG.A)
 lang – *-vā* ‘in the most tiresome way’ (tiresome=SUPL)

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

- (29) × × ||
 cā – *nang* ‘love’ (love-A)

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

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

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

- (31) $\grave{\times}$ $\acute{\times}$ ||
 mā - *sāy* 'I traveled' (PST-travel=ISG.TOP)

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

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

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

To summarize the above findings:

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

5. Secondary stress is assigned to syllables that are eligible for word stress but which are not in the final foot.

Analysis of stress patterns in trisyllabic words

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

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

Figure 1.6: Stress patterns for $[+ \text{HEAVY}, - \text{LONG}]$ in trisyllabic words

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

It becomes clear from Figure 1.6 that the rules stated at the end of the previous section (p. 35) also hold in the case of trisyllabic words whose syllables alternate short syllables based on the $[\pm \text{HEAVY}]$ feature: ပြဏတၢ်လၢၤ *prantanley*, ရၢလၢၣ်ဂၢၤ *ralanghay*, ကၢၤပၢၣ်နီၤ *kaybunay*, and ပၢၣ်လၢၣ်လၢၤ *panglay-kay* receive final-syllable stress since this is their last heavy syllable. The first syllables of တၢၢ်ကၢၤတၢၢ် *taykondam* and တၢၢ်မၤကၢၤ *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. တၢၢ်ကၢၤတၢၢ် *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.

Carrying out the same analysis as above and moving the feature [+ LONG] through the various positions, we receive the results depicted in Figure 1.7.²² Since long syllables override stress of both light and heavy syllables as pointed out above (p. 35), the example words in this chart contain both of these syllable types. It was not too hard finding examples for all the slots in this case either, except that words with two long syllables in succession are rather rare. Still, only the case of three long syllables must remain unattested.²³

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

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

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

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

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

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

Stress in compounds

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

- (33) a. *Ang bengay kardangirayya ya Litareng.*
 Ang beng=ay.Ø kardang.iray-ya ya Litareng
 AT attend=1SG.TOP school.high-LOC LOC= Litareng
 ‘I attend university in Litareng.’
- b. *Ang pasyye Pila sungkoranyam kibas.*
 Ang pasy-ye Ø Pila sungkoran-yam kahas
 AT be.interested.in-3SG.F TOP Pila science-DAT map
 ‘Pila is interested in geography.’

For purposes of surveying stress patterns, we will only deal with the kind in (33a), though it may be noted that when not being overtly inflected, the second kind of compound will be treated as a word as well: ၵ်း ၵ်း *sungkorankihas* ‘geography’. Another kind of indivisible compound as in (33a) is one formed from reduplication, for instance, ၵ်း ၵ်း *kusang-kusang* ‘model’, from ၵ်း *kusang* ‘double’. Figure 1.8 gives several examples along with their stress patterns. As a reference for the various rules in operation, consider the list above (p. 35).

The first word, ၵ်း ၵ်း *apan-apan*, is not very noteworthy but I included it nonetheless as a reference for regular stress assignment to light syllables. The word decomposes into two feet, each of them has trochaic stress, which does not change when reduplicated. Per rule, the first syllable of the word receives secondary stress while the penultimate syllable bears primary stress. ၵ်း ၵ်း *kusang-kusang* is following the normal rules as well in that a heavy syllable takes stress from a light one. This does not change in reduplication.

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

- (34) $(*dépang \rightarrow depáng) + cāti \rightarrow \left\{ \begin{array}{l} *depàng + cāti \\ dèpang + cāti \end{array} \right\}$

Figure 1.8: Examples of stress patterns in compounds

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

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

Besides the shortening of the second component of the compound, *မာသင်္ဂရာ* *māva-ganeng* retains the stress pattern of its constituents. Since /ma:/ is not in a final foot, it receives secondary stress. Moreover, *မာသင်္ဂရာ* *māvaganeng* and *ပိကုနင်္ဂ* *pikunanding* both show that it is acceptable for two light syllables to follow each other.

1.3.2 Intonation

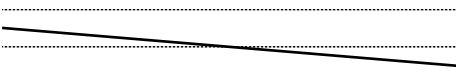
Peterson (2015: 66) writes that if “you’re creating a language on your own and you’re the only speaker, intonation is usually not high on the list of features to focus on, but intonational flavoring is well worth it (read: crucial) when it comes to making an authentic language.” Indeed, this has so far been a rather neglected topic in my work on Ayeri. Even though I made a handful of recordings in the past, I have never considered intonation much. Yet, of course, the spoken words in those recordings do not sound like robot speech either, so there must be intonational patterns that I have been subconsciously applying.²⁴

Since intonation contours are notoriously difficult to display in print, I will give very approximate graphs of pitch in the respective examples for each surveyed pattern. Certainly there will be other patterns as well which would require more detailed gradations, but for the time being, I will only try to briefly describe those that are most prominent.

Declarative statements

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

(35)



<i>Ang</i>	<i>gibayo</i>		<i>Pintemis</i>		<i>minganeri-hen</i>	<i>yona.</i>
Ang	giha-yo	Ø	Pintemis		mingan-eri=hen	yona
AT	blow-3SG.N	TOP	North Wind		ability-INS=all	3SG.N.GEN.

‘The North Wind blew with all of his might.’

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

Yes-no questions

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

(36)



Ang gihayo Pintemis minganeri-ben yona?
 Ang giha-yo Ø Pintemis mingan-eri=hen yona
 AT blow-3SG.N TOP North Wind ability-INS=all 3SG.N.GEN.

‘Did the North Wind blow with all of his might?’

‘Wh-’ questions

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

(37)



Ang engyo mico sinya luga toya sam?
 Ang eng-yo mico sinya-Ø luga toya sam
 AT be.more-3SG.N strong who-TOP among 3PL.N.LOC two

‘Who was the stronger of the two?’

Lists

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

(38)



Le vacyeng seygo, disu, betay nay vasra.
 Le vac=yeng seygo-Ø, disu-Ø, betay-Ø nay vasra-Ø
 PT.INAN like=3SG.F.A apple-TOP, banana-Ø, berry-Ø and nut-Ø

‘She likes apples, bananas, berries and nuts.’

Complement and relative clauses

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

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

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

- (40) a. 
- | | | | | | |
|---------------|----------------|-----------|--------------------------|-------------|----------------|
| <i>Lugaya</i> | <i>asāyāng</i> | <i>si</i> | <i>sitang-naykonyāng</i> | <i>kong</i> | <i>tovaya.</i> |
| Luga-ya | asāya-ang | si | sitang=naykon=yāng | kong | tova-ya |
| pass-3SG.M | traveler-A | REL | self=wrap=3SG.M.A | inside | cloak-LOC |
- ‘A traveler passed who had wrapped himself into a cloak.’
- b. 
- | | | | | | |
|----------------|---------------|-----------|-----------|----------------|--------------|
| <i>Adareng</i> | <i>asāyās</i> | <i>si</i> | <i>le</i> | <i>ninyāng</i> | <i>tova?</i> |
| Ada-reng | asāya-as | si | le | nin=yāng | tova-Ø |
| that-A.INAN | traveler-P | REL | PT.INAN | wear=3SG.M.A | coat-TOP |
- ‘Is that the traveler who wore the coat?’

Contrast

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

(41)



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

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

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

2 Writing system

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

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

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

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

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

Figure 2.1: Box script and Hikamu



(a) Old and aborted draft: Box script



(b) Ayeri's native script: Tahano Hikamu

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

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

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

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

2.1 Consonants

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

Figure 2.2: The consonant graphemes

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

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

- (1) ᠠ + ᠨᠠᠭᠢᠰ
 /pa/ /ɲis/
 ᠠᠨᠠᠭᠢᠰ *pangis* /paŋ.is/ ‘money’

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

- (2) ᠨ + ᠨ → ᠨᠠᠨᠠ
 /na/ /na/ /nana/

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

⟨na⟩ is not deleted, and each ᳵ ⟨na⟩ retains the ability to be modified by diacritics. Tahano Hikamu also has a few ligatures of the kind you would find in Brāhmī scripts, however:

- (3) a. ᳵ᳚ ⟨kwa⟩ ← ᳵ ⟨ka⟩ + ᳚ ⟨va⟩,
 b. ᳵ᳛ ⟨tsa⟩ ← ᳵ ⟨ta⟩ + ᳛ ⟨sa⟩, and
 c. ᳵ᳜ ⟨ksa⟩ ← ᳵ ⟨ka⟩ + ᳜ ⟨sa⟩.

These conjunct letters are, however, not normally employed by Ayeri. Figure 2.3 shows all additional consonants, added to write other languages. Individual languages may adapt the sound values slightly to fit their own purposes.

Figure 2.3: Additional consonant graphemes

/fa/	/wa/	/tsa/	/za/	/ja/	/ʒa/
ᳶ	᳷	᳸	᳹	ᳺ	᳻
/ɕa/	/ksa/	/kwa/	/xa/	/ɣa/	
᳼	᳾	᳿	ᳺ	᳻	

2.2 Vowels

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

- (4) $\text{ᳵ} \rightarrow \text{ᳶ}$
 /pa/ /pe/

Figure 2.4: Primary vowel graphemes

	/i/	/e/	/a/	/o/	/u/	/ə/	/au/
Diacritics	ᳶ	᳷	᳸	᳹	ᳺ	᳻	᳼
Independent	ᳶ	᳷	᳸	᳹	ᳺ	᳻	᳼

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

$$(5) \quad \text{ပ} \rightarrow \text{ပျ} \rightarrow \text{ပျဲ} \\ \quad \quad \quad /ja/ \quad \quad /je/ \quad \quad /jea/$$

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

$$(6) \quad \text{ရ} \rightarrow \text{ရွာ} \quad (\text{ရွဲ}) \\ \quad \quad \quad /ru/ \quad \quad /rwa:/ \quad \quad /ru:a/$$

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

Figure 2.5: Secondary vowel graphemes

/i/	/e/	/a/	/o/	/u/	/ə/	/au/

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

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

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

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

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

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

2.3 Diacritics

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

2.3.1 Subscript diacritics

Figure 2.6 shows the bottom-attaching diacritics. The ‘large diacritics’ cause the secondary slot of consonants to move down below the diacritic. ‘Small diacritics’

Figure 2.6: Bottom-attaching diacritics

Native name	Function	Example
Large diacritics		
 ၵိကခိ <i>tupasati</i> ‘long-maker’	Lengthens the primary vowel of the syllable	ၵ <i>pa</i> → ပျ <i>pā</i>
 ယရိ <i>ya eyra</i> ‘low ya’	⟨ya⟩ following another consonant, also across syllables. Marks palatalization of ၵ ⟨ta⟩, ပျ ⟨da⟩, ခိ ⟨ka⟩, ခိ ⟨ga⟩ and ယ ⟨ya⟩ in Ayeri.	ခိ <i>ara</i> → ခိယ <i>arya</i> ; ၵ <i>ta</i> → ခိယ <i>ca</i>
 ခိယ <i>ringaya</i> ‘raiser’	Palatalizes a consonant (not used in Ayeri)	ခိ <i>ta</i> → ခိယ <i>/tʃa/</i> , <i>/tʃa/</i>
 ၵိယ <i>ulangaya</i> ‘breather’	Aspiration or frication of a consonant (not used in Ayeri)	ခိ <i>ta</i> → ခိယ <i>/tʰa/</i> , <i>/θa/</i>
 ခိယရိ <i>raypāya eyra</i> ‘low stopper’	Glottal stop coda or glottalization of a consonant (consonant letters with ascenders; not used in Ayeri)	ခိ <i>ka</i> → ခိယ <i>/kaʔ/</i> ; ပျ <i>da</i> → ပျ <i>/dʰa/</i>
Small diacritics		
 ခိယ <i>gondaya</i> ‘extinguisher’	Deletes the inherent /a/ of a consonant, e.g. in consonant clusters or closed syllables	ၵ <i>para</i> → ပျ <i>pra</i> , ပျ <i>par</i>
 ခိယ <i>vināti</i> ‘nasalizer’	Indicates a homorganic nasal or nasalizes the vowel, depending on the language	ၵ <i>pada</i> → ပျ <i>panda</i> /panda/ or /pāda/
 ခိယရိ <i>kusangisāti</i> ‘duplicator’	Indicates a geminated or otherwise double consonant	ၵ <i>pala</i> → ပျ <i>palla</i>

can attach in this place as well as secondary vowels, as does the homorganic nasal diacritic ၂ in this diacritic-fraught example:

- (10) $\text{တၢ်} + \text{ပုလၢ်} \rightarrow \text{တၢ်ပုလၢ်}$
 /tʃa:n/ /pu'lɔɪ/ /tʃa:mpu'lɔɪ/
 တၢ်ပုလၢ် *cāmpuluy* 'heterosexual'

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

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

- (11) $\text{ပ} \quad \text{နီၤပၤ} \quad \text{ကၢၤ} \quad \text{နီၤပၤနီၤပၤ}$
Ya nimreng pangan narānyena.
Ya nim-reng pangan-Ø narān-ye-na
 LOCT appear=3SG.INAN.A end-TOP word-PL-GEN
 'It appears at the end of words.'

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

2.3.2 Prepended diacritics

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

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

- (12) a. $\text{ယၢ်} \text{ baday 'hero'} \rightarrow \text{ယၢ်ပၤ} (*\text{ယၢ်ယၢ်}) \text{ badayang 'the hero' (hero-A);}$
 b. $\text{ယၢ်} \text{ tipuy 'grass'} \rightarrow \text{ယၢ်ပၤ} (*\text{ယၢ်ယၢ်}) \text{ tipuyya 'in the grass' (grass-LOC).}$

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

Figure 2.7: Obligatorily prepended diacritics

Native name	Function	Example
ၢ် ၵုၵ်းလၢ <i>lentankusang</i> 'double-sound'	Marks a diphthong with /l/	ၵ် <i>pe</i> → ၵ် <i>pey</i>
ၢ် ၵုၵ်းလၢ <i>tilamaya</i> 'changer'	Marks raised vowels (i.e. umlaut; not used in Ayeri)	ၵ် <i>po</i> → ၵ် <i>pø</i>
ၢ် ၵုၵ်းလၢ <i>hiyamaya</i> 'roller'	Marks retroflex consonants (not used in Ayeri) ⁴	ၵ် <i>ta</i> → ၵ် <i>ṭa</i>

Figure 2.8: Allographically prepended diacritics

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

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

1. when there is no stem or bowl for the regular subscript diacritic to attach to, which is the case for ᳵ <na>, ᳶ <nga>, ᳷ <va>, and ᳸ <wa>:

$$\begin{array}{ll}
 (13) \quad \text{a. } \text{ᳵ} \rightarrow \text{ᳵᳵ} & \text{c. } \text{᳷} \rightarrow \text{᳷᳷} \\
 \quad \quad \quad /na/ & \quad \quad \quad /va/ \\
 \quad \quad \quad /na:/ & \quad \quad \quad /va:/ \\
 \quad \quad \quad \text{b. } \text{ᳶ} \rightarrow \text{ᳶᳶ} & \text{d. } \text{᳸} \rightarrow \text{᳸᳸} \\
 \quad \quad \quad /ŋa/ & \quad \quad \quad /wa/ \\
 \quad \quad \quad /ŋa:/ & \quad \quad \quad /wa:/
 \end{array}$$

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

$$\begin{array}{ccccccc}
 (14) & & + \text{ᳶ} & & + \text{᳷} & & + \text{᳸} \\
 \text{ᳵ} & \rightarrow & \text{ᳵᳶ} & \rightarrow & 2\text{ᳵᳶ} & \rightarrow & 2\text{ᳵᳶ᳷} \\
 /ta/ & & /tʰa/ & & /tʰja/ & & /tʰji/ \\
 & & & & & & /tʰji:/
 \end{array}$$

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

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

$$\begin{array}{ccccccc}
 (15) & & + \text{ᳶ} & & + \text{᳸} & & + \text{ᳵᳵ} \\
 \text{ᳵ} & \rightarrow & 2\text{ᳵ} & \rightarrow & 22\text{ᳵ} & \rightarrow & 3ᳵ2ᳵ \\
 /na/ & & /nʲa/ & & /nʲa:/ & & /nʲa:ɿ/
 \end{array}$$

3. with consonants directly following ᳵ <na>, to avoid a clash with its swash:

$$(16) \quad \underset{/na/}{\text{န}} + \underset{/pa:/}{\text{ပ}} \rightarrow \underset{/napa:/}{\text{နပ}} \quad (*\text{နပ})$$

An exception to this exception occurs, however, when the consonant is not directly following. In this case, no reordering happens, only န ⟨na⟩ *may* reduce its swash in size to accommodate the following prepended diacritic:⁵

$$(17) \quad \underset{/na/}{\text{န}} + \underset{/pai/}{\text{ပိ}} \rightarrow \underset{/napai/}{\text{နပိ}} \quad (? \text{နပိ})$$

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

$$(18) \quad \underset{/di/}{\text{သိ}} + \underset{/pa:/}{\text{ပ}} \rightarrow \underset{/dipa:/}{\text{သိပ}} \quad (*\text{သိပ})$$

Alternatively, the following solution is permissible:

$$(19) \quad \underset{/di/}{\text{သိ}} + \underset{/pa:/}{\text{ပ}} \rightarrow \underset{/dipa:/}{\text{သိပ}}$$

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

$$(20) \quad \begin{array}{l} \underset{/ba:/}{\text{မိ}} + \underset{/ma:/}{\text{မ}} \rightarrow \underset{/ba:ma:/}{\text{မိမ}} \quad (? \text{မိမ}) \\ \text{or:} \quad \underset{/ba:/}{\text{မိ}} + \underset{/ma:/}{\text{မ}} \rightarrow \underset{/ba:ma:/}{\text{မိမ}} \end{array}$$

⁵ The font I am using here is designed so that the reduced combination looks nicer, but if unreduced, န ⟨na⟩’s swash is not so long as to cross the descender of ပ either in this particular case.

2.3.3 Superscript diacritics

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

Figure 2.9: Superscript diacritics

	Native name	Function	Example
◌̇	ၵၵၵၵၵၵ <i>godaya ling</i> 'upper extinguisher'	Deletes inherent /a/ of consonant, e.g. in consonant clusters or closed syllables	$\text{ၶၵ} \textit{vara} \rightarrow \text{ၶၵ} \textit{vra}$
◌̈	ၵၵၵၵၵၵ <i>vināti ling</i> 'upper nasalizer'	Indicates a homorganic nasal or nasalizes the vowel, depending on language/context	$\text{ၵၵ} \textit{naka} \rightarrow \text{ၵၵ̈} \textit{nanka} / \text{ၵၵၵ} \textit{nanka} / \text{ၵၵၵ̈} \textit{nanka}$
◌̄	ၵၵၵၵၵၵၵၵ <i>kusangisāti ling</i> 'upper duplicator'	Indicates a geminated or otherwise double consonant	$\text{ၵၵ} \textit{pana} \rightarrow \text{ၵၵ̄} \textit{panna}$
◌̣	ၵၵၵၵ <i>raypāya</i> 'stopper'	Glottal stop coda or glottalization of a consonant (not used in Ayeri)	$\text{ၵ} \textit{ta} \rightarrow \text{ၵ̣} / \text{taʔ}/;$ $\text{ၶ} \textit{sa} \rightarrow \text{ၶ̣} / \text{s'a}/$

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

2.4 Numerals

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

Figure 2.10: The numerals

1 12	2 𐌒	3 1𐌒	4 𐌗	5 𐌙	6 𐌚
7 𐌛	8 𐌜	9 1𐌛	10 𐌛𐌚	11 𐌛𐌛	12 𐌛𐌚𐌛

2.5 Punctuation and abbreviations

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

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

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

⁶ And one possibly overrepresented by fictional languages due to its rarity in natural languages.

Figure 2.11: Common punctuation marks

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

Figure 2.12: Less common punctuation marks

	Native name	Function	Example
၎်	သံသရာဇ် danarān ‘speaking point’	Quotation marks	ဓာယျာ ‘မိန့်ရ်’ Narayāng “Manisu!” ‘He says, “Hello!”’
၎်	သံသရာဇ် dankayvo ‘beside-point’	Bracketing of text	ဓာယျာ ဘဏ်း babis (larau) ‘a (nice) day’
၎်	သံသရာဇ် dangaran ‘name-point’	Explicitly marks a name as such. For the closing name parenthesis, ၎် can be found as well.	အိမ်အိမ်အိမ် Ajān Savati; ပိလ်လယ် Pila Lay Maran
၎်	သံသရာဇ် dansinday ‘number-point’	Marks (duo)decimal fractions	၁၇.၄၅၈၂၁၀ 17.45B82 ‘19.37482’
၎်	သံသရာဇ် adrumaya ‘breaker’	Marks line breaks within a phrase	

Two very common abbreviations are symbolic in nature, like the ampersand & in the Latin alphabet, and incidentally, they correspond to it in that the very common small word *nay* ‘and’ may be abbreviated as *n̄*. Based on this, its reduplicated form *naynay* ‘furthermore, also’ may be abbreviated as *n̄̄*.

⁷ For examples, see Everson (2008), or *Wikipedia*.

it to Tahano Hikamu at some point. As mentioned above as well, there are no subscript letters and in Ayeri, and the number of large swirling diacritics is also rather low, so there is still definitely a difference in appearance. The ‘angular’ style is also the one that is comparable in function to our bold face or italic style letters, since it is used in captions or to highlight special text within running text. This letter style (ဆွဲ၃ *binya* ‘angular’) is displayed in Figure 2.14.

Figure 2.14: Tahano Hikamu, ‘angular style’



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

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

Figure 2.15: Tahano Hikamu, ‘hand style’



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

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

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

Figure 2.16: Tahano Hikamu, ‘blackletter style’



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

3 Morphological typology

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

3.1 Typology

For the largest part, Ayeri is an *agglutinative* language. Comrie (1989) says of agglutinating languages that in these, typically,

a word may consist of more than one morpheme, but the boundaries between morphemes in the word are always clear-cut; moreover, a given morpheme has at least a reasonably invariant shape, so that the identification of morphemes in terms of their phonetic shape is also straightforward. [...] As is suggested by the term agglutinating (cf. Latin *gluten* 'glue'), it is as if the various affixes were just glued on one after the other (or one before the other, with prefixes). (43–44)

In Ayeri, root morphemes are modified by affixes for the purposes of inflection and derivation, and these affixes, in the form of suffixes more specifically, can be stacked, especially on verbs. Indeed, they vary little, so that they are always easily recognizable. Suffixation in Ayeri is especially prominent on verbs:

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

'He always eats his eggs without salt and pepper.'

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

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

The principle of not conflating several grammatical functions into a single suffix can be observed in (2a) regarding the word 𐌆𐌶𐌵𐌹𐌶𐌵 *tamanyeley* ‘letters’, in which the plural marker 𐌶𐌵 -*ye* is distinct from the inanimate-patient case marker 𐌶𐌵𐌹𐌶𐌵 -*ley* (the latter, however, conflates animacy and case). Strictly speaking, the pronoun 𐌶𐌵𐌹𐌶𐌵 *yeyam* ‘to her’ is also composed, namely of the third person feminine base form 𐌶𐌵𐌹𐌶𐌵 *ye* and the dative case marker 𐌶𐌵𐌹𐌶𐌵 *yam*. Example (2b) is one we have already encountered before (p. 11). Here, the relative pronoun, 𐌶𐌵𐌹𐌶𐌵 *sinā* ‘of/about which’ is inflected for genitive case, and stress on the usually unstressed last syllable suprasegmentally marks that this form is contracted from 𐌶𐌵𐌹𐌶𐌵𐌹𐌶𐌵 *sileyena* (*si-ley-ena*, REL-P.INAN-GEN).

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

- (3) *Le no intoyyang eda-peham.*
 Le no int-oy=yang eda=peham-Ø
 PT.INAN want buy-NEG=ISG.A this=carpet-TOP
 ‘I do not want to buy this carpet.’

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

always bound and always unaccented show considerable syntactic freedom, in the sense that they can be associated with words of a variety of morphosyntactic categories. Frequently, such a *bound word* is semantically associated with an entire

consituent while being phonologically attached to one word of this constituent, and ordinarily the bound word is located at the very margins of the word, standing outside even inflectional affixes. (6)

This is the case with the marker ᄃᆞᆫ *manga*, which is treated as an independent word, but can modify verbs and prepositions—heads of verb phrases (VPs) and prepositional phrases (PPs), respectively—is unstressed and appears at the margin of its modification target:

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

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

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

Perpendicular to the axis isolation–agglutination runs the axis analytic–synthetic. On the latter axis, Ayeri scores mostly as *synthetic*, since it prefers compactness over spreading a construction over several words, though it does not incorporate object noun phrases (NPs) and it is not possible to form ‘sentence-words’

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

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

either, so it is not going so far as to be polysynthetic (Comrie 1989: 45–46). It is nonetheless theoretically possible, due to suffixation being a prominent pattern, to form foot-long words like

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

Cases of analytic morphology are compound prepositions as we have seen with *ᄒᆞᆫᆫᆫᆫᆫᆫ manga luga* ‘through’ in (4b), but verbs as well show analytic structures not only with the progressive marker, but also with modals:

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

Most of the information the VP contains in this example is marked on the content verb, *ᄒᆞᆫᆫᆫᆫᆫᆫ saba-* ‘come’, except for ability, which is expressed by the particle *ᄒᆞᆫᆫᆫᆫᆫᆫ ming* ‘can’. *ᄒᆞᆫᆫᆫᆫᆫᆫ ming* is an uninflected form of the verb expressing ability and may be counted as an auxiliary verb in that the full semantic content of the VP is spread out over two verb forms, one major, one minor—this probably should not be understood as a serial verb construction, however (Aikhenvald 2006).³ Consider also the following example in which *ᄒᆞᆫᆫᆫᆫᆫᆫ ming* is inflected like a regular verb:

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

3.2 Morphological processes

3.2.1 Prefixation

Prefixes in Ayeri apply mainly to verbs, but nouns, pronouns, adjectives and conjunctions as well can appear with them, some of which may be clitics; reasons for their being clitics will be given at the appropriate cases in the sections on the various

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

parts of speech. With verbs, prefixes that are most certainly ‘true’ prefixes—that is, morphemes that have been semantically bleached by grammaticalization to the point where they only express grammatical functions (Lehmann 2015: 157ff.) and which subcategorize words rather than phrases (Klavans 1985: 117)—are the tense prefixes marking both three degrees of past and future tense, for example:

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

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

- (9) *Da-mingya ang Diyan.*
 Da-ming-ya ang Diyan.
 so-can-3SG.M A Diyan
 ‘Diyan can (do it).’
- (10) *Sitang-kecāng.*
 Sitang=ket-yāng
 self-wash=3SG.M.A
 ‘He washes himself.’

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

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

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

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

The demonstrative prefixes are also used to form the demonstrative pronouns မ်ၤ *edanya* ‘this one’, မ်ၤ *adanya* ‘that one’ and မ်ၤ *danya* ‘such one’. A special case in this regard is the postposition မ်ၤ *da-nārya* ‘in spite of, despite’ where မ်ၤ *da-* combines with the conjunction ခၢၣ် *nārya* ‘but, although, except’. Originally, မ်ၤ *dikapisa* ‘respective’ is derived from မ်ၤ *da-* + မ်ၤ *ikapisa* ‘bound, dependent’, which is an example of a combination with an adjective. There is also a fixed adverbial expression using one of these prefixes, မ်ၤ *eda-tadayyam* ‘for the time being, for now’ (this=time-DAT).

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

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

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

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

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

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

3.2.2 Suffixation

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

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

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

As mentioned above, quantifiers appear as enclitics on almost any type of content word, like on the adverb 𐌆𐌆𐌆𐌆 *para* ‘fast’ in the following example:

- (16) *Tigalyeng para-ma.*
 Tigal=yeng para=ma
 swim=3SG.F.A fast=enough
 ‘She swims fast enough.’

3.2.3 Reduplication

There are two patterns of reduplication for verbs, one with complete reduplication of the imperative form to create a hortative statement (17a), and one with partial

reduplication as a way to express that an action takes place again, that is, partial reduplication expresses a frequentative, as it were (17b). The imperative iterative, then, has a hortative function as well (17c):

- (17) a. *naru-naru*
 naru~nara-u
 HORT~speak-IMP
 ‘let us speak’
- b. *na-narayeng*
 na~nara=yeng
 ITER~speak=3SG.F.A
 ‘she speaks again’
- c. *na-naru*
 na~nara-u
 ITER~speak-IMP
 ‘let us speak again’

With nouns, full reduplication is used to create a diminutive form (18a), though some reduplications are also lexicalized and may use roots from other parts of speech as well to form nouns, for instance, the words in (18b–d). There are also a number of adjectives for which there exists a lexical reduplication with an intensifying meaning; (19) lists a few examples. This, however, is not a productive derivation strategy.

- (18) a. *veney* ‘dog’ → *veney-veney* ‘little dog, doggie’
 b. *gan* ‘child’ → *gan-gan* ‘grandchild’
 c. *kusang* ‘double (adj.)’ → *kusang-kusang* ‘model’
 d. *veba* ‘build’ → *veba-veba* ‘tinkering’
- (19) a. *apan* ‘wide’ → *apan-apan* ‘extensive’
 b. *kebay* ‘alone’ → *kebay-kebay* ‘all alone’
 c. *pisu* ‘tired’ → *pisu-pisu* ‘exhausting’

3.2.4 Suprasegmental modification

As written above (section 1.1.2), case agreement on a complex-marked relative pronoun can drop out under certain circumstances and is replaced by compensatory stress on the secondary case marker, which lengthens the syllable’s nucleus vowel:

- (20) ... *tamanley sinā* (**sina*) *ang ningay tamala vās*
 ... [taman-ley]₁ si-Ø₁-na (*si-na₁) ang ning=ay.Ø tamala vās
 ... letter-P.INAN REL-PT.INAN-GEN (*REL-GEN) AT tell=ISG.TOP yesterday 2SG.P
 ‘... the letter which (*whose) I told you about yesterday’

This can be reinterpreted so that vowel length/stress itself is what signifies the agreement of the relativizer with the preceding NP. Which grammatical role the relativizer’s head instantiates as an agreement controller is essentially underspecified, hence I will gloss it as -AGR in the following example instead of as full -P.INAN:

- (21) /si.lei.'ena/ → /si.'na(:)/
 /si-lei-ena/ → /si-'-na(:)/
 REL-P.INAN-GEN REL-AGR-GEN

Since ɹ *na* as a light syllable cannot be stressed in word-final position under normal circumstances, it has to lengthen to ɹ̄ *nā*.

3.3 Marking strategies

With regards to the dichotomy head–dependent marking, Ayeri is rather thoroughly dependent marking, albeit with the exception of agreement morphology on the verb. Dependent marking is exhibited, for instance, in the expression of possessive relationships, where the dependent is marked for genitive case:

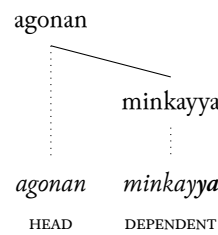
- (22) a. *dema*

dema na Tuvo
 HEAD DEPENDENT
dema na Tuvo
 dema na Tuvo
 aunt GEN Tuvo
 ‘Tuvo’s aunt’
- b. *kasu*

kasu bariri nā
 HEAD DEPENDENT
kasu bariri nā
 kasu bari-ri nā
 basket meat-INS ISG.GEN
 ‘my basket of meat’

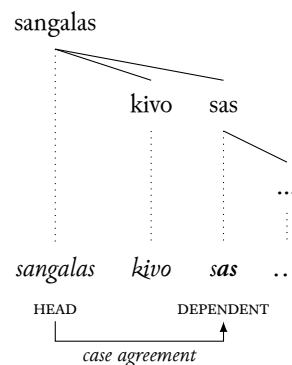
In (22a), 𐌆𐌋𐌆 *Tuvo* is grammatically in possession of her 𐌆𐌋𐌆 *dema* ‘aunt’; the possessee forms the head of the phrase while it is modified by the possessor, which receives the marking. In (22b), 𐌆𐌋𐌆 *kasu* ‘basket’ forms the head and thus also the possessee while ɹ̄ *nā* ‘my’ serves as the dependent possessor; the genitive case is, then again, marked on the dependent. A further example of dependent marking is the locative case, which is marked on the prepositional object while the preposition itself, as the head of the PP, does not receive marking:

- (23) *agonan minkayya*
 agonan minkay-**ya**
 outside village-**LOC**
 ‘outside of the village’



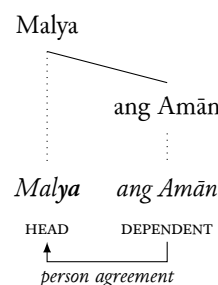
The relativizer, likewise, may agree in case with the NP in the matrix clause to which it links the relative clause. This typically happens—in terms of linear succession of words at the surface level of the clause—if the relativizer cannot be immediately adjacent to the NP which the relative clause modifies, for example, when an adjective or a possessive pronoun is following the noun:

- (24) *sangalas kivo sas ...*
 sangal-as kivo s-**as** ...
 room-P small REL-**P** ...
 ‘the small room which ...’



The only instance of head-marking there is in Ayeri is person-marking on the verb, which manifests when the NP following the verb (agent or patient) is not pronominal and thus there is no pronoun to cliticize to the verb stem, but the verb still receives a suffix that indicates a relation with, usually, the agent NP:

- (25) *Malya ang Amān.*
 Mal-**ya** ang Amān
 sing-**3SG.M** A Amān
 ‘Amān sings.’



Sentences containing more than one NP also have topic marking on the verb, so that the verb agrees with one of the NPs in topicality. This may be the NP it has person agreement with or any other NP. The topicalized NP as a dependent of the verb is, in turn, zero-marked, so that the marking relationship for topics is bilateral and thus mixed:

- (26) *Sa manya ang Ajān Pila.*
 Sa man-ya ang Amān Ø Pila
 PT greet-3SG.M A Ajān TOP Pila
 ‘Pila, Ajān greets her.’



In this example, the verb exhibits canonic agreement with the agent, *ang Ajān*, in person, gender, and number. It is additionally marked for a patient topic, *Pila*, and thus serves as an agreement target for two different controller NPs. The agreement relation is of a different kind for each of the two NPs, however.

4 Grammatical categories

While the previous chapter was about general mechanisms of marking in Ayeri, this chapter will dive into the various parts of speech in order to define their morphology with a closer look. I will begin with nouns as the main carriers of meaning, then deal with other parts of speech that regularly feature in or in combination with the noun phrase—pronouns, adjectives, and adpositions. Following this, there will be a discussion of verbs and adverbs before moving on to numerals and conjunctions.

4.1 Nouns

Nouns in Ayeri have *gender* and *number* as their inherent grammatical properties. Besides common nouns, there are, of course, also proper nouns (i.e. names) and nominalizations. Nouns, as the heads of NPs, are also assigned *case* by the verb, which is a third grammatical property they display. For an illustration of the declension paradigms, compare Figures 4.1–4.4.

Figure 4.1: Declension paradigm for Ayeri  *badan* ‘father’ (animate; consonantal root)

		Singular		Plural
TOP	<i>badan</i>	‘the father’	<i>badanye</i>	‘the fathers’
A	<i>badanang</i>	‘father’	<i>badanjang</i>	‘fathers’
P	<i>badanas</i>	‘father’ (obj.)	<i>badanjas</i>	‘fathers’ (obj.)
DAT	<i>badanyam</i>	‘to the father’	<i>badanjyam</i>	‘to the fathers’
GEN	<i>badanena</i>	‘of the father’	<i>badanyena</i>	‘of the fathers’
LOC	<i>badanya</i>	‘at the father’	<i>badanjya</i>	‘at the fathers’
CAUS	<i>badanisa</i>	‘due to the father’	<i>badanjisa</i>	‘due to the fathers’
INS	<i>badaneri</i>	‘with the father’	<i>badanyeri</i>	‘with the fathers’

Figure 4.2: Declension paradigm for Ayeri māva ‘mother’ (animate; vocalic root)

		Singular		Plural
TOP	<i>māva</i>	‘the mother’	<i>māvaye</i>	‘the mothers’
A	<i>māvāng</i>	‘mother’	<i>māvajang</i>	‘mothers’
P	<i>māvās</i>	‘mother’ (obj.)	<i>māvajas</i>	‘mothers’ (obj.)
DAT	<i>māvayam</i>	‘to the mother’	<i>māvajyam</i>	‘to the mothers’
GEN	<i>māvana</i>	‘of the mother’	<i>māvayena</i>	‘of the mothers’
LOC	<i>māvaya</i>	‘at the mother’	<i>māvajya</i>	‘at the mothers’
CAUS	<i>māvaisa</i>	‘due to the mother’	<i>māvajisa</i>	‘due to the mothers’
INS	<i>māvāri</i>	‘with the mother’	<i>māvayeri</i>	‘with the mothers’

Figure 4.3: Declension paradigm for Ayeri kirin ‘street’ (inanimate; consonantal root)

		Singular		Plural
TOP	<i>kirin</i>	‘the street’	<i>kirinye</i>	‘the streets’
A	<i>kirinreng</i>	‘street’	<i>kirinyereng</i>	‘streets’
P	<i>kirinley</i>	‘street’ (obj.)	<i>kirinyeley</i>	‘streets’ (obj.)
DAT	<i>kirinyam</i>	‘to the street’	<i>kirinjyam</i>	‘to the streets’
GEN	<i>kirinena</i>	‘of the street’	<i>kirinyena</i>	‘of the streets’
LOC	<i>kirinya</i>	‘at the street’	<i>kirinjya</i>	‘at the streets’
CAUS	<i>kirinisa</i>	‘due to the street’	<i>kirinjisa</i>	‘due to the streets’
INS	<i>kirineri</i>	‘with the street’	<i>kirinyeri</i>	‘with the streets’

4.1.1 Gender

Grammatical gender in Ayeri consists of two tiers which are subdivided into four classes based on a mixture of semantic and epistemic properties, see Figure 4.5. The animate gender refers, broadly speaking, to entities that are considered alive or are closely associated with living things, such as events, concepts, or activities executed or connected to them. The ‘masculine’ and ‘feminine’ subcategories are applied to humans, animals whose sex is known (for example on behalf of breeding them or keeping them as pets), and gods—basically anything that shows sexual dimorphism or is assumed to be an exponent of it as well as nouns referring to such entities in a functional way, for instance, badan ‘father’ and māva ‘mother’. The remainder falls into the ‘neuter’ category—plants, for instance, body parts, or animals whose

Figure 4.4: Declension paradigm for Ayeri ^{၁၁} *pera* ‘measure’ (inanimate; vocalic root)

	Singular		Plural	
TOP	<i>pera</i>	‘the measure’	<i>peraye</i>	‘the measures’
A	<i>perareng</i>	‘measure’	<i>perayereng</i>	‘measures’
P	<i>peraley</i>	‘measure’ (obj.)	<i>perayeley</i>	‘measures’ (obj.)
DAT	<i>perayam</i>	‘to the measure’	<i>perajyam</i>	‘to the measures’
GEN	<i>perana</i>	‘of the measure’	<i>perayena</i>	‘of the measures’
LOC	<i>peraya</i>	‘at the measure’	<i>perajya</i>	‘at the measures’
CAUS	<i>peraisa</i>	‘due to the measure’	<i>perajisa</i>	‘due to the measures’
INS	<i>perari</i>	‘with the measure’	<i>perayeri</i>	‘with the measures’

sex is unknown. The ‘inanimate’ category typically contains materials and things, such as tools. Furthermore, animals and plants change their category to inanimate as well if they serve as food. There are exceptions to either group, where elements appear in them for no obviously discernable reason. In order to illustrate, here are a few examples for each category:

- (i) a. Animate masculine:
^{၁၂} *badan* ‘father’, ^{၁၃} *netu* ‘brother’, ^{၁၄} *aguyan* ‘rooster’, ^{၁၅} *Ajān*, ^{၁၆} *Latun*;
- b. Animate feminine:
^{၁၇} *māva* ‘mother’, ^{၁၈} *kina* ‘sister’, ^{၁၉} *aguvay* ‘hen’, ^{၂၀} *Maba*, ^{၂၁} *Trānay*;
- c. Animate neuter:
^{၂၂} *adang* ‘palm tree’, ^{၂၃} *mino* ‘color’, ^{၂၄} *ikam* ‘deer’, ^{၂၅} *kadān* ‘harvest’, ^{၂၆} *cān* ‘love’, ^{၂၇} *nanga* ‘house’, ^{၂၈} *tampu* ‘luck’, ^{၂၉} *yila* ‘foot’;
- d. Inanimate:
^{၃၀} *abal* ‘sand’, ^{၃၁} *bema* ‘egg’, ^{၃၂} *kahan* ‘spear’, ^{၃၃} *melung* ‘yogurt’, ^{၃၄} *nusān* ‘damage’, ^{၃၅} *payutān* ‘mathematics’.

Figure 4.5: Grammatical genders in Ayeri



There are also a number of doublets like French *le livre* ‘the book’ and *la livre* ‘the pound’, for instance, ႁႏ *banan* (an.) ‘kindness, charity’ or ႁႏ *binu* (an.) ‘color’ on the one hand, and ႁႏ *banan* (inan.) ‘quality’ or ႁႏ *binu* (inan.) ‘paint’ on the other. Gender is reified by case marking as well as verb agreement; it is not possible to read the gender of a noun from its phonological makeup. The following example illustrates differences in case marking and agreement (inherent information on grammatical features underneath the NPs):

- (2) a. *Ang konja badan hemale.*
 Ang kond-ya badan-Ø hema-ley
 [3SG.M.AN] [3SG.INAN]
 AT.AN eat-3SG.M.AN father-TOP egg-P.INAN
 ‘Father eats an egg.’
- b. *Sa tombara kahanreng burang.*
 Sa tomb-ara kahan-reng burang-Ø
 [3SG.INAN] [3SG.N.AN]
 PT.AN kill-3SG.INAN spear-A.INAN animal-TOP
 ‘The animal, the spear kills it.’

In example (2a), the noun in the agent NP, ႁႏ *badan* ‘father’, bears the features [+ ANIMATE, + MASCULINE], which triggers the animate agent topic agreement marker ႁႏ *ang* on the verb, since the agent NP is also topicalized. The verb also agrees in person and number with the agent NP by way of the person marker ႁႏ *-ya* for third person singular masculine. The object of the sentence, ႁႏ *hema* ‘egg’, on the other hand bears the feature [− ANIMATE], so it receives the inanimate patient case marker ႁႏ *-ley* rather than its animate counterpart ႁႏ *-as*.

In (2b), on the other hand, we see an inanimate agent, ႁႏ *kahan* ‘spear’, so the verb receives the marker ႁႏ *-ara* for third person singular inanimate rather than its animate neuter counterpart ႁႏ *-yo*. The (non-topicalized) NP’s case marking shows that the agent of the clause is inanimate: ႁႏ *kahan* carries the marker ႁႏ *-reng*, which marks it as an inanimate agent. The object of the sentence, ႁႏ *burang* ‘animal’, is also the topic, hence topic agreement on the verb uses the marker ႁႏ *sa* according to the NP being animate, rather than its inanimate counterpart ႁႏ *le*.

4.1.2 Number

Ayeri only distinguishes singular and plural in nouns, which receive plural marking; verbs, then, agree with agent NPs in number in the canonical case. Ordinarily, nouns in Ayeri are countable, however, there is also a group of uncountable nouns as well as a (small) group of nouns which are always plural. As above, I will list a few words from each group for illustration:

(3) a. Countable nouns:

- ၼံ့ဗျဲ *ajam* ‘toy’ — ၼံ့ဗျဲ *ajamye* ‘toys’,
 ပျီ *devo* ‘head’ — ပျီ *devoye* ‘heads’,
 ၼံ့ဇီ *inun* ‘fish’ — ၼံ့ဇီ *inunye* ‘fish’ (pl.),
 နဲ *netu* ‘brother’ — နဲ *netuye* ‘brothers’;

b. Uncountable nouns:

ၼံ့ဗျဲ *abal* ‘sand’, ၼံ့ဗျဲ *bakay* ‘stuff’, ၼံ့ဗျဲ *gabān* ‘hope’, ၼံ့ဗျဲ *mingan* ‘ability’;

c. Plurale tantum nouns:

ၼံ့ဗျဲ *burang* ‘lifestock, cattle’,¹ ၼံ့ဗျဲ *ganengan* ‘siblings’, ၼံ့ဗျဲ *keynam* ‘people’, ၼံ့ဗျဲ *tang* ‘ears’.

Most concrete things that exist as discrete entities are countable, also, for instance, animals and livestock—fish, deer, sheep etc. are thus countable, unlike in English; pants, pliers, scissors, glasses, etc. are by default singular as well. Uncountable, on the other hand, are materials in general or abstract concepts. There are also a number of nouns which are plural by default, most notably entities which often occur in groups, but there is as well the odd word for which there seems to be no reason to be included in this group, for instance, ၼံ့ဗျဲ *bino* ‘paint’, and ၼံ့ဗျဲ *gimbay* ‘sorrows’. A few body parts are also plurale tantum nouns, especially those which occur in pairs (ၼံ့ *niva* ‘eye’ is a notable exception).

As demonstrated in (3a), the noun plural marker is ပျီ *-ye*, which in native orthography also occurs in the variant နဲ or နဲ. As described above (section 1.1.1, p. 7), the plural marker may also be reduced to [dʒ] <-j> before case suffixes that begin with /j/ or with a vowel other than /e/, like ၼံ့ဗျဲ *-ang* (A) or ပျီ *-yam* (DAT):

- (4) a. ပျီ *diranang* (uncle-A) + ပျီ *-ye* (PL) → ပျီ *diranjang* (uncle-PL-A),
 b. ပျီ *diranena* (uncle-GEN) + ပျီ *-ye* (PL) → ပျီ *diranyena* (uncle-PL-GEN),
 c. ပျီ *diranyam* (uncle-DAT) + ပျီ *-ye* (PL) → ပျီ *diranjyam* (uncle-PL-DAT).

For plurale tantum, to express a singular entity, it is always possible to use a genitive phrase like —ၼံ့ဗျဲ *...-ena men* ‘one of ...’ (...-GEN one), for instance:

- (5) a. *Nupayon tangang nā.*
 Nupa-yon tang-ang nā
 hurt-3PL.N ears-A ISG.GEN
 ‘My ears hurt.’

¹ Specifically in this meaning; ၼံ့ဗျဲ *burang* can also simply mean ‘animal’, in which case there is a plural form ၼံ့ဗျဲ *burangye* ‘animals’.

- b. *Na nupareng tang nā men.*
 Na nupa=reng tang-Ø nā men
 GENT hurt=3SG.INAN.A ears-TOP 1SG.GEN one
 ‘Of my ears, it hurts one.’

Number in nouns can also be manipulated by quantifiers which attach to declined nouns as suffixes. In this case, when plurality is indicated by the quantifier, the noun is not additionally marked for number; the verb, however, keeps agreeing in number:

- (6) a. *Ajayan ganjang kivo.*
 Aja-yon gan-ye-ang kivo
 play-3SG.N child-PL-A small
 ‘The small children are playing.’
 b. *Ajayan ganang-ikan kivo.*
 Aja-yon gan-ang=ikan kivo.
 play-3SG.N child-A=many small
 ‘Many small children are playing.’

Likewise, when nouns are modified by numerals, plurality is not normally marked again on the noun. In example (7a), we see a plural noun, 𐌂𐌆𐌇 *nanga* ‘house’, and in (7b) the same phrase is repeated again with plurality implied by the use of a numeral, 𐌂𐌆𐌇 *sam* ‘two’; the plural noun itself appears unmarked in its singular form in this case.

- (7) a. *Ang no vehya sitang-yām nangajas veno nay hiro.*
 Ang no veh=ya.Ø sitang=yām nanga-ye-as veno nay hiro
 AT want build-3SG.M.TOP self=3SG.M.DAT house-PL-P pretty and new
 ‘He wants to build himself pretty new houses.’
 b. *Ang no vehya sitang-yām nangās sam veno nay hiro.*
 Ang no veh=ya.Ø sitang=yām nanga-as sam veno nay hiro
 AT want build-3SG.M.TOP self=3SG.M.DAT house-P two pretty and new
 ‘He wants to build himself two pretty new houses.’

An exception to this is the use of words for the numeral powers, like 𐌂𐌆𐌇 *lan* ‘dozen’, 𐌂𐌆𐌇𐌆𐌇 *menang* ‘gross’, 𐌂𐌆𐌇𐌆𐌇𐌆𐌇 *samang* ‘myriad’, etc. in an unspecified way like ‘dozens of people’. In this case, to convey that the numeral is not to be understood as a precise value, the modified noun will appear in the plural—even if it is a plurale tantum like 𐌂𐌆𐌇𐌆𐌇 *keynam* ‘people’:

- (8) *Bengyon keynamjang menang.*
 Beng-yon keynam-ye-ang menang
 attend-3SG.N people-PL-A gross
 ‘Hundreds of people attended.’

As we have seen in various examples above, proper nouns in Ayeri do not receive inflection for case by suffixes as common nouns do, and for the purpose of number they are treated as uncountable in Ayeri—they resist inflection by suffixation, marking their special status.² However, they can still be modified by quantifiers and quantifying suffixes; verb agreement as well can be used to indicate plurality:

- (9) a. *Sabayan cabo ekeng ang Yan.*
 Saha-yan cabo ekeng ang Yan
 come-3PL.M late too A Yan
 ‘The Yans are coming too late.’
 b. *Ang apatang sa Yan-ikan.*
 Ang apa=teng sa Yan=ikan
 AT laugh=3PL.F.A P Yan=all
 ‘They laughed at (all) the Yans.’

4.1.3 Case

As demonstrated in the declension tables at the beginning of this section (Figures 4.1–4.4), Ayeri’s NPs are marked for case, which is governed by the verb. Since Ayeri uses a split alignment system with some additional complications, it is not very straightforward, in my opinion, to use the classical labels of nominative (S/A) and accusative (O), or of absolutive (S/P) and ergative (O) for the first two core roles. Hence, I will be using the terms ‘agent’ and ‘patient’, which I hope brings about some more clarity, especially when discussing the mentioned complications later on.

Agent

To quote Fillmore (2003 [1968]), what I call ‘agent’ here is “the case of the typically animate perceived instigator of the action identified by the verb” (46). Fillmore himself qualifies this definition, however, in that the “escape qualification ‘typically’ expresses my awareness that contexts which I will say require agents are sometimes

² Many common names in Ayeri are derived from regular words in the language, so the language needs to have a way to distinguish between regular use and use as a name. For instance, the name *uꞑ Yan* also means ‘boy, son’ as a common noun.

occupied by ‘inanimate’ nouns like robot or ‘human institution’ nouns like nation” (Fillmore 2003 [1968]: 46, footnote 31). Payne (1997) summarizes on prototypical agents with regards to their topicality that a “less technical way of expressing this fact is to say that people identify with and like to talk about things that act, move, control events, and have power” (151).

Agents in Ayeri frequently embody the properties quoted by both Fillmore and Payne in this regard, including Fillmore’s caveat. However, importantly, ‘agent’ in Ayeri is a macrorole that may be applied to, for instance, instruments, experiencers, and less typical actors as well, specifically, in absence of more prototypical candidates for agenthood in a sentence. It thus comes very close to a nominative, except that it does not need to be locus of the sentence’s topic—although agents very typically are topics, as Payne (1997: 151) goes on to note.³ Thus, the first NP after the verb in all of the following examples is treated as an agent; the agent is marked by the suffix ᐱᐅᐅ *-ang* for animate referents and the suffix ᐱᐅᐅᐅ *-reng* for inanimate referents; names and verbal topic agreement are marked by ᐱᐅᐅ *ang* and ᐱᐅᐅᐅ *eng*, respectively:

- (10) a. *Ang tinkaya Yan kunangley.*
 Ang tinka-ya \emptyset **Yan** kunang-ley
 AT open-3SG.M TOP Yan door-P.INAN
 ‘Yan opens the door.’
- b. *Le tinkaya ayonang kunang.*
 Le tinka-ya **ayon-ang** kunang- \emptyset
 PT open-3SG.M **man-A** door-TOP
 ‘The door is opened by a/the man’,
 or: ‘The door, a/the man opens it.’
- c. *Eng tinkāra tinkay kunangley.*
 Eng tinkā-ara **tinkay- \emptyset** kunang-ley
 AT.INAN open-3SG.INAN **key-TOP** door-P.INAN
 ‘The key opens the door.’
- d. *Tinkāra kunangreng.*
 Tinka-ara **kunang-reng**
 open-3SG.INAN **door-A.INAN**
 ‘The door opens.’

³ This is the main reason I spoke of ‘complications’ above: Ayeri’s notion of ‘subject’ is somewhat problematic due to topicalization, which is why I try to avoid complicating terminology by using ‘nominative’ for agent topics and ‘ergative’ for agent non-topics, and ‘accusative’ for patient non-topics and ‘absolutive’ for patient topics, respectively.

- e. *Sā tinkaya ang Yan kunangley yan.*
Sā tinka-ya ang Yan kunang-ley yan.Ø
 CAUT open-3SG.M A Yan door-P.INAN 3SG.M.TOP
 ‘They make Yan open a/the door’,
 or: ‘Because of them, Yan opens the door.’

In predicative constructions, the constituent which a quality is assigned to or about which a judgement is made is also assigned the agent case:

- (11) a. *Tado tinkayreng.*
Tado tinkay-reng
 old key-A.INAN
 ‘The key is old.’
 b. *Ang Yan nimpayās ban.*
Ang Yan nimpaya-as ban
 A Yan runner-P good
 ‘Yan is a good runner.’

With regards to constituents’ roles in ditransitive verb frames, donors are represented by agents in Ayeri as well, since they are the origin of whatever is conceptually passed on to the recipient party:

- (12) *Le ilya ang Yan tinkay yam Cānlay.*
Le il-ya ang Yan tinkay-Ø yam Cānlay
 PT give-3SG.M A Yan key-TOP DAT Cānlay
 ‘The key, Yan gives it to Cānlay.’

Patient

Patients are less of a definitional problem than agents, since in transitive sentences, they are very typically undergoers, that is, the constituent that is acted on, affected, or produced by the action expressed by the verb. The patient case is thus the one assigned by default to direct objects—but also to predicative nominals. In ditransitive sentences, the theme is represented by the patient. Animate patients are marked by 𑀅𑀲 -as, inanimate ones by 𑀅𑀲 -ley; for names and verbal topic agreement, the markers are 𑀅 sa and 𑀅 le, respectively:

- (13) a. *Ang silvye Briha sa Taryan.*
Ang silv-ye Ø Briha sa Taryan
 AT see-3SG.F TOP Briha P Taryan
 ‘Briha sees Taryan.’

- b. *Sa manye ang Briha Taryan.*
Sa man-ye ang Briha Ø Taryan
 PT greet-3SG.F A Briha TOP Taryan

‘Taryan is greeted by Briha’,
 or: ‘Taryan, Briha greets him.’

- (14) a. *Ang rimaye Briha kunangley.*
Ang rima-ye Ø Briha kunang-ley
 AT close-3SG.F TOP Briha door-P.INAN

‘Briha closes a/the door.’

- b. *Le rimaye ang Briha kunang.*
Le rima-ye ang Briha kunang-Ø
 PT.INAN close-3SG.F A Briha door-TOP

‘The door is closed by Briha’,
 or: ‘The door, Briha closes it.’

- (15) *Ang ilya Taryan koyaley yam Kandan.*
Ang il-ya Ø Taryan koya-ley yam Kandan
 AT give-3SG.M TOP Taryan book-P.INAN DAT Kandan

‘Taryan gives Kandan a book.’

As the translations of the examples above show, topicalizing the patient can be used to create an effect similar to English’s passive voice, except that the patient will not become marked by the agent case for logical reasons—this is a notable difference from the nominative. Even if the agent NP is omitted, the patient NP will not be changed to the agent case, since that would reverse the direction of action:

- (16) *Manya sa Taryan. ≠ Manya ang Taryan.*
Man-ya sa Taryan Man-ya ang Taryan
 greet-3SG.M P Taryan greet-3SG.M A Taryan

‘Taryan is greeted.’ ≠ ‘Taryan greets.’

This example shows that the case of the NP will not change, however, the verb will: it now agrees with the next argument in line, the patient NP. It will not do so, however, if the order of arguments is just scrambled, as exemplified by (17). This is to say that the verb does not simply agree with whichever NP follows it, even if it can be assumed that verb agreement in Ayeri developed along similar lines in-world, which will become especially apparent in the discussion of pronouns.⁴

⁴ Mismatches in agreement in connection to scrambling such as exemplified by (17b) are to be expected, however, since the brain can only handle so much information between the controller and the target of an agreement relationship. Corbett (2006), notes that with regards

- (17) a. *Sa manye Taryan ang Briha.*
 Sa man-ye Ø Taryan ang Briha
 PT greet-3SG.F TOP Taryan A Briha
 ↑
 person agreement
 ‘Taryan is greeted by Briha’,
 or: ‘Taryan, Briha greets him.’
- b. **Sa manya Taryan ang Briha.*
 Sa man-ya Ø Taryan ang Briha
 PT greet-3SG.M TOP Taryan A Briha
 ↑
 *person agreement

Besides being the default case for direct objects, the patient case is also assigned to predicative nominals, by analogy with transitive sentences and in spite of the likening nature of the construction:

- (18) *Ang Yan nimpayās ban.*
 Ang Yan nimpaya-as ban
 A Yan runner-P good
 ‘Yan is a good runner.’

Dative

The most typical use of the dative is for the recipient NP in a ditransitive clause; as such, it may be a recipient proper or the entity to whose benefit the action is carried out. A number of transitive verbs also use the dative for their object, for example, when it is the target of address. The dative can furthermore be used to mark movement toward a place. The case suffix for datives is ꠘꠗ -yam for both animate and inanimate entities. Names and verbal topic agreement are marked equally by ꠘꠗ *yam*. Verbs do not exhibit person agreement with dative NPs, since experiencers are treated as agents.

- (19) a. *Ang ilya Taryan koyaley ayonyam.*
 Ang il-ya Ø Taryan koya-ley ayon-yam
 AT give-3SG.M TOP Taryan book-P.INAN man-DAT
 ‘Taryan gives a book to the man.’

to agreement in NP conjuncts, “distant agreement is rare, and that agreement with the nearest noun phrase or agreement with all (resolution) is much more common” (62). If there were an extensive corpus of texts written by Ayeri speakers, it might be interesting to gather statistics on the number of words between target and controller in relation to the prevalence of agreement mismatches.

- b. *Ang ilya Taryan koyaley yam Kandan.*
 Ang il-ya Ø Taryan koya-ley **yam Kandan**
 AT give-3SG.M TOP Taryan book-P.INAN **DAT Kandan**
 ‘Taryan gives Kandan a book.’
- c. ***Yam** ilya ang Taryan koyaley ayon.*
Yam il-ya ang Taryan koya-ley **ayon-Ø**
DATT give-3SG.M A Taryan book-P.INAN **man-TOP**
 ‘The man is given a book by Taryan’,
 or: ‘The man, Taryan gives him a book.’

The three examples in (19) show the regular use of the dative as the case the recipient of the theme appears in. What distinguishes Ayeri from a pure split-S language is that all constituents can serve as topics, not just agents and patients with regards to their function as syntactic subjects. Thus, it is also possible for dative NPs to appear as topics—person agreement is unaffected by this, though. The following example shows the addressee of a speech act in the dative case; the message is treated as the theme which is passed on:

- (20) *Ang ningye māva ninganas **ganyam** yena.*
 Ang ning-ye māva-Ø ningan-as **gan-yam** yena
 AT tell-3SG.F mother-TOP story-P **child-DAT** 3SG.F.GEN
 ‘The mother tells her child a story.’

As mentioned above, the dative can also take on an allative meaning insofar as it marks the target of a motion, as displayed in (21a). As an extension of this means, the adpositional object may as well appear in the dative, since Ayeri cannot distinguish, for instance, ‘up’ from ‘to the top of’ with just the preposition, in this case *ṛṅṛ* *ling* ‘on top of’. With the adpositional object in the locative case (see below), the phrase in (21b) would imply that the man were literally going to the top of the temple, that is, possibly ending up on its roof.

- (21) a. *Ang nimpye lay **māvayam** yena.*
 Ang nimp-ye lay-Ø **māva-yam** yena
 AT run-3SG.F girl-TOP **mother-DAT** 3SG.F.GEN
 ‘The girl runs to her mother.’
- b. *Ang saraya ayon manga ling **natrangyam.***
 Ang sara-ya ayon-Ø manga ling **natrang-yam**
 AT go-3SG.M man-TOP DYN top **temple-DAT**
 ‘The man goes up to the temple.’

Genitive

The genitive is used to mark possessors; attributive genitives follow the possessee. It can also be used for ablative meanings, that is, to mark the place from which a motion originates, in analogy to the dative's allative use. The genitive is marked on common nouns with the suffix :᳚ -*na*. If a noun stem ends in a consonant, the marker becomes :᳚᳚ -*ena*, compare Figures 4.1–4.4 above. Names and verbal topic agreement are marked by ᳚ *na*. There is no animacy distinction in the genitive case.

- (22) a. *Pakur ledanang netuna nā.*
 Pakur ledan-ang netu-na nā
 sick friend-A brother-GEN ISG.GEN
 'My brother's friend is sick.'
- b. *Kopo dilengyereng ajānena.*
 Kopo dileng-ye-reng ajān-ena
 difficult rule-PL-A.INAN game-GEN
 'The rules of the game are difficult.'
- c. *Ang nakasyo tamo ibangya na Niyas.*
 Ang nakas-yo tamo-Ø ibang-ya na Niyas
 AT grow-3SG.N wheat-TOP field-LOC GEN Niyas
 'There is wheat growing on Niyas's field.'
- d. *Na nakasyo tamoang ibangya Niyas.*
 Na nakas-yo tamo-ang ibang-ya Ø Niyas
 GENT grow-3SG.N wheat-A field-LOC TOP Niyas
 'Regarding Niyas, there is wheat growing on his field.'

Futhermore, Ayeri does not make a distinction between alienable and inalienable possession, so that typically inalienable things such as body parts, relatives and family members, or personal items and tools are all treated as described above. Consider the following example for illustration:

- (23) *Ang puntaye māva nā mitrangas yena sembari yena.*
 Ang punta-ye māva-Ø nā mitrang-as yena semba-ri yena
 AT brush-3SG.F mother-TOP ISG.GEN hair-P 3SG.F.GEN comb-INS 3SG.F.GEN
 'My mother is brushing her hair with her comb.'

The above examples show the regular use of the genitive as a marker of possession. The following examples, on the other hand, show the genitive in its ablative function, first without qualification by a preposition, then with the preposition ᳚᳚ *avan* 'at the bottom of', which together with the genitive assumes the meaning 'down from':

- (24) a. *Ang sabaya Vetayan rimanena.*
 Ang saha-ya Ø Vetayan riman-ena
 AT come-3SG.M TOP Vetayan city-GEN
 ‘Vetayan comes from the city.’
- b. *Sabu manga avan mehirena, Niva!*
 Saha-u manga avan mehir-ena, Niva
 come-IMP DYN at.bottom tree-GEN, Niva
 ‘Come down from the tree, Niva!’

Locative

The locative marks basic locations, often the default that is associated with a verb. It is also the case in which adpositional objects normally appear, besides the special cases using the dative and the genitive mentioned above. Common nouns are marked by *-ya*;⁵ names and verbal topic agreement use the marker *u ya*. There is no difference made between animate and inanimate referents in the locative.

- (25) a. *Ang nedraya paray binya.*
 Ang nedra-ya paray-Ø hin-ya
 AT sit-3SG.M cat-TOP box-LOC
 ‘The cat sits in the box.’
- b. *Ang naraya Ajān ya Kaman.*
 Ang nara-ya Ø Ajān ya Kaman
 AT speak-3SG.M TOP Ajān LOC Kaman
 ‘Ajān speaks to Kaman.’
- c. *Ya mica ang Kaman Visamhinang.*
 Ya mit-ya ang Kaman Ø Visamhinang
 LOCT live-3SG.M A Kaman TOP Visamhinang
 ‘Kaman lives in Visamhinang’,
 or: ‘Visamhinang is where Kaman lives.’

The example sentences in (25) show locative NPs that are not further specified by adpositions so that the correct interpretation may be dependent on context and the experience of the addressee. Example (25a) is an instance of this circumstance, insofar as experience tells that cats like to sit inside boxes, so further specifying the position with the preposition *kong* ‘inside’ would be emphasizing that the cat is not sitting just anywhere, but really *inside* the box as opposed to on top of it, for instance. The following example has the cat sitting on top of the box:

⁵ Older texts still exhibit an allomorph *-ea*, used especially in combination with the plural suffix *-u*, giving *-yea*. The modern language uses *-ya*.

- (26) *Ang nedraya paray ling binya.*
 Ang nedra-ya paray-Ø ling hin-ya
 AT sit-3SG.M cat-TOP on.top box-LOC
 ‘The cat sits on the box.’

Ayeri also has a number of postpositions, which do not change marking on the adpositional object, however:

- (27) *Ang mican edaya tenyanya tan pesan.*
 Ang mit-yan edaya **tenyan-ya** tan pesan
 AT live-3PL.M here **death-LOC** 3PL.M.GEN until
 ‘They lived here until their death.’

Causative

The causative marks the cause or causer of an action, the instigator or the reason on behalf of which an agent is acting. It is thus similar to the agent case, though it does not replace it in Ayeri; verbs do not exhibit person agreement with causers even though their action logically supersedes or precedes that of the agent in the embedded event. Dixon (2000) writes that a “causer refers to someone or something (which can be an event or state) that initiates or controls the activity. This is the defining property of the syntactic–semantic function A (transitive subject)” (30). According to Comrie (1989: 176), the causee—the agent of the event controlled by the causer—normally takes the highest place in the hierarchy of syntactic constituents that is not already filled, in this case, by the causer. This observation, however, is complicated by Ayeri’s more or less semantics-based case marking as well as topicalization. In the following, I will give examples of nominal marking for cause as before; a discussion of the morphosyntax of Ayeri’s morphological causative constructions will be deferred to the section on valency-increasing operations.

Causers or causes are marked by *isa* for common nouns; names and verbal topic agreement use the marker *sā*. As stated above, verbs do not agree with causers even though they have agent-like semantics. There is no animacy distinction in the marking of causers.

- (28) a. *Ang rua sarāyn seyaranisa.*
 Ang rua sara=ayn.Ø **seyaran-isa**
 AT must leave=IPL.TOP **rain-CAUS**
 ‘We had to leave due to the rain.’
 b. *Ang yomāy edaya sā Apican.*
 Ang yoma=ay.Ø edaya **sā Apican**
 AT be=ISG.TOP here **CAUS Apican**
 ‘I am here because of Apican.’

- c. *Sā nimpvāng hakasley yan.*
Sā nimp=vāng hakas-ley yan.Ø
 CAUT run=2SG.A mile-P.INAN 3PL.M.TOP
 ‘You run a mile because of them’,
 or: ‘Due to them, you run a mile’,
 or: ‘They make you run a mile.’

Regarding the typological oddities mentioned above, example (28c) shows what happens in Ayeri with regards to the marking of causers. Essentially, the causer topic was grammaticalized to express a causation relationship.

Instrumental

The instrumental marks the means by which an action is carried out by an agent. This can be a tool as well as an animate being by whose help the action is brought about. The instrumental thus, in effect, marks secondary agents; verbs, however, never show person agreement with instrumental NPs. Common nouns are marked by *-ri* when ending in a vowel and with *-eri* when ending in a consonant; names and verbal topic agreement are marked by *ri*. With nouns ending in *-e*, as well as the plural marker *-ye*, there is variation regarding whether *-ri* or *-eri* is used, so that in the case of the plural marker both *-yeri* and *-yēri* occur. In passive-like constructions, it is not grammatical to reintroduce the agent as an instrumental; the agent simply remains in the clause in this case, though as a non-topic constituent.

- (29) a. *Ang visye Pila seygoley tibangeri yena.*
Ang vis-ye Ø Pila seygo-ley tihang-eri yena.
 AT cut-3SG.F TOP Pila apple-P.INAN knife-INS 3SG.F.GEN
 ‘Pila cuts an apple with her knife.’
- b. *Ang liboyya-ma badan nibanyeri (nibanyēri).*
Ang liha-oy-ya=ma badan-Ø nihan-ye-ri (nihan-ye-eri)
 AT earn-NEG-3SG.M=enough father-TOP nihan-PL-INS (nihan-PL-INS)
 ‘Father did not earn enough with his fruits.’
- c. *Ang lingya Mindan mebiras ri Kadijān.*
Ang ling-ya Ø Mindan mehir-as ri Kadijān.
 AT climb.up-3SG.M TOP Mindan tree-P INS Kadijān
 ‘Mindan climbs a tree with Kadijān’s help.’
- d. *Ri tavya gino ang Kan nimpur.*
Ri tav-ya gino ang Kan nimpur-Ø
 INST become-3SG.M drunk A Kan wine-TOP
 ‘Kan becomes drunk on the wine’,
 or: ‘The wine, Kan becomes drunk on it.’

The instrumental may also be used for comitative meanings where the instrumental NP describes an attribute of its antecedent, for example:

- (30) *Ang pegayo sinyā kasuley bariri nā?*
 Ang pega-yo sinyā-Ø kasu-ley **bari-ri** nā
 AT steal-3SG.N who-TOP basket-P.INAN **meat-INS** 1SG.GEN
 ‘Who stole my basket of meat?’

In this case, *bari* is marked as an instrumental since it is an attribute of sorts to *kasu*: the instrumental NP describes what its antecedent contains or entails more specifically: it is a basket *with* meat in it. Note, however, that this comitative use of the instrumental is different from mere accompaniment. Thus, it is not possible to say

- (31) **Ang sabaya Ajān ri Pila.*
 Ang saha-ya Ø Ajān **ri** Pila
 AT come-3SG.M TOP Ajān **INS** Pila

to express ‘Ajān comes (together) with Pila’. The sentence in (31) would instead imply that Pila helps Ajān to come, for example, because he has a sprained ankle and thus needs support to go places. To express accompaniment, instead, the preposition *kayvo* ‘with, along, beside’ has to be used; the prepositional object appears in the locative case:

- (32) *Ang sabaya Ajān kayvo ya Pila.*
 Ang saha-ya Ø Ajān **kayvo ya** Pila
 AT come-3SG.M TOP Ajān **with Loc** Pila
 ‘Ajān comes (together) with Pila.’

Theoretically, it should be possible as well to use the instrumental together with prepositions for some kind of prolocative meaning. The adposition would indicate the place *by way of* a motion is happening:

- (33) *Ang pukay manga luga labaneri.*
 Ang puk=ay.Ø manga luga **lahan-eri**
 AT jump=1SG.TOP DYN top **fence-INS**
 ‘I jump over the fence.’

This use of the instrumental is unattested in previous translations into Ayeri, however, but could be considered a stylistic alternative—in the case of the example above, to a construction with the word for ‘over’, *eyrarya*:

- (34) *Ang pukay manga eyrarya labanya.*
 Ang puk=ay.Ø manga eyrarya lahan-ya
 AT jump=ISG.TOP DYN over fence-LOC
 ‘I jump over the fence.’

A more literal translation of ᐱᐅᐅᐅᐅᐅᐅᐅᐅᐅᐅ *manga luga labaneri* is ‘by way of the top of the fence’, though without the verbosity of the English translation, as both ways to express the circumstance are about equally long in Ayeri.

Case-unmarked nouns

Case endings are applied to nouns in Ayeri only if the word is actually in a syntactic context where case can be applied. Thus, the unmarked form is the citation form, not the one declined for agent. This is the case when addressing people—one might speak of an unmarked vocative:

- (35) a. *Raypu, petāya!*
 Raypa-u, petāya
 stop-IMP, idiot
 ‘Stop it, you idiot!’
 b. *Sabu edaya, Diras!*
 Saha-u edaya, Diras
 come-IMP here, Diras
 ‘Come here, Diras!’

Imperative forms have underlying second-person agents, so both the ‘idiot’ in (35a) and Diras in (35b) would be the implied agents of their sentences, yet neither the noun nor the name are marked by the agent markers ᐱᐅᐅ -*ang* and ᐱᐅᐅ *ang*, respectively. Another case where nouns are not necessarily marked for case is attested in translations for the prefix ᐱᐅᐅ *ku-* ‘like, as though’ when the phrase acts as an adverb or an object complement:

- (36) a. ... *nay ang mya rankyon sitanyās ku-netu.*
 ... *nay ang mya rank=yon.Ø sitanya-as ku=netu*
 ... and AT be.supposed.to treat=3PL.N.TOP each.other-P like=brother
 ‘... and they shall treat each other like brothers.’⁶ (Becker 2011)
 b. ... *ang nunaya ku-vipin ...*
 ... *ang nuna=ya.Ø ku=vipin ...*
 ... AT fly=3SG.M.TOP like=bird ...
 ‘... he (would) fly like a bird ...’ (Becker 2012: 14)

Strikingly, in example (36a), ᲛᲗ᲏ *netu* ‘brother’ in ᲛᲗ᲏Თ᲏ *ku-netu* ‘like brothers’ is not even inflected for plural; its placement after the object is likely an effect of translation: adverbs have a strong tendency to appear right after the verb, and a position immediately to the right of the verb is attested for adjectival object predicates as well. In (36b), on the other hand, ᲛᲗ᲏Თ᲏Თ᲏ *ku-vipin* ‘like a bird’ is feasibly interpretable as an adverb, since it follows the verb and acts as a modifier to it, not as a complement.

Nouns may also be unmarked if they act as modifiers in a compound and the head is marked for the NP’s case and number, for instance:

- (37) *ralanyeri mapang*
 ralan-ye-ri mapang
 nail-PL-INS finger
 ‘with the fingernails’

And lastly and probably most importantly, nouns appear superficially unmarked if topicalized, since the topic marker is $-\emptyset$:

- (38) *Saru-nama, ang nupoyya veney aruno vās.*
 Sar-u=nama, ang nupa-oy-ya **veney- \emptyset** aruno vās
 go-IMP=just, AT hurt-NEG-3SG.M **dog-TOP** brown 2SG.P
 ‘Just go, the brown dog won’t hurt you.’

4.1.4 Prefixes on nouns

All of the nominal morphology we have so far dealt with in this section was suffixing. As mentioned in the previous section already (p. 67), however, there are also a number of prefixes that can be applied to nouns. I have just given two examples of the prefix ᲛᲗ᲏ *ku-* ‘like, as though’ above, but ᲛᲗ᲏ *ku-* applies not only to nouns themselves. In fact, it rather attaches to whole NPs, which makes it a clitic according to Klavans (1985: 117), and a special clitic in Zwicky’s terminology, since no corresponding full form exists in its place, comparable to the English possessive clitic *’s*, for instance (Zwicky 1977: 3, 13; 1985: 295; Zwicky and Pullum 1983: 510). To cite from the Ayeri translation of Kafka’s short story “Eine kaiserliche Botschaft” again:

⁶ The original English text this was translated from has “and should act towards one another in a spirit of brotherhood” (United Nations 1948: Article 1).

- (39) ... *saylingyāng kovaro naynay, ku-ranyāng palung.*
 ... sayling=yāng kovaro naynay, ku=ranya-ang palung
 ... progress=3SG.M.A easy also, like=nobody-A else

‘... he also got on easily, like nobody else.’

(Becker 2012: 12)

In this example, we can see ku- attaching to a properly inflected NP adjunct. The NP is case-marked for agent since it can be understood to refer to the verb sayling- ‘progress’ in the main clause, insofar ranyāng palung ‘nobody else’ can replace -yāng ‘he’ in the main clause.

Besides ku- , there are also the demonstrative prefixes da- ‘such’, eda- ‘this’, and ada- ‘that’, which have already been mentioned in the previous section as well. The demonstrative prefixes undergo coalescence with nouns beginning with $a-$, that is, they form phonological words with their hosts for all means and purposes. The demonstrative prefixes are special clitics as well, since no contemporary free form exists.

- (40) a. *da-nanga kāryo*
 da=nanga kāryo
 such=house big
 ‘such a big house’
 b. *edāyon nake*
 eda=ayon nake
 this=man tall
 ‘this tall man’
 c. *ada-envan alingo*
 ada=envan alingo
 that=woman clever
 ‘that clever woman’

Moreover, there is a prefix mə- in complementary distribution with the demonstrative prefixes, which adds a meaning along the lines of ‘just any’, ‘whatsoever’, ‘some’ to the noun. Note that this prefix is distinct from the morpheme indicating an inspecific quantity, -aril ‘some’.

- (41) a. *Ang lampyo mə-veney kayvo kirinya.*
 Ang lamp-yo mə=veney-Ø kayvo kirin-ya
 AT walk-3SG.N some=dog-TOP along street-LOC
 ‘Some dog is walking along the street.’
 b. *Ang noyan mēntānley pegamayayam.*
 Ang no=yan mə=entān-ley pegamaya-yam
 AT want=3SG.M.TOP some=punishment-P.INAN thief-DAT
 ‘They demanded some kind of punishment for the thief.’

4.1.5 Compounding

With regards to the classification of compounds, Bauer (2001) gives some helpful typological guidelines. Besides the compound types recognized by Sanskrit grammarians—endocentric (*tatpuruṣa*), coordinative (*dvandva*), adjectival-endocentric (*karmadhāraya*), and exocentric (*bahuvrīhi*)—he also adds synthetic compounds, which Sanskrit did not have (697). Overall, he finds that determinative, or endocentric, compounds are the most common ones in the languages of the world (697), especially if the head refers to a location or source of sorts (702).

Gaeta (2008), then, adds to Bauer’s research, based on a larger sample of grammars surveyed, that compounds for the largest part correlate with the constituent order of the language, both regarding the order of verb and object and that of noun and genitive (129–133). Mismatches in headedness occur, but appear to constitute the minority of cases and may often be explained through historical changes in syntax; he discerns for one that “morphology is not autonomous from syntax” (135), and that secondly, “[s]yntax seems to be the motor of change, which may be then reflected in compounds” (135), and that thirdly, lexical conservatism causes atavisms to linger on, reflecting the syntax of earlier stages of the language (138–139).

For the purpose of gaining at least a little insight into which types of compounds Ayeri allows—besides endocentric compounds—I conducted a small (non-exhaustive) survey based on 130 compounds from the Ayeri dictionary (Becker 2016: Dictionary); Table 4.1 shows the various compound classes and the number of words for each. ‘Harmonic’ and ‘disharmonic’, respectively, refer to the order of elements; the order is ‘harmonic’ if it is following the normal constituent order of the language and ‘disharmonic’ if it is at odds with it (Gaeta 2008).

Table 4.1: Compounds in the Ayeri dictionary (Becker 2016) and their classification (n = 130)

Type	Harmonic		Disharmonic		Total	
Endocentric (N + N)	67	51.54%	2	1.54%	69	53.08%
Endocentric (N + Adj)	18	13.85%	4	3.08%	22	16.92%
Synthetic (V + N)	16	12.31%	4	3.08%	20	15.38%
Coordinative (N + N)	9	6.92%	—	—	9	6.92%
Exocentric (N + N)	1	0.77%	3	2.31%	4	3.08%
Unclear	6	4.62%	—	—	6	4.62%
Total	117	90.00%	13	10.00%	130	100%

Unsurprisingly, the largest number of compound nouns in the sample were endocentric compounds of the regular kind, which means that, just like genitive attributes follow nouns, noun compounds are headed left. Especially compounds

with adjectives are interesting insofar as this is also the normal order for free adjectives, so to illustrate, some tests will be necessary to show that these adjectives form a unit with the head noun and are unable to undergo comparison, for instance. Synthetic compounds exist in Ayeri and produce nouns. These are compounds in which “the modifying element in the compound is (usually) interpreted as an argument of the verb from which the head is derived” (Bauer 2001: 701). There are also a number of coordinative compounds; this group, however, is lexicalized and not productive. Exocentric compounds constitute the minority of the sample. In the following I will give examples for each type.

It needs to be noted that unlike Germanic languages, Ayeri does not allow compounds of arbitrary length to be strung together, like in the following ridiculous but no less real example from (former) German legislation (see, for instance, *Süd-deutsche Zeitung* 2013):

- (42) *Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz* [German]
 Rind-fleisch-etikettierung-s-überwachung-s-aufgabe-n-übertragung-s-gesetz
 cow-meat-labeling-LNK-surveillance-LNK-duty-LNK-delegation-LNK-law
 ‘Law on the delegation of duties in the surveillance of beef labeling’

In stark contrast, Ayeri allows only two elements in compounds. Furthermore, this section on compounds is located within the section on nouns because Ayeri almost only possesses compounds involving nouns, and the majority of these also results in a noun.

Endocentric compounds

To start with the largest group, endocentric/*tatpuruṣa* compounds, the bulk of these compounds combines two nouns, one of which is the head which is modified by a dependent noun. As Ayeri exhibits a rather strict head-first word order, it comes as no surprise, according to Gaeta (2008), that most of these compounds follow this order strictly: the second noun modifies the first, which is opposite of how English, for instance, typically operates:

- (43) a. *betaynimpur* ‘grape’ ← *betay* ‘berry’ + *nimpur* ‘wine’
 b. *karirayan* ‘vertigo’ ← *kar* ‘fear’ + *irayan* ‘height’⁷
 c. *pikunanding* ‘mustache’ ← *piku* ‘beard’ + *nanding* ‘lips’
 d. *tapayperin* ‘sunblind’ ← *tapay* ‘screen’ + *perin* ‘sun’

The example words in (43) show that the relationships between the modifier and the head are various: a grape is a berry *used* to make wine from (compare

⁷ *irayan*, however, is a transparent nominalization of *iray* ‘high’.

Bauer 2001: 702); vertigo is the fear *of* height; a mustache is a beard *located* over the lips (702); and a sunblind is a screen *against* the sun. Bauer (2001) mentions that “there may be special morphophonemic processes which apply between the elements of compounds”, such as “phonological merger[s] between the elements of the compound” (695). This also occasionally happens in Ayeri, as the next few example words show:

- (44) a. ၼာၼာၼ် *avararan* ‘wetland’
 ← ၼာၼ် *avan* ‘ground’ + ၼာၼ် *raro* ‘wet’ + ၼ် *-an* (NMLZ)
- b. မိမိမိမိ *mehimitrang* ‘fiber tree’
 ← မိမိ *mehir* ‘tree’ + မိမိ *mitrang* ‘hair, fiber’
- c. နိဂံၼ် *ningampinam* ‘bedtime story’
 ← နိဂံၼ် *ningan* ‘story’ + နိဂံၼ် *pinam* ‘bed’
- d. ပုၼ်းမိမိ *padilamican* ‘gravitational force’
 ← ပုၼ်း *padilan* ‘attraction’ + မိမိ *mican* ‘force, power’

There is a modicum of alteration happening in all of the heads of the example words in (44), mostly nasals assimilating to the stop or nasal which the modifier begins with (/n/ + /p/ → /mp/, /n/ + /m/ → /m/), though ၼာၼာၼ် *avararan* and မိမိမိမိ *mehimitrang* even delete whole coda segments. Bauer (2001: 703) notes that very commonly, genitive and plural markers may form linking elements, though he also gives examples of languages which allow other case markers on the modifying element in languages with head-right order; individual languages may allow even more case inflection. However, this appears not to happen in Ayeri. The only element that comes up time and again in between the two halves of compounds is the nominalizer ၼ် *-an*, which signifies that the head is being formed by a nominalized root, such as in ပုၼ်းမိမိ *padilamican*, where ပုၼ်း *padilan* ‘attraction’ is a nominalization of ပုၼ်း *padil-* ‘attract’, or in နိဂံၼ် *ningampinam*, where နိဂံၼ် *ningan* ‘story’ is derived from the verb နိဂံၼ် *ning-* ‘tell’. However, since Ayeri is head-first and possessive phrases are dependent marking, genitive or other case marking would be expected on the second element, not the first. Case marking on a compound, however, does not inflect just the modifier, but the whole NP:

- (45) *Ang ningya sipikanena koyabahisena.*
Ang ning-ya sipik-an-ena koyabahis-ena
 AT talk-3SG.M.TOP keep-NMLZ-GEN book.day-GEN
 ‘He talks about keeping a journal.’

နိဂံၼ် *koyabahisena* in this example is not to be interpreted as ‘book of day(s)’ but as ‘of a day-book’. Inflection between the parts of a compound can happen nonetheless, though. In compounds which are formed ad hoc or which are otherwise transparent in their composition, inflection often is deferred to the noun head

noun instead of the edge of the compound as a whole; the modifier is treated as an adjunct in this case, and stays uninflected:

- (46) *Sa trayeng tipin ralanyeri mapang yena.*
 Sa tra=yeng tipin-Ø ralan-ye-ri mapang yena
 PT scratch=3SG.F.A itch-TOP nail-PL-INS finger 3SG.F.GEN

‘The itch, she scratches it with her fingernails.’

Besides noun modifiers, there are also compounds where the modifier is an adjective. In classical Sanskrit terminology, this type is called *karmadhāraya* (Bauer 2001: 698–699).⁸ Examples in Ayeri include:

- (47) a. *ခဲကပ်ကပ်ကပ် kardangiray* ‘university’ ← *ခဲကပ်ကပ် kardang* ‘school’ + *ခဲကပ် iray* ‘high’
 b. *မာရနဟိ marashati* ‘witticism’ ← *မာရန maras* ‘phrase’ + *ဟိ hati* ‘pithy’
 c. *ခဲကပ်ကပ်ကပ် silvanikan* ‘overview’ ← *ခဲကပ်ကပ် silvan* ‘view’ + *ခဲကပ်ကပ် ikan* ‘whole’
 d. *ဝိပိမာကာရီ vipimakārya* ‘crow’ ← *ဝိပိမာ vipin* ‘bird’ + *မာကာရီ makārya* ‘black’

In all of these cases, the adjective forms a unified lexeme with the head noun, hence it is not comparable, for example:

- (48) a. **kardangiray-eng kardangiray-vā*
kardang-iray=eng kardang-iray=vā
school-high=COMP school-high=SUPL
 ‘*higher-school’ ‘*highest-school’
 b. **marashati-eng *marashati-vā*
maras-hati=eng maras-hati=vā
phrase-pithy=COMP phrase-pithy=SUPL
 ‘*pithier-phrase’ ‘*pithiest-phrase’

In fact, it is possible to form *ခဲကပ်ကပ်ကပ်ကပ် kardangiray vā* and *မာရနဟိဟိ marashati vā*, but they mean ‘most universities’ and ‘most witticisms’, respectively; *ခဲကပ် -eng* ‘rather’ as a quantifier does not combine with nouns. Since the meaning composed from noun–adjective compounds is often idiomatic, they also cannot be divided as shown above in (46), since a *ခဲကပ်ကပ်ကပ် kardangiray* ‘university’ is not a *ခဲကပ်ကပ် kardang* ‘school’ which is *ခဲကပ် iray* ‘high’ in the literal sense, but a school of the highest tier. *ခဲကပ်ကပ်ကပ်ကပ် kardangena iray* (school-GEN high), then, can only be interpreted in the literal sense, ‘of the high school’, but not as ‘of the university’, which thus can only be *ခဲကပ်ကပ်ကပ်ကပ် kardangirayena*.

⁸ Bauer also mentions that appositional compounds like *maid-servant*, *woman doctor* and *fighter-bomber* are counted in this category (Bauer 2001: 699). Ayeri, however, does not possess such formations in particular.

In the sample, there were also a few compounds I categorized as noun–noun combinations, which look as though they violate head-first order. All of these involve ṣṭang *sitang* ‘self’ as a modifier:

- (49) a. ṣṭangṭentan *sitanglentan* ‘vowel’ ← ṣṭang *sitang* ‘self’ + ṭentan *lentan* ‘sound’
 b. ṣṭangṭaronān *sitangparonān* ‘self-confidence’ ← ṣṭang *sitang* ‘self’ + ṭaronān *paronān* ‘faith’
 c. ṣṭangṭenyan *sitangtenyan* ‘suicide’ ← ṣṭang *sitang* ‘self’ + ṭenyan *tenyan* ‘death’

ṣṭang *sitang* does not exist as a noun by itself in Ayeri, the word for ‘self’ is its nominalization ṣṭangan *sitangan*. Nonetheless, it looks like it could have plausibly been a noun once. However, this noun may have been grammaticalized into a reflexive morpheme of a more general kind, which in turn birthed the form ṣṭangan *sitangan* as a renovation.⁹ The reflexive ṣṭang *sitang* is used—as we have seen in the previous chapter—as a prefix, so there are two ways to interpret these formations: first, ṣṭang *sitang* may be the reflexive prefix here and thus the compound follows the normal syntactic order; or second, the order of elements is reversed and thus may reflect an earlier stage of Ayeri where ṣṭang *sitang* was still a noun and modifiers could still appear in front of their heads, at least optionally so (Gaeta 2008: 133–137).

There are a number of genuinely reversed endocentric compounds as well, however, in which the modifier comes first and the head last. Since there are only a few of these, I will give all of them in the following example:

- (50) a. ṭaripata *baripata* ‘ground meat’ ← ṭari *bari* ‘meat’ + ṭata *pata* ‘mash’
 b. ṭayvolentan *kayvolentan* ‘consonant’ ← ṭayvo *kayvo* ‘with’ + ṭentan *lentan* ‘sound’
 c. māṭaganeng *māvaganeng* ‘mother’s siblings’ ← māṭa *māva* ‘mother’ + ṭaganeng *ganeng* ‘siblings’
 d. ṭatinanding *matinanding* ‘labia’ ← ṭatikan *matikan* ‘hot’ + ṭanding *nanding* ‘lips’
 e. ṭuyavirang *muyavirang* ‘brass’ ← ṭuya *muya* ‘false’ + ṭavirang *avirang* ‘gold’
 f. ṭonisayang *tonisayang* ‘self-assured’ ← ṭonisa *tonisa* ‘assured’ + ṭitangan *sitangan* ‘self’

Given the discussion of ṣṭang *sitang* above, one word among the examples above that is not too clear is ṭonisayang *tonisayang*, which appears to contain a deviant form

⁹ A little bit of language history would certainly simplify things here and lend them credence. Let us simply assume that ṣṭang *sitang* used to be a noun meaning something like ‘self’ at a previous stage of Ayeri and was repurposed as a reflexive prefix. Lehmann (2015) quotes a few examples of what he calls ‘autophoric’ nouns that came to be used as reflexive pronouns in their respective language: “Typical examples are Sanskrit *tan* ‘body, person’ and *ātmán* ‘breath, soul’, Buginese *elena* ‘body’, Okinawan *dūna* ‘body’, !Xu *l’esi* ‘body’, Basque *burua* ‘head’, Abkhaz *a-xə* ‘the head’. In their respective languages, all these nouns are translation equivalents of English *self*” (45–46). Thus, it would not be out of line at all to assume such a grammaticalization path for Ayeri as well.

of either *sitang* or *sitangan*, which is preceded by the adjective *tonisa* ‘assured, ascertained’.

All of the previously mentioned compounds involving nominal elements formed nouns, though, there are also a few denominal compounds in the sample. This process is not productive, however, and interestingly, only noun–adjective combinations appear in this group:

- (51) a. *mirampaluy* ‘otherwise’
 ← *miran* ‘way’ + *palung* ‘different’
 b. *padabanya* ‘insane’
 ← *padang* ‘mind’ + *banaya* ‘sick’
 c. *tenkarisa-* ‘be frightened to death’
 ← *ten* ‘life’ + *karisa* ‘frightened’

mirampaluy is an adverb, the modifier probably a mangling of *palung*; *padabanya* is an adjective meaning ‘insane’ rather than the expected ‘insanity’ (instead: *padabanyān*); and *tenkarisa-* acts as a verb, possibly from conversion or reinterpretation, since the suffix *-isa* also forms morphological causatives of a number of verbs. Besides these irregularities, there is also at least one noun compound which uses a postposition as an adjectival modifier:

- (52) *silvankayvay* ‘blindness’ ← *silvan* ‘sight’ + *kayvay* ‘without’

This compound must be derived from the phrase *silvanya kayvay* ‘without sight’ (see-NMLZ-LOC without), though here as well, the word roots are simply juxtaposed, as noted above is the common way to form compounds in Ayeri.

Synthetic compounds

According to Bauer (2001), (semi-)synthetic compounds, or verbal(-nexus) compounds, are compounds that have “been variously defined as being based on word-groups or syntactic constructions (Botha 1984: 2), or as compounds whose head elements are derived from verbs (Lieber 1994: 3607)” (Bauer 2001: 701). Examples of this type in English would include *truck-driver*, *peace-keeping*, and *home-made*. He mentions also that synthetic compounds have been mainly discussed with regards to Germanic languages, but that according to Lieber (1994: 3608), the phenomenon is much more widespread. Ayeri possesses compounds like this as well, and the regular case again follows the constituent order, here that of verbs and nouns: Ayeri is a VO language, and thus the verb as the head of the compound is usually found on the left side with its nominal modifier following it (Gaeta 2008: 129–133):

- (53) a. ၼံၼ်ၵၢၼ် anlāgonan ‘pronunciation’ ← ၼံၼ်ၵ: anl- ‘bring’ + ၼံၼ်ၵ anagonan ‘outside’
 b. ဘၢၼ်ၵ်း napakaron ‘acid’ ← ဘၢ: nap- ‘burn’ + ၼ်ၵ်း karon ‘water’
 c. ဘၢၼ်ၵ်း napaperin ‘sunburn’ ← ဘၢ: nap- ‘burn’ + ၼ်ၵ်း perin ‘sun’
 d. ၼံၼ်ၵၢၼ် telbasasān ‘waysign’ ← ၼံၼ်ၵ: telba- ‘show’ + ၼံၼ်ၵ sasān ‘way’

The relations between the verb and the noun are various again, that is, the nominal modifier is not simply the direct object of the verb: to pronounce something means to bring it *to* the outside; a sunburn is a burn *caused by* the sun; and a waysign shows the way (ၼံၼ်ၵ sasān is the object here). Even though ၼ်ၵ်း karon may serve as an agent (or a causer) of the burning effect of acid (similarly for ဘၢၼ်ၵ်း napaperin ‘sunburn’), the verb-first order is justified here as well, since verbs always go first in Ayeri sentences, and any other NPs, whether actor or undergoer, are following.

Just as with endocentric compounds, there are a number of seeming exceptions to the verb-first order of synthetic compounds as well. These are just as far and few between, however, and whether they should all be counted as noun-verb combinations is also questionable as they appear to all be formed with nominalized verbs. The verbal element may thus be only indirectly verbal for the purposes of compounding. If interpreted as noun-noun combinations, the nominal first element would reasonably form the head again for some of the below example words.

- (54) a. ၼံၼ်ၵၢၼ် maripuntayam ‘spread’
 ← ၼံၼ်ၵ marin ‘surface’ + ၼံၼ်ၵ punta- ‘stroke’ + :ပဃ -yam (DAT)
 b. ၼံၼ်ၵၢၼ် sasanlekān ‘labyrinth’
 ← ၼံၼ်ၵ sasān ‘way’ + ၼံၼ်ၵ leka- ‘guess’ + :ၼံၼ် -an (NMLZ)
 c. ၼံၼ်ၵၢၼ် selangnunān ‘plane’
 ← ၼံၼ်ၵ selang ‘machine’ + ၼံၼ်ၵ nuna- ‘fly’ + :ၼံၼ် -an (NMLZ)
 d. ၼံၼ်ၵၢၼ် sinturān ‘radio’
 ← ၼံၼ်ၵ sinto ‘wave’ + ၼံၼ်ၵ tura- ‘send’ + :ၼံၼ် -an (NMLZ)

ၼံၼ်ၵၢၼ် maripuntayam is special in that it contains the dative suffix :ပဃ -yam which is lexicalized as part of the word: something made or intended for spreading on a surface. A few more such verbal derivations can be found, though not compounds, among others:

- (55) a. ၼံၼ်ၵၢၼ် grenyam ‘extremity’ ← ၼံၼ်ၵ gren- ‘reach out’
 b. ၼံၼ်ၵၢၼ် lugayam ‘password’ ← ၼံၼ်ၵ luga- ‘go through’
 c. ၼံၼ်ၵၢၼ် sabayam ‘future’ ← ၼံၼ်ၵ saba- ‘come’

There is also ၼံၼ်ၵ maripunta- ‘spread over’ as the verb corresponding to ၼံၼ်ၵၢၼ် maripuntayam, though its meaning is less specific. Here as well, however, the verbal part is last instead of first. For the other example words (54b–d), an interpretation

of the second part as a deverbal noun is possible: a labyrinth as a way or path which requires guessing, a plane a machine for flight, and radio as a sending of waves. In the latter case, *နီတုရ်* *shinturān*, however, the head is still on the wrong side even if one interprets all of the above examples as noun–noun compounds with a deverbal element.

Coordinative compounds

Coordinative compounds are a very small group among the sample drawn from the dictionary, and not a very productive one. Bauer (2001) defines this class as having “two or more words in a coordinate relationship, such that the entity denoted is the totality of the entities denoted by each of the elements” (699). He cautions that they are very easy to confuse with appositional (also *karmadhāraya*) compounds in that both types of compound allow inserting an *and* between both elements. The following nominal coordinative compounds are included in the dictionary sample:

- (56) a. *ဘုံမာ* *bāmā* ‘mom-and-dad’ ← *ဘုံ* *bā(bā)* ‘dad’ + *မာ* *mā(mā)* ‘mom’
 b. *ပုရ်ဉာ* *pruynapay* ‘seasoning’ ← *ပုရ်* *pruy* ‘salt’ + *ဉာ* *napay* ‘pepper’
 c. *နာပုယ* *sapayyila* ‘hands-and-feet’ ← *နာ* *sapay* ‘hand’ + *ပုယ* *yila* ‘foot’
 d. *နီမာ* *simileno* ‘horizon’ ← *နီ* *simil* ‘country’ + *မာ* *leno* ‘sky’
 e. *နီမာရုဂ* *sitemrugon* ‘thunderstorm’ ← *နီမာ* *sitem* ‘lightning’ + *ရုဂ* *rugon* ‘thunder’
 f. *နီမာ* *vekamdekey* ‘dishes’ ← *နီမာ* *vekam* ‘plate’ + *နီမာ* *dekey* ‘fork’

None of the two elements recognizably forms the head in these examples, but both elements are typical exponents of the thing the compound signifies. Bauer (2001: 699) mentions that coordinative adjective compounds are rare, or at least rarely documented in the grammars he surveyed. In the sample I took, only the following compound is included, which forms a noun from the combination of two adjectives, insofar it is relevant to this section even though the component parts are not nouns:

- (57) *မာဂါ* *makagisu* ‘twilight’ ← *မာ* *maka* ‘light’ + *ဂါ* *gisu* ‘dark’

The sample also includes the following two words, however, which are neither made up from nouns, nor do they form a noun in combination. Instead, they are technically verbs combining to form directional adverbs and have been exceptionally included here for completeness:

- (58) a. *မာဂါ* *mangasaba* ‘towards’ ← *မာ* *manga-* ‘move’ + *ဂါ* *saba-* ‘come’
 b. *မာဂါ* *mangasara* ‘away’ ← *မာ* *manga-* ‘move’ + *ဂါ* *sara-* ‘go’

Exocentric compounds

In exocentric compounds, the modifier is not a hyponym of its head (700), which means that the modifier is not describing a property that more closely determines its head. So while a *dog kennel* is a type of kennel made for dogs, the head of an *egghead* is neither for eggs, nor containing eggs, nor made of eggs; instead, it refers to an egg-shaped skull metaphorically. And while a *bluecollar* may wear a blue shirt professionally, the referent it signifies is not a type of collar, but the relationship is metonymical in that the blue collar is part of the guise of the signified entity as a whole. The sample from the Ayeri dictionary contains a few compounds of this kind as well, though again, it is not a very productive group:

- (59) a. ၼ်းၼ်းၼ်း *avanyonang* ‘artery’ ← ၼ်းၼ်း *avan* ‘bottom, down’ + ပိၼ်း *yonang* ‘stream’
 b. ဃုၼ်းၼ်း *baytandevo* ‘headache’ ← ဃုၼ်း *baytang* ‘blood’ + ပိၼ်း *devo* ‘head’
 c. ၼ်းၼ်းၼ်း *linyongang* ‘vein’ ← ၼ်းၼ်း *ling* ‘top, up’ + ပိၼ်း *yonang* ‘steam’
 d. ၼ်းၼ်းၼ်း *sindaynanga* ‘address’ ← ၼ်းၼ်း *sinday* ‘number’ + ၼ်း *nanga* ‘house’

What is striking here is that only one out of four examples shows the expected head-left order: ၼ်းၼ်းၼ်း *sindaynanga*. The other three examples all have the head component on the right side, preceded by a modifier. However, what all of these have in common, is that they are only metaphorically or metonymically describing the thing they signify: veins and arteries are not literally streams going up or down (they are a kind of stream flowing in different directions, however, so these are probably on the borderline between exocentric and endocentric); a headache is related to the head, but has not directly to do with being made of or containing blood (the rationale behind this being a superstition that you have too much blood in your head, which is said to cause the pain); and a house number may be part of an address, but is in a *pars pro toto* relationship to it.

A few mysterious cases

The following words from my sample were either undeterminable as to their composition due to parts of the word not being clear regarding one of their constituent parts, either because I tweaked the constituent so much as to not be readily recognizable anymore, or because I forgot to make an entry in the dictionary, or even deleted or changed that. The words in question are the following:

- (60) a. ဃုၼ်းၼ်း *batangiman* ‘mosquito’ ← ဃုၼ်း *baytang* ‘blood’ + ?
 b. ၼ်းၼ်းၼ်း *kirinalang* ‘avenue’ ← ၼ်းၼ်း *kirin* ‘street’ + ?
 c. ၼ်းၼ်းၼ်း *ningambakar* ‘telltale’ ← ၼ်းၼ်း *ningan* ‘story’ + ?
 d. ၼ်းၼ်းၼ်း *ragayesuy* ‘grid’ ← ၼ်းၼ်း *ragan* ‘line’ + ?

- e. ၵံးရံး ၵံးရံး *teraymino* ‘melancholic’ ← ? + ၵံးရံး *mino* ‘happy’
 f. ၵံးရံးရံး *vetaysano* ‘fare’ ← ? + ၵံးရံး *sasān* (earlier ၵံးရံး *sasano*) ‘way’

For all of the components represented by a question mark, there is no corresponding dictionary entry. At least in ၵံးရံးရံး *baytangiman*, the *ၵံးရံး **iman* part looks as though it could be a noun due to the ၵံးရံး *-an* nominalizer suffix. *ၵံးရံး **teray* in ၵံးရံးရံး *teraymino* might also be an adjective supposed to mean ‘sad’ (which would make it an adjectival coordinative compound), although the dictionary entry for that is ၵံးရံး *giday*. Even though parts of all these words are unclear, they all seem to follow the correct syntactic order, judging by those parts that are identifiable. And even in the case of ၵံးရံးရံး *vetaysano*, which is missing the first part, it can be reasonably assumed that the identifiable part, *ၵံးရံး **sano*, is the modifier, and *ၵံးရံး *vetay* may have once been intended to mean ‘money’ or ‘fee’ or something along these lines.

With the exception of ၵံးရံးရံး *ningambakar*, all of the mystery words were entered into the dictionary in 2006. Digging through old archives and translations, I could determine at least that *ၵံးရံး **bakar* was once intended to mean ‘lie’, and *ၵံးရံး **teray* was indeed meant to mean ‘sad’.

4.1.6 Reduplication

Wiltshire and Marantz (2000) write that it has been suggested that reduplication serves an iconic function, “with the repetition of phonological material indicating a repetition or intensity in the semantics” (561), so with regards to nouns it mainly serves to indicate plurality of various kinds. However, they find that in fact, reduplication serves all kinds of functions, also ones without iconic meanings, and mention Agta, an Austronesian language of the Philippines, which uses reduplication to form diminutives (Healey 1960: 6–9). As we have seen in subsection 3.2.3 above, so does Ayeri, and it is justified in doing so since there is real-world evidence for this use of reduplication. Examples for diminutive reduplication in Ayeri include:

- (61) a. ၵံးရံး *limu* ‘shirt’ → ၵံးရံးရံး *limu-limu* ‘little shirt’
 b. ၵံးရံး *nanga* ‘house’ → ၵံးရံးရံး *nanga-nanga* ‘little house’
 c. ၵံးရံး *sapay* ‘hand’ → ၵံးရံးရံး *sapay-sapay* ‘little hand’
 d. ၵံးရံး *veney* ‘dog’ → ၵံးရံးရံး *veney-veney* ‘little dog’

Diminutive reduplication involves full stem reduplication in Ayeri. Besides the productive use of reduplication for diminutive marking, there are a number of diminutive formations which have been lexicalized, such as in the following examples:

- (62) a. ၼ် ၼ် *agu* ‘chicken’ → ၼ်:ၼ်:ၼ် *agu-agu* ‘chick’
 b. ၼ် *gan* ‘child’ → ၼ်:ၼ် *gan-gan* ‘grandchild’
 c. ၼ် *pasing* ‘tube’ → ၼ်:ၼ် *pasing-pasing* ‘straw’
 d. ၼ် *poyu* ‘cheek; bacon’ → ၼ်:ၼ် *poyu-poyu* ‘butt’

There are also at least two documented cases where the reduplicated root is not a noun, but the reduplication results in a noun:

- (63) a. ၼ် *kusang* ‘double (adj.)’ → ၼ်:ၼ် *kusang-kusang* ‘model’
 b. ၼ် *veh-* ‘build’ → ၼ်:ၼ် *veha-veha* ‘tinkering’

Reduplicated nouns behave like regular nouns with regards to inflection, that is, they receive prefixes and suffixes just like the simplexes from which they are derived:

- (64) *Puco mino veney-veneyang.*
Puk-yo mino veney~veney-ang
 jump-3SG.N happily DIM~dog-A

‘The little dog is jumping happily.’

In this example, the reduplicated noun ၼ်:ၼ် *veney-veney* is marked as an agent in that the agent suffix ၼ် *-ang* is appended to the noun as a unit *after* reduplicating the noun stem. In other words, the following formation in which the root is reduplicated along with its declension suffix is ungrammatical for the purpose of forming a diminutive:

- (65) *ၼ်:ၼ် **veneyang-veneyang*

Likewise, the reduplicated form is not treated in the way an endocentric compound would be, so case and plural marking cannot be appended to the first element:

- (66) *ၼ်:ၼ် **veneyang veney*

While ordinary nouns undergo full reduplication to form a diminutive, in compounds, only the head is reduplicated, unless the compound is strongly lexicalized or has an idiomatic meaning going beyond that of its components. The following example shows the simple case of a transparent endocentric compound:

- (67) *Ya yomayo mehir-mehirang seygo veno kay pang nanga nana.*
Ya yoma-yo mehir~mehir-ang seygo veno kay pang nanga-Ø nana
 LOCT be-3SG.N DIM~tree-A apple pretty three back house-TOP IPL.GEN

‘There are three pretty little apple trees behind our house.’

In this example, being endearing or otherwise small is treated as a property of the head, မဲဃ် *mehir* ‘tree’, not of the whole compound မဲဃ်ဒုန် *mehirseygo* ‘apple tree’, or the dependent, ဒုန် *seygo* ‘apple’—after all, an apple tree which is small is rather a small tree with apples on it than a tree with small apples. The avoidance of the fully reduplicated form မဲဃ်ဒုန်ဒုန် *mehirseygo-mehirseygo* is probably related to the notion of economy of expression.

4.1.7 Nominalization

Some accidental ways of deriving nouns have been mentioned above, for instance, some reduplicated non-nominal roots like နွန် *kusang* ‘double’ or ဖဲဃ် *veha* ‘build’ may form nouns. However, Ayeri also has some dedicated morphology to derive nouns from other parts of speech. The most common and highly productive way to derive a noun, is the suffix -an. The examples in (68) illustrate some derivations from verbs, and (69) shows derivations from adjectives to nouns. As နွန် *kuban* ‘oar’ shows, the nominalization may have an idiomatic meaning.

- (68) a. ခာဏ် *balang* ‘search (v.)’ → ခာဏ်ဒ် *balangan* ‘search (n.)’
 b. နွန် *kub* ‘row’ → နွန်ဒ် *kuban* ‘oar’
 c. ဂိန် *rig* ‘draw’ → ဂိန်ဒ် *rigan* ‘drawing’
 d. ဖဲဃ် *veh* ‘build’ → ဖဲဃ်ဒ် *vehan* ‘building’
- (69) a. ပါဏ် *apitu* ‘pure’ → ပါဏ်ဒ် *apituan* ‘purity’
 b. ဂိာ *gira* ‘urgent’ → ဂိာဒ် *girān* ‘hurry’
 c. ပန် *pakis* ‘serious’ → ပန်ဒ် *pakisan* ‘seriousness’
 d. ဝာ *vapa* ‘skillful’ → ဝာဒ် *vapan* ‘skill’

Occasionally, it may even happen that a noun is derived from a noun with a related but sometimes more basic meaning using the nominalizer -an. This process, however, is not productive, so compared to deverbalization and deadjectivization, examples of this derivation strategy are few.

- (70) a. ခာဏ် *ajam* ‘toy’ → ခာဏ်ဒ် *ajaman* ‘game’
 b. နွန် *kelang* ‘chain’ → နွန်ဒ် *kelangan* ‘connection’
 c. ဂိာ *nanga* ‘house’ → ဂိာဒ် *nangān* ‘household’
 d. ဝာ *ten* ‘life’ → ဝာဒ် *tenan* ‘soul’

There are also some apparent nominalizations in -am and -ang, although these are irregular and non-productive:

- (71) a. ၎်ဗျ: *aja*- ‘play’ → ၎်ဗျဗျ *ajam* ‘toy’
 b. ၎်ဇ်: *gin*- ‘drink’ → ၎်ဇ်ဗျ *ginam* ‘glass’
 c. ၎်ဇ်: *mik*- ‘poison (v.)’ → ၎်ဇ်ဗျ *mikam* ‘poison (n.), venom’
 d. ၎်ဇ်: *nuna*- ‘fly’ → ၎်ဇ်ဗျ *nunam* ‘feather’
- (72) a. ၎်ဗျ: *bayha*- ‘rule’ → ၎်ဗျဗျ *bayhang* ‘government’
 b. ၎်ဗျ *bapa* ‘remaining’ → ၎်ဗျဗျ *bapang* ‘remainder’
 c. ၎်ဇ်: *kada*- ‘collect’ → ၎်ဇ်ဗျ *kadang* ‘committee; alliance’
 d. ၎်ဗျ *mima* ‘possible’ → ၎်ဗျဗျ *mimang* ‘access’

Agentive nouns can be formed from regular nouns with the suffix ဗျ *-maya*, compare the examples in (73). An epenthetic /a/ may be introduced to break up consonant clusters that would otherwise be either difficult to pronounce or violating phonotactics. When the stem of the word the agentive suffix is attached to ends in a consonant or /Ca/, it is also often found fused with the root, sometimes with the first /a/ of *-Caya* lengthened, see (74). Specifically feminine agentive nouns can be derived with the related suffix ဗျ *-vaya*; two examples are given in (75).

- (73) a. ၎်ဇ်: *anl*- ‘bring’ → ၎်ဇ်ဗျ *anlamaya* ‘waiter’
 b. ၎်ဇ်: *bora* ‘sin’ → ၎်ဇ်ဗျ *boramaya* ‘sinner’
 c. ၎်ဇ်: *nasy*- ‘follow’ → ၎်ဇ်ဗျ *nasyamaya* ‘follower’
 d. ၎်ဇ်: *teba*- ‘bake’ → ၎်ဇ်ဗျ *tebamaya* ‘baker’
- (74) a. ၎်ဇ်: *asa*- ‘travel’ → ၎်ဇ်ဗျ *asāya* ‘traveler’
 b. ၎်ဇ်: *ibut*- ‘trade’ → ၎်ဇ်ဗျ *ibutaya* ‘trader, merchant’
 c. ၎်ဇ်: *lant*- ‘lead’ → ၎်ဇ်ဗျ *lantaya* ‘leader; driver’
 d. ၎်ဇ်: *tang*- ‘listen’ → ၎်ဇ်ဗျ *tangaya* ‘listener’
- (75) a. ၎်ဇ်: *gan* ‘child’ → ၎်ဇ်ဗျ *ganvaya* ‘governess’
 b. ၎်ဇ်: *lanya* ‘king’ → ၎်ဇ်ဗျ *lanvaya* ‘queen’

Besides these, there is also a derivative suffix for makers of things, ၎်ဇ် *-ati*, though this is not too productive, and sometimes irregular, as ၎်ဇ်ဗျ *sirtangati* ‘youth’ shows:

- (76) a. ၎်ဇ်: *gindi* ‘poem’ → ၎်ဇ်ဗျ *gindati* ‘poet’
 b. ၎်ဇ်ဗျ *sirtang* ‘young’ → ၎်ဇ်ဗျ *sirtangati* ‘youth’
 c. ၎်ဇ်: *taban*- ‘write’ → ၎်ဇ်ဗျ *tabanati* ‘scribe’
 d. ၎်ဇ်: *vehim* ‘piece of clothing’ → ၎်ဇ်ဗျ *vehimati* ‘tailor’

(77) a. ḡur : *gur*- ‘turn’ → ḡuryan *guryan* ‘coil, cylinder’
 b. mis : *mis*- ‘behave’ → miseryan *miseryan* ‘method, strategy’
 c. nap : *nap*- ‘burn’ → naperyan *naperyan* ‘tinder’
 d. pra : *pra*- ‘glitter, gleam’ → praryan *praryan* ‘spark’

(78) a. Manhattan is famous for its tall **buildings**.
b. **Building** a house is an expensive endeavor.

(79) a. *Lesāra maritay vebānreng tado.*
 Lesa-ara maritay vebān-reng tado
 collapse-3SG.INAN about.to building-A.INAN old
 ‘The old building is about to collapse.’

b. *Le vacyang eda-vebān.*
 Le vac=yang eda=vebān-Ø
 PT.INAN like=1SG.A this=building-TOP
 ‘This building, I like it.’

c. *Ang latayo bayhang vebānyeley yona.*
 Ang lata-yo bayhang-Ø vebān-ye-ley yona
 AT sell-3SG.N government-TOP building-PL-P.INAN 3SG.N.GEN
 ‘The government is selling its buildings.’

d. *Le ming kuysāran vebān-kay dirasyam*
 Le ming kuysa-aran vebān-Ø=kay diras-yam
 PT.INAN can compare-3PL.INAN building-TOP=few splendor-D
 ‘Few buildings can compare to its splendor.’

The above examples condense several properties into one for illustration. Thus, (79a) shows that ṛṣṣṛ *vehān* can serve as basically a subject of a clause, and that it can as well be modified by an adjective—the choice of adjectives is not subject to any distributional restrictions other than those imposed by the semantic frame of HOUSE. In the next example, (79b), ṛṣṣṛ *vehān* serves as the object of the clause and is being determined by the demonstrative prefix ṣṣṣ *eda-* ‘this’. The third example, (79c), shows ṛṣṣṛ *vehān* both pluralized and modified by a possessive pronoun, ṣṣṣ *yona* ‘of it’. And finally, in (79d) we see ṛṣṣṛ *vehān* quantified by the suffix ṣṣṣ *-kay* ‘few’.

Similar to the English example in (78b), Ayeri can also derive nouns from the participle of a verb describing the action as such—a gerund. For an example, I will again draw on the Ayeri translation of Kafka’s short story “Eine kaiserliche Botschaft” (Becker 2012: 2, 14):

- (80) ... *nay ang pətangongva* *ankyū haruyamanas* *nanang megayena*
 ... *nay ang pə-tang-ong=va.Ø* *ankyū haru-yam-an-as* *nanang mega-ye-na*
 ... and AT NFUT-hear-IRR=2SG.TOP truly **beat-PTCP-NMLZ-P** great fist-PL-GEN
yana kunangya vana.
yana kunang-ya vana
 3SG.M.GEN door-LOC 2SG.GEN

‘... and you would indeed hear his magnificent beating at your door very soon.’

The annotations to this translation contain a comment on the grammatical rules which operate in this passage, more specifically also on the gerund derivation ṣṣṣ *haruyaman* ‘beating’:

Furthermore, I wrote *haruyaman* ‘beating’ instead of *haruan* ‘beat(ing)’ because I wanted to emphasize the process of beating as an incomplete action. This is possible here because the word is not topicalized and neither is it marked as a dative, which would also require *haruyamanyam* ‘beat-PTCP-NMLZ-DAT’ to become *haruanyam* ‘beat-NMLZ-DAT’ (the participle marker *-yam* is derived from the dative case ending *-yam*).
 (14–15)

We can read from this description that the participle marker in Ayeri has possibly been grammaticalized from the dative case marker, or that it is at least synchronically homonymous. In order for case marking to operate, this formation has to be nominalized, which is done the usual way by appending ṣṣṣ *-an*, thus yielding the suffix cluster ṣṣṣ *-yaman* for the derivation of verbs as gerunds. If the gerund is marked for dative case, the suffix cluster *ṣṣṣ *-yamanyam* basically undergoes haplogy to a simple nominalized form with the suffix cluster ṣṣṣ *-anyam*:

- (81) *haru-* *haruyam* *haruyaman* **haruyamanyam* *haruanyam*
 haru- → haru-yam → haru-yam-an → haru-yam-an-yam → haru-an-yam
 beat beat-PTCP beat-PTCP-NMLZ beat-PTCP-NMLZ-DAT beat-NMLZ-DAT

The comment on the translation also makes a little note on the gerund being possible because the word is not topicalized. This is based on an old rule that gerunds cannot be topicalized unless nominalized first, however, usage has since changed so that earlier, *ၵၢၢၢၢၢၢ haruyam* would have constituted the gerund form, while even by the time of translating the short story, it had changed to *ၵၢၢၢၢၢၢ haruyaman*. Thus, it is no surprise to see the following example, from the partial translation of Saint-Exupéry's story "Le petit prince" (Becker 2015 [2013]: 3, 13):

- (82) *Sa koronyang palungyaman na Baysānterpeng nay na Bayokivo*
 Sa koron=yang **palung-yam-an-Ø** na Baysānterpeng nay na Bayokivo
 PT knew=1SG.A **distinguish-PTCP-NMLZ-TOP** GEN Realm.Middle and GEN Spring.Little
menaneri nivānyena.
 menan-eri nivān-ye-na
 first-INS glimpse-PL-GEN

'I knew how to distinguish between China and Arizona at first sight.'

A more literal translation of this sentence would be 'The distinguishing of China and Arizona, I knew it at first sight', so the whole passage *ၵၢၢၢၢၢၢ palungyaman ... na Bayokivo* forms the topic of the sentence here, headed by the gerund *ၵၢၢၢၢၢၢ palungyaman* 'distinguishing'. According to the old rule obliquely quoted in the comment to the passage in (80), this should not be possible. As said before, though, usage has changed.

A rule we can gather from the above example from Saint-Exupéry is that gerunds are treated as animate nouns. Since they are impersonal, they trigger neuter agreement on verbs. They can also be the objects of sentences. The passage in (80) furthermore illustrates that gerunds can be modified by The following example shows a gerund used as an agent—basically a subject—as well (Becker 2014):

- (83) *Dilayamanang kalamena babalanas ayonena ...*
 Dila-yam-an-ang kalam-ena bahalan-as ayon-ena ...
 find.out-PTCP-NMLZ-A truth-GEN goal-P man-GEN ...

'(If) finding out the truth is the goal of the man ...'

What all the passages on gerunds quoted before show is that gerunds in Ayeri do not behave like transitive verbs as in English. Thus, what would be the object of the former verb appears in the genitive case in Ayeri. As in English, however, gerunds in Ayeri cannot be pluralized:

- (84) **Noyo vebayamanjang nangayena.*
 Noyo **veha-yam-an-ye-ang** nanga-ye-na
 expensive **build-PTCP-NMLZ-PL-A** house-PL-GEN

'*The buildings of houses are expensive.'

It is possible, however, to quantify gerunds insofar as the quantifier does not imply countable quantities of the action. Moreover, it is possible for gerunds to be modified by possessors. The following two sentences exemplify this use:

- (85) *Ang lugayan delacamanas-ikan kayanya pang.*
 Ang luga=yan.Ø delak-yam-an-as=ikan kayan-ya pang
 AT go.through=3PL.M.TOP suffer-PTCP-NMLZ-P=much war-LOC after

‘They went through a lot of suffering after the war.’

- (86) *Krico malyyamanang muya tan.*
 Krit-yo maly-yam-an-ang muya tan
 annoy-3SG.N sing-PTCP-NMLZ-A wrong 3PL.M.GEN

‘Their off singing is annoying.’

4.2 Pronouns

...

4.2.1 Personal pronouns

...

Bibliography

- Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In *Serial verb constructions: A cross-linguistic typology*, edited by Alexandra Y. Aikhenvald and Robert M. W. Dixon, 1–69. Explorations in Linguistic Typology 2. Oxford: Oxford University Press.
- Bauer, Laurie. 2001. Compounding. In *Language typology and language universals: An international handbook*, edited by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, 1:695–707. Handbooks of Linguistics and Communication Science 20. Berlin: Walter de Gruyter.
- 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>.
- . 2011. Article 1 of the United Nations 'Universal declaration of human rights'. November 22. Accessed July 23, 2016. http://benung.nfshost.com/wp-content/uploads/2011/12/xmp_humanrights.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>.
- . 2014. Translation challenge: The scientific method. *Benung: The Ayeri Language Resource* (April 25). Accessed August 16, 2016. <http://benung.nfshost.com/archives/4092>.
- . 2015 [2013]. Translation challenge: The first two chapters from saint-exupéry's "Le Petit Prince". Accessed August 16, 2016. <http://benung.nfshost.com/wp-content/uploads/2015/05/2013-05-25-stexupery-petitprince-ayeri.pdf>.

- Becker, Carsten. 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>.
- Botha, Rudolf P. 1984. *Morphological mechanisms*. Oxford: Pergamon Press.
- Buch, Armin. 2016. Kunst- und Plansprachen – von Esperanto bis Dothraki. Accessed July 2. <http://www.sfs.uni-tuebingen.de/-abuch/16ss/conlang.html>.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. 2nd ed. London: Blackwell.
- Corbett, Greville G. 2006. *Agreement*. Cambridge Textbooks in Linguistics 52. Cambridge: Cambridge University Press.
- Court, Christopher. 1996. The spread of Brahmi scripts into southeast Asia. In Daniels and Bright 1996, 445–449.
- Daniels, Peter T., and William Bright, eds. 1996. *The world's writing systems*. New York, NY: Oxford University Press.
- Dixon, Robert M. W. 2000. A typology of causatives. In *Changing valency: Case studies in transitivity*, edited by Robert M. W. Dixon and Aleksandra Y. Aikhenvald, 30–83. Cambridge: Cambridge University Press.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Accessed July 12, 2016. <http://wals.info>.
- Eisenberg, Peter, Jörg Peters, Peter Gallmann, Cathrine Fabricius-Hansen, Damaris Nübling, Irmhild Barz, Thomas A. Fritz, Reinhard Fiehler, and Mathilde Hennig. 2016. *Die Grammatik: Unentbehrlich für richtiges Deutsch*. 9th ed. Edited by Angelika Wöllstein. Duden 4. Berlin: Dudenverlag.
- Everson, Michael. 2008. *Proposal for encoding the Javanese script in the UCS*. International Organization for Standardization, January 28. Accessed July 24, 2016. <http://www.evertype.com/standards/iso10646/pdf/n3319r2-javanese.pdf>.
- Fillmore, Charles J. 2003 [1968]. The case for case. Reprint. In *Papers on semantic roles*, vol. 1 of *Form and meaning in language*, edited by Charles J. Fillmore, 21–119. Chicago, IL: University of Chicago Press. Accessed August 1, 2016. <http://linguistics.berkeley.edu/-syntax-circle/syntax-group/spro8/fillmore.pdf>.

- Gaeta, Livio. 2008. Constituent order in compounds and syntax: Typology and diachrony. *Morphology* 18 (2): 117–141. doi:10.1007/s11525-009-9125-x.
- Halle, Morris. 1998. The stress of English words 1968–1998. *Linguistic Inquiry* 29 (4): 539–568.
- Healey, Phyllis M. 1960. *An Agta grammar*. Manila: Manila Bureau of Printing. Accessed August 12, 2016. https://archive.org/details/rosettaproject_dgc_book-1.
- Kager, René. 2007. Feet and metrical stress. In *The Cambridge handbook of phonology*, edited by Paul de Lacy, 195–227. Cambridge: Cambridge University Press.
- Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61 (1): 95–120.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Edited by Alexander Lubotsky. Leiden Indo-European Etymological Dictionary Series 11. Leiden: Brill.
- Kuipers, Joel C., and Ray McDermott. 1996. Insular southeast Asian scripts. In Daniels and Bright 1996, 474–484.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*. 3rd ed. Classics in Linguistics 1. Berlin: Language Science Press. Accessed July 18, 2016. <http://langsci-press.org/catalog/book/88>.
- Lieber, Rochelle. 1994. Root compounds and synthetic compounds. In *The encyclopedia of language and linguistics*, edited by Ronald E. Asher, 7:3607–3610. Oxford: Pergamon Press.
- 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.
- Payne, Thomas E. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Peterson, David J. 2015. *The art of language invention: From horse-lords to dark elves, the words behind world building*. New York, NY: Penguin.
- Salomon, Richard G. 1996. Brahmi and Kharoshthi. In Daniels and Bright 1996, 373–383.

- Süddeutsche Zeitung. 2013. Abschied vom “RkReÜAÜG”-Ungetüm: Längstes deutsches Wort verschwindet. *Süddeutsche Zeitung* (June 3). Accessed August 10, 2016. <http://www.sueddeutsche.de/kultur/laengstes-deutsches-wort-verschwindet-abschied-vom-rkreueaueg-ungetuem-1.1687160>.
- United Nations. 1948. Universal declaration of human rights. United Nations. Accessed July 23, 2016. <http://www.un.org/en/universal-declaration-human-rights/>.
- Wiltshire, Caroline, and Alec Marantz. 2000. Reduplication. In *Morphology: An international handbook on inflection and word-formation*, edited by Geert Booij, Christian Lehmann, and Joachim Mugdan, 1:557–567. Handbooks of Linguistics and Communication Science 17. Berlin: Walter de Gruyter.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington, ID: Indiana University Linguistics Club. Accessed July 22, 2016. https://web.stanford.edu/~zwicky/on_clitics.pdf.
- . 1985. Clitics and particles. *Language* 61 (2): 283–305. Accessed July 26, 2016. <https://web.stanford.edu/~zwicky/cliticsparticles.pdf>.
- Zwicky, Arnold M., and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *N'T*. *Language* 59 (3): 502–513. Accessed July 21, 2016. <https://web.stanford.edu/~zwicky/ZPCliticsInfl.pdf>.

Index

- agglutination, 63
- allophony
 - of consonants, 7–10, 40
 - of vowels, 10–12, 16, 19
- bound words, 64
- Box script, 45
- Brāhmī scripts, 45
- cases, 81–93
 - agent, 81–83
 - causative, 89–90
 - dative, 9, 85–86
 - genitive, 71, 87–88
 - instrumental, 90–92
 - locative, 88–89
 - patient, 83–85
- clitics, 64, 93
- complement clause, 42
- compounds, 38–40, 95–104
 - coordinative, 102
 - endocentric, 96–100
 - exocentric, 103
 - synthetic, 100–102
- consonants, 7–10, 16, 19, 20, 22–24, 26, 28, 47–48
 - clusters, 16, 23, 28–30
- dependent marking, 71
- diacritics, 50–56
- diminutive, 70
- diphthongs, 10, 19, 33, 52
- gender, 76–78
- gerund, 108–111
- goals, 3, 4
- marking strategies, 71
- morphophonology
 - of relative pronouns, 9, 11, 70
 - of the plural marker, 7, 9, 79
- nominalization, 106–111
- nouns, 38, 75–111
- number, 78–81
- numerals, 80
 - digits, 57
- plural, 7, 9, 79
- prefixes, 64, 66–69
 - on nouns, 93–95
- pronouns, 111
 - demonstrative, 68
 - personal, 65, 111
 - reflexive, 67
 - relative, 70
- punctuation, 57–59
- quantifiers, 37
- questions
 - closed, 41
 - open, 41

- reduplication, 69–70, 104–106
- relative clause, 42
- stress, 31–38
- suffixes, 63, 69
- suprasegmental, 64, 70–71
- syllable
 - final, 23–26
 - heavy, 33–38
 - initial, 16–20
 - medial, 20–23
 - monosyllabic words, 26–28, 37
 - superheavy, 34, 36–38
 - types, 14, 33
- Tahano Hikamu, 3, 45–61
 - angular style (ṣūṣ), 59
 - blackletter style, 61
 - hand style (ḥaṣṣ), 60
 - regular/book style, 59
- tense, 67
 - past, 10
- topic, 42, 82
- typology
 - of compounds, 95
 - of morphemes, 63–73
- vowels, 10–12, 16, 20, 23, 28, 48–50