The Syllable Structure of Ayeri's Stems: A Statistical Analysis

Carsten Becker, 3 Apr 2010
carbeck@gmail.com

1 Introduction

In the following analysis I will try to shed more light on the syllable structure of Ayeri's[1] word stems. This analysis has been conducted in order to be able to generate new random words with a weighted-probability word generation algorithm called *kwet*[2] which are as close as possible to already existing words. For this reason only unique dictionary entries, or "stems", have been included in this analysis; fully inflected words would be useless in this case, so set expressions have been disregarded for the matter of this work. The present analysis also serves the purpose of gathering additional information about Ayeri's phonotactic structure for a written grammar of Ayeri.

In the following a short introduction to the method by means of which the data sample was generated and an introduction to the the sound inventory of Ayeri.

1.1 Methodology

In order to gain a base for the following statistical analysis, the pronunciation fields of the Ayeri dictionary database[3] as of 16th March 2010 have been queried and converted into a text file. The pronunciations were given in IPA according to the literary standard dialect, Common Ayeri (Ayeri Vihay), with syllable borders indicated; stress marks have been removed for this analysis, that is, word stress has not been taken into consideration. Furthermore, multiple instances of the same word have been omitted, the pseudo-infinitive ending *-ao* of verbs has been dropped. Several lists have been extracted from this master list:

(1) **All syllables:** Words were split according to the syllable border indications, multiple instances of the same syllable have not been deleted.

(2) **Initial syllables:** Only the initial syllable of each unique word was kept for investigation. Monosyllabic words have been included.

1) Ayeri is an artificial language made by myself to try out something different from the foreign languages I learned in school – English and French. As it is a so-called *artlang* it also serves the purpose of artistic expression (cf. Becker: "Meta Information", 2010).

2) Cf. Hoffman: 2004. Of course the findings of this analysis can be used for other similar programs as well, only that *kwet* has been chosen by means of an example.

3) Cf. Becker: "The Dictionary", 2010.

(3) **Medial syllables:** The initial and final syllables of each word were deleted. Monosyllabic words have not been included.

(4) **Final syllables:** Everything but the final syllable of each word has been deleted. Monosyllabic words have been included.

(5) **Monosyllabic words:** Only monosyllabic words are included.

(6) **Syllable Borders:** A list of all *C.C(C)* occurrences.

1.2 Sound inventory

Ayeri has the following *phonemic* sounds; the second line in italic print shows their orthographic representation:

Consonants: m n ŋ p b t d k g r l j v s h (w ʔ)
m n ng p b t d k g r l y v s h (u –)

Vowels: a aː e eː ə i iː o oː u (uː)
a ā e ē e/ə i ī o ō u (ū)

Diphthongs: ɑɪ ɛɪ ɔɪ ɑʊ
ay ey oy au

[w] and [ʔ] only occur marginally, [w] being an allophone of /u/ before another vowel, [ʔ] in words beginning with a vowel. /uː/ is theoretically possible but does practically never occur. The lax vowels [ɑɛɪɔʊ] are not phonemic, or only negligibly so, because they mostly occur in closed syllables, whereas open syllables usually feature the tense counterparts [aeiou]. Hence, the appearance of each kind of vowel is mostly triggered by its environment (diphthongs not counted: 4319 = 100%):

	a	**e**	**i**	**o**	**u**		**Σ**
open	1515 35.08%	324 7.50%	603 13.96%	252 5.83%	265 6.14%		2959 68.51%
closed	135 3.13%	11 0.25%	51 1.18%	5 0.12%	7 0.16%		209 4.84%

	ɑ	**ɛ**	**ɪ**	**ɔ**	**ʊ**	**ə**	**Σ**
open	7 0.16%	3 0.07%	4 0.09%	3 0.07%	0 0.00%	13 0.30%	30 0.69%
closed	721 16.69%	128 2.96%	149 3.45%	82 1.90%	40 0.93%	1 0.02%	1121 25.96%

[ə] only occurs phonemically in the tense prefixes *kə-, mə-, və-, pə-, sə-,* otherwise it is an allophone of unstressed /e/. As far as tense prefixes are concerned, the vowel arch-phoneme is *a, except for *sə-*, where it is *e, which is reflected in the

combination of a tense prefix with a verb starting in either of the mentioned vowels:

(1) medial past tense (*mə-*):

a. *mə-* + apa- ‘to laugh’ → *māpa-* [ma:pa] ‘laughed’

b. *mə-* + *int-* ‘to buy’ → *məint-* [mə.ɪnt] ‘bought’

(2) medial future tense (*sə-*):

a. *sə-* + *epa-* ‘to refuse’ → *sēpa-* [se:pa] ‘will refuse’

b. *sə-* + *ur-* ‘to crawl’ → *səur-* [sə.ur] ‘will crawl’

Examples (1b) and (2b) demonstrate as an example that [ə] plus another vowel leads to a vowel hiatus usually. Note however that only 41 verb stems (1.99%) of 2061 surveyed stems begin with a vowel.

Note as well that the combinations /kj/ and /tj/ may be palatalized to [t͡ʃ] in colloquial speech; /gj dj/ may become [d͡ʒ] accordingly. For this analysis however, palatalization has been ignored, as it is not a phonemic feature.

2 General observations

2.1 Stems in general

Firstly, we will examine to which percentage the following number of syllables per word occurs in the sample (2,061 words with 4906 syllables):

1 syllable	147	7.13%	**4 syllables**	157	7.62%
2 syllables	1168	56.67%	**5 syllables**	15	0.73%
3 syllables	573	27.80%	**6 syllables**	1	0.05%

Words with more than three syllables are usually compounds; words with more than six syllables have not been encountered. Clearly, bi- (56.67%) and trisyllabic (27.8%) stems are favoured.

2.2 Syllables in general

Ayeri’s syllables may generally be of the following form:

(1) $\sigma_O \rightarrow \begin{Bmatrix} C \\ C+C \\ \sigma_N \end{Bmatrix}$ (2) $\sigma_N \rightarrow \begin{Bmatrix} V \\ V+V \\ V: \end{Bmatrix}$ (3) $\sigma_C \rightarrow \begin{Bmatrix} C \\ \sigma_N \end{Bmatrix}$

(1) constitutes that the onset *O* of a syllable *σ* consists of a consonant and optionally another consonant or it may be identical to the nucleus *N*.

(2) states that the nucleus *N* of a syllable *σ* consists of a vowel and optionally another vowel, where two vowels refer to a diphthong. It may be a long vowel as well. Note that *āy* may occur, but only marginally so (see the table below).

(3) finally constitutes that the coda *C* of a syllable *σ* may either be a consonant or the nucleus *N*. Two consonants or geminated consonants do not exist in codas, which can be observed in the statistics below. Note that σ_N is included for onsets and codas since syllables do not necessarily need to contain these elements separately; they may just as well only consist of a vowel.

In the following shall be discussed which syllable types can occur in the language, and to which percentage of the surveyed sample. The figures for the different syllable types – regardless of their position in a word – are as follows:

***C-**[4]	250	5.10%	***CC-**	30	0.61%
C	1	0.02%	**CCV**	79	1.61%
CCVC	61	1.24%	**CCVV**	6	0.12%
CCV:C	10	0.20%	**CV**	2513	51.22%
CVC	1357	27.66%	**CVV**	284	5.79%
CV:	30	0.61%	**CVVC**	2	0.04%
CV:C	69	1.41%	**CV:V**	2	0.04%
V	161	3.28%	**VC**	37	0.75%
VV	13	0.26%	**VVC**	1	0.02%
CCV:	0	0.00%	**CCV:CC**	0	0.00%
CCVCC	0	0.00%	**CCVVC**	0	0.00%
CCVVCC	0	0.00%	**CV:CC**	0	0.00%
CVCC	0	0.00%	**CVVCC**	0	0.00%
V:	0	0.00%	**V:C**	0	0.00%
V:CC	0	0.00%	**VCC**	0	0.00%
VVCC	0	0.00%			

The above chart shows that *CV* (51.22%) syllables are by far the most common type, followed by *CVC* (27.66%) and *CVV* (5.79%). Syllables with two final consonants generally do not occur; a long vowel does not occur in all positions where two vowels are possible either, or two vowels do not occur in all positions where there may be a long vowel: For example *CCVV* but no **CCV:*, *CCV:C* but no **CCVVC* although these are not impossible.

[4] These are due to verb stems ending in arbitrary consonants, like e.g. *tahan-* ‘to write’. They can be counted with closed syllables mostly, however this poses the difficulty of some codas not being allowed, e.g. *-k* – though note that forms like *tik-* ‘to dig’ → *tikvāng* ‘you dig’ are common nonetheless, however such final syllables cannot be counted together with all other closed syllables, as depending on the following consonant they may slip into the next syllable or not: *tik]-nang* ‘we dig’, but *ti-k]yāng* ‘he digs’, especially when plosives are involved.

In the following we will investigate how common these syllable types are in different positions (initial, medial, final; monosyllabic words) and which sounds are typically filled in the slots for consonants and vowels in order to be able to generate more words with the same structure or aesthetics of those already existing. Syllable patterns which never occur will not be listed any more hereafter.

3 Analysis of syllables in all possible positions

3.1 Initial syllables

The figures for the different syllable patterns for initial syllables are as follows – the sample size is 2061 syllables; monosyllabic words are included:

C-	0	0.00%	**CC-**	1	0.05%
C	1	0.05%	**CCV**[5]	62	3.01%
CCVC	14	0.68%	**CCVV**	2	0.10%
CCV:C	5	0.24%	**CV**[6]	1319	64.00%
CVC	357	17.32%	**CVV**	67	3.25%
CV:	16	0.78%	**CVVC**	2	0.10%
CV:C	12	0.58%	**CV:V**	2	0.10%
V	152	7.38%	**VC**	35	1.70%
VV	13	0.63%	**VVC**	1	0.05%

The following table gives an overview of the frequency of initial single consonants, the percentages are relative to the number of syllables as given above:

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
Σ	565 27.41%	873 42.36%	64 3.11%	223 10.82%	252 12.23%	1977 95.92%
plosives (vl.)	150 7.28%	235 11.40%		163 7.91%	201[7] 9.75%	749 26.59%
plosives (v.)	117 5.68%	103 5.00%		59 2.86%		279 13.54%
nasals	193 9.36%	134 6.50%		1 0.05%		328 15.91%
fricatives	104 5.05%	236 11.45%			51 2.47%	391 18.97%
taps		82 3.98%				82 3.98%
approxs.	1[8] 0.05%	83 4.03%	64 3.11%			148 7.13%

5) 6 of which CCV- (0.29%).

6) 4 of which CV- (0.19%).

7) The glottal stop [ʔ] indicates syllables beginning with a vowel. See the chart about vowels below.

8) The voiced labiovelar approximant [w] is an allophone of [u] before a vowel.

Combinations of two syllable-initial consonants occur to the following percentages in relation to the overall number of initial syllables:[9]

	C+w	**C+r**	**C+j**	**Σ**
Σ	7 0.34%	50 2.43%	27 1.31%	84 4.08%
p		25 1.21%	2 0.10%	27 1.31%
b	3 0.15%	6 0.29%		9 0.44%
m			3 0.15%	3 0.15%
t		8 0.39%	7 0.34%	15 0.73%
d			2 0.10%	2 0.10%
n	1 0.05%		5 0.24%	6 0.29%
s			3 0.15%	3 0.15%
r	2 0.10%			2 0.10%
k	1 0.05%	8 0.39%	5 0.24%	14 0.68%
g		3 0.15%		3 0.15%

Now let us have a look at the nuclei of initial syllables (2058 syllables surveyed[10]):

i	469	22.79%	**u**ː	0	0.00%
iː	2	0.10%	**ə**	3	0.15%
e	334	16.23%	**aɪ**	53	2.58%
eː	1	0.05%	**a**ːɪ	2	0.10%
a	778	37.80%	**eɪ**	23	1.12%
aː	27	1.31%	**ɔɪ**	4	0.19%
o	119	5.78%	**ʊɪ**	1	0.05%
oː	1	0.05%	**ɑʊ**	1	0.05%
u	155	7.53%	**ea**[11]	1	0.05%

And finally, the sounds which occur in codas of initial syllables (sample size: 2060 syllables[12]):

9) All other combinations turned out 0 (0.00%).

10) Due to the issues mentioned in footnote 4, 3 syllables did not have a vowel (0.15%) – the overall count has thus been reduced, since syllables in "-" are not supposed to occur regularly in the context of automatic word generation.

11) This is clearly a mistake in the result set, as *ea* would turn into *[e.a]*. *-ea* is an allomorph of the locative case marker *-ya* which occurs when the word to be marked ends in *-e*, so that *-e + -ya → -ēa*.

12) For the same reason as stated in footnote 10, however this only applies to 1 syllable here (0.05%).

p	1	0.05%	**r**	47	2.28%
b	0	0.00%	**l**	32	1.55%
m	58	2.82%	**j**	0	0.00%
v	0	0.00%	**k**	2	0.10%
w	0	0.00%	**ŋ**	128	6.21%
t	1	0.05%	**h**	0	0.00%
d	1	0.05%	**V**	1535	74.51%
n	139	6.75%	**VV**	82	3.98%
s	18	0.87%	**V:**	16	0.78%

The combinations including *V* refer back to the chart of syllable nuclei regarding probabilities.

3.2 Medial syllables

3.2.1 The sound structure of full syllables

The different syllable patterns have been found in medial syllables in the following amounts (sample size: 1081 syllables):

C-	0	0.00%	**CC-**	0	0.00%
C	0	0.00%	**CCV**	26	2.41%
CCVC	3	0.28%	**CCVV**	2	0.19%
CCV:C	5	0.46%	**CV**	737	68.18%
CVC	217	20.07%	**CVV**	34	3.15%
CV:	20	1.85%	**CVVC**	2	0.19%
CV:C	14	1.30%	**CV:V**	2	0.19%
V	7	0.65%	**VC**	8	0.74%
VV	3	0.28%	**VVC**	1	0.09%

The distribution of syllable patterns is similar to that of initial syllables, however single-vowel syllables only occur in 0.65% of cases, while *CCV* syllables are on third place at 3.15%. *CV* and *CVC* are most common again, at 68.18% and 20.07%, respectively. In the following, an overview of initial single consonants of medial syllables (sample size: 1081 syllables).

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
Σ	242 22.39%	500 46.25%	66 6.11%	172 15.91%	65 6.01%	1045 96.67%
plosives (vl.)	67 6.20%	111 10.27%		101 9.34%	19 1.76%	298 27.57%
plosives (v.)	49 4.53%	72 6.66%		71 6.57%		192 17.76%
nasals	72 6.66%	76 7.03%				148 13.69%

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
fricatives	53 4.90%	54 5.00%			46 4.26%	153 14.15%
taps		117 10.82%				117 10.82%
approxs.	1 0.09%	70 6.48%	66 6.11%			137 12.67%

Again, dentals are especially common, followed by labials. A list of consonant clusters with their respective numbers of occurrence and percentages relative to the number of initial consonant clusters surveyed as above:

	C+w	**C+r**	**C+j**	**Σ**
Σ	4 0.37%	16 1.48%	16 1.48%	36 3.33%
p		3 0.28%		3 0.28%
m			1 0.09%	1 0.09%
t		7 0.65%	4 0.37%	11 1.02%
d		3 0.28%	3 0.28%	6 0.56%
n			3 0.28%	3 0.28%
s			2 0.19%	2 0.19%
r	2 0.19%			2 0.19%
k		2 0.19%	2 0.19%	4 0.37%
g	2 0.19%	1 0.09%	1 0.09%	4 0.37%

Combinations of *b* × *w, r, j* do not occur in this case.

Another look at syllable nuclei in medial syllables:

i	215	19.89%	**u**ː	0	0.00%
iː	2	0.19%	**ə**	4	0.37%
e	89	8.23%	**aɪ**	28	2.59%
eː	0	0.00%	**a**ːɪ	2	0.19%
a	554	51.25%	**eɪ**	5	0.46%
aː	36	3.33%	**ɔɪ**	5	0.46%
o	56	5.18%	**ʊɪ**	1	0.09%
oː	1	0.09%	**ɑʊ**	2	0.19%
u	60	5.55%	**ea**	1	0.09%

In comparison to initial syllables there is a striking lack of /e/, however /a/ is still about twice as frequent as /i/.

And finally, a table giving the frequencies of final consonants in medial syllables:

p	0	0.00%	**r**	14	1.30%
b	0	0.00%	**l**	6	0.56%
m	25	2.31%	**j**	0	0.00%
v	0	0.00%	**k**	1	0.09%
w	1	0.09%	**ŋ**	87	8.05%
t	1	0.09%	**h**	0	0.00%
d	0	0.00%	**V**	772	71.42%
n	48	4.44%	**VV**	39	3.61%
s	0	0.00%	**V:**	20	1.85%

As expected, apart from syllables ending in a vowel, the nasals /n m ŋ/ are common, /r/ and /l/ appear as well. Although /s/ is possible in syllable-final position however, at least as compared to overall tendencies, it does not occur as the coda consonants of medial syllables.

3.2.2 Consonant clusters across the syllable border

The table of the final consonants of medial syllables only offers limited information on which combinations of consonants across syllable borders are typical of the language, since only words were counted which actually do have a middle syllable – that is, they are at least trisyllabic. In the following a table which shows the figures for *C.C* combinations overall, including bisyllabic words (sample size: 469 syllables).

	m.C	**t.C**	**n.C**	**s.C**	**r.C**	**l.C**	**k.C**	**ŋ.C**	**Σ**
Σ	65 13.86%	2 0.43%	162 34.54%	5 1.07%	52 11.09%	27 5.76%	1 0.21%	155 33.05%	469 100.00%
p	23 4.90%				5 1.07%	1 0.21%			29 6.18%
b	33 7.04%					5 1.07%		1 0.21%	39 8.32%
m								1 0.21%	1 0.21%
w	1 0.21%		2 0.43%		2 0.43%				5 1.07%
v	2 0.43%		5 1.07%		1 0.21%	11 2.35%			19 4.05%
t			47 10.02%		3 0.64%	3 0.64%			53 11.30%
d	1 0.21%		54 11.51%		7 1.49%				62 13.22%
n					3 0.64%	1 0.21%			4 0.85%
s					3 0.64%		1 0.21%		4 0.85%
r				1 0.21%					1 0.21%

	m.C	**t.C**	**n.C**	**s.C**	**r.C**	**l.C**	**k.C**	**ŋ.C**	**Σ**
l		2 0.43%	4 0.85%		1 0.21%			1 0.21%	8 1.71%
j	4 0.85%		50 10.66%	4 0.85%	20 4.26%	6 1.28%			84 17.91%
k	1 0.21%				5 1.07%			27 5.76%	33 7.04%
g					1 0.21%			124 26.44%	125 26.65%
h					1 0.21%			1 0.21%	2 0.43%

Note that a combination of *nasal + stop* is rather common: It appears in 65.67% of all surveyed cases.

3.3 Final syllables

The figures for the various possible syllable patterns of final syllables (Sample size: 2057 syllables[13]):

C-	250	12.15%	**CC-**	30	1.46%
C	0	0.00%	**CCV**[14]	30	1.46%
CCVC	30	1.46%	**CCVV**	6	0.29%
CCV:C	10	0.49%	**CV**[15]	687	33.40%
CVC	719	34.95%	**CVV**	207	10.06%
CV:	6	0.29%	**CVVC**	2	0.10%
CV:C	61	2.97%	**CV:V**	2	0.10%
V[16]	5	0.24%	**VC**	8	0.39%
VV	3	0.15%	**VVC**	1	0.05%

Note that verbs ending in *Ca-* (pseudo-infinitive ending in *-āo*, from *-a + -ao*) have been counted as *CV* syllables, however *C-* syllables (pseudo-infinitive ending in *-ao*) have been counted separately. Since verbs are a rather common part of speech, the percentage of this group is quite large (12.15%). As regards issues of grouping, see footnote 4 above.

The single-consonant onsets behave as follows (sample size: 2057 syllables):

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
Σ	305 14.83%	840 40.84%	230 11.18%	224 10.89%	102 4.96%	1701 82.69%

[13] Cf. footnote 4 – those lines which came out as just "-" have been omitted, which was the case four times, hence the difference in number between initial and final syllables.

[14] 17 of which CCV- (1.46%)

[15] 186 of which CV- (9.04%)

[16] 1 of which V- (0.05%)

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
plosives (vl.)	89 4.33%	158 7.68%		123 5.98%	17 0.83%	387 18.81%
plosives (v.)	55 2.67%	95 4.62%		101 4.91%		251 12.20%
nasals	93 4.52%	171 8.31%				148 12.83%
fricatives	64 3.11%	153 7.44%			85 4.13%	302 14.68%
taps		169 8.22%				169 8.22%
approxs.	4 0.19%	94 4.57%	230 11.18%			328 15.95%

While /j/ is the single most common initial consonant in final syllables (11.18%), dentals are the largest group at 40.84%.

As I said earlier, syllables in *Cao* need to be differentiated, which is why the above table does not include them. Again a table of initial consonants, for the mentioned *C*- syllables (sample size: 2057 syllables):

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
Σ	44 2.14%	133 6.47%	18 0.88%	47 2.28%	8 0.39%	250 12.15%
plosives (vl.)	12 0.58%	22 1.07%		12 0.58%		46 2.24%
plosives (v.)	9 0.44%	16 0.78%		35 1.70%		60 2.92%
nasals	16 0.78%	36 1.75%				148 2.53%
fricatives	4 0.19%	26 1.26%			8 0.39%	38 1.85%
taps		19 0.92%				19 0.92%
approxs.	3 0.15%	14 0.68%	18 0.88%			35 1.70%

The high frequency of /g/ is due to /ŋ.g/ being a frequent combination, however the combination *ŋ+V* is not necessarily always realized as *ŋ.g+V* in spoken language.

The frequencies for *C+C* clusters in the onsets of final syllables are listed in the table below (sample size: 2057 syllables). Since *C+w* is so common, note again that [w] is strictly speaking not a phoneme, but an allophone of /u/ before vowels.

	C+w	**C+r**	**C+j**	**Σ**
Σ	25 1.22%	28 1.36%	53 2.58%	106 5.15%
p	1 0.05%	5 0.24%	3 0.15%	9 0.44%
b	1 0.05%	4 0.19%		5 0.24%

	C+w	C+r	C+j	Σ
m	1 0.05%		1 0.05%	2 0.10%
t	3 0.15%	9 0.44%	17 0.83%	29 1.41%
d		2 0.10%	6 0.29%	8 0.39%
n	3 0.15%		4 0.19%	7 0.34%
s	4 0.19%		5 0.24%	9 0.44%
r	4 0.19%			4 0.19%
k	4 0.19%	4 0.19%	12 0.58%	20 0.97%
g	3 0.15%	4 0.19%	5 0.24%	12 0.58%
h	1 0.05%			1 0.05%

Vowel nuclei of final syllables (sample size: 1777 syllables[17]):

i	132	7.43%	**uː**	0	0.00%
iː	1	0.06%	**ə**	13	0.73%
e	49	2.76%	**aɪ**	148	8.33%
eː	1	0.06%	**aːɪ**	2	0.11%
a	1001	56.33%	**eɪ**	25	1.41%
aː	73	4.11%	**ɔɪ**	26	1.46%
o	182	10.24%	**ʊɪ**	12	0.68%
oː	1	0.06%	**ɑʊ**	7	0.39%
u	103	5.80%	**ea**	1	0.06%

/a/ and /aɪ/ are exceedingly common as compared to other syllable positions.

Coda consonants in final syllables (sample size: 1777 syllables again, see below):

p	0	0.00%	**r**	16	0.90%
b	0	0.00%	**l**	27	1.52%
m	81	4.56%	**j**	0	0.00%
v	0	0.00%	**k**	1	0.06%
w	0	0.00%	**ŋ**	209	11.76%
t	0	0.00%	**h**	0	0.00%
d	0	0.00%	**V**	724	40.74%
n	437	24.59%	**VV**	216	12.16%
s	59	3.32%	**Vː**	6	0.34%

[17] Due to the mentioned issues with *Cao* syllables, 280 syllables (13.61%) in this set did not have a vowel but ended in "-".

3.4 Monosyllabic words

Finally, we will look at monosyllabic words. Note that monosyllabic words are mostly function words like adverbs, articles (case markers, indefinite article), particles (for example TAM markers) or pronouns. Sample size: 180 words.[18]

adjectives	10	5.56%	**interjections**	5	2.78%
adpositions	2	1.11%	**nouns**	36	20.00%
adverbs	22	12.22%	**particles**	19	10.56%
articles	13	7.22%	**pronouns**	60	33.33%
conjunctions	2	1.11%	**verbs**	11	6.11%

As in the chapters above, the table below displays the various possible syllable patterns and their respective frequencies (sample size: 149 syllables):

C-	4	2.68%	**CC-**	6	4.03%
C	0	0.00%	**CCV**	5	3.36%
CCVC	2	1.34%	**CCVV**	2	1.34%
CCV:C	5	3.36%	**CV**	19	12.75%
CVC	61	40.94%	**CVV**	12	8.05%
CV:	7	4.70%	**CVVC**	2	1.34%
CV:C	10	6.71%	**CV:V**	1	0.67%
V	1	0.67%	**VC**	7	4.70%
VV	3	2.01%	**VVC**	1	0.67%

As in final syllables of polysyllabic words, there is a prevalence of *CVC* syllables (40.94%), followed by *CV* syllables (12.75%). Syllables ending in a pseudo-infinitive are comparatively rare (6.71%).

Initial single consonants in monosyllabic words (sample size: 149 syllables):

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
Σ	24 16.11%	58 38.93%	19 12.75%	11 7.38%	16 10.74%	128 85.91%
plosives (vl.)	5 3.36%	19 12.75%		8 5.37%	4 2.68%	36 24.16%
plosives (v.)	1 0.67%	4 2.68%		3 2.01%		8 5.37%
nasals	8 5.37%	9 6.04%				148 11.41%
fricatives	10 6.71%	10 6.71%			12 8.05%	32 21.48%

[18] Because the original data set did not contain any information of the part of speech, another query had to be made on 3 April 2010 – though with a filter to only include records entered before 16 March 2010. The discrepancy in the number of results is due to zero-derivations for example from nouns to adjectives where the word class differs, but not the surface form.

	labials p b m v w	**dentals** t d n s l r	**palatals** j	**velars** k g ŋ	**glottals** ʔ h	**Σ**
taps		7 4.70%				7 4.70%
approxs.	0 0.00%	9 6.04%	19 12.75%			28 18.79%

Initial consonant clusters in monosyllabic words (sample size: 149 syllables):

	C+w	**C+r**	**C+j**	**Σ**
Σ	3 2.01%	7 4.70%	11 7.38%	21 14.09%
p		2 1.34%		2 1.34%
m			1 0.67%	1 0.67%
t		3 2.01%	4 2.68%	7 4.70%
d			2 1.34%	2 1.34%
n	1 0.67%		3 2.01%	4 2.68%
r	2 1.34%			4 1.34%
k		2 1.34%	1 0.67%	3 2.01%

Combinations of *b, s, g, h* × *w, r, j* do not occur in monosyllabic words, which makes them even more restrictive than medial syllables, where only combinations with /b/ as the first part do not occur.

Vowel nuclei of monosyllabic words (sample size: 148 syllables[19]):

i	17	11.49%	**u**ː	0	0.00%
iː	2	1.35%	**ə**	3	2.03%
e	16	10.81%	**aɪ**	10	6.76%
eː	0	0.00%	**a**ːɪ	1	0.68%
a	56	37.84%	**eɪ**	5	3.38%
aː	19	12.84%	**ɔɪ**	3	2.03%
o	10	6.76%	**ʊɪ**	1	0.68%
oː	1	0.68%	**ɑʊ**	0	0.00%
u	3	2.03%	**ea**	1	0.68%

/a/ and /i/ are still the most common vowels (37.84% and 11.49% respectively), however they are closely followed by /e/ (10.81%).

[19] 1 syllable (0.67%) ended in *Cao*, so there was no vowel in the syllable.

Terminal consonants in monosyllabic words (sample size: 148 syllables[20]):

p	0	0.00%	**r**	2	1.35%
b	0	0.00%	**l**	4	2.70%
m	7	4.73%	**j**	0	0.00%
v	0	0.00%	**k**	1	0.68%
w	0	0.00%	**ŋ**	25	16.89%
t	0	0.00%	**h**	0	0.00%
d	0	0.00%	**V**	37	25.00%
n	33	22.30%	**VV**	17	11.49%
s	15	10.14%	**V**:	7	4.73%

The results of the above table are not surprising, they fit the overall tendency of the only consonants in codas being /m n ŋ s r l/ well.

4 *kwet* sample output

The *kwet* rule file made from the data gathered in this analysis can be found in the Appendix. Unfortunately *kwet* does not support entering probabilities to the degree of accuracy presented here, also I could see no practical way to make the program prefer certain consonant clusters straddling the syllable border while not allowing others.

Running the rule file returned the following sample output of 100 randomly generated words. Improper words are marked with an asterisk, questionable ones with a question mark at the beginning of the line; the respective word in orthography has been added manually.

Monosyllabic

daŋ	(dang)	joŋ	(yong)	san	(san)	to	(to)
ja	(ya)	pas	(pas)	taɪ	(tay)		
jeŋ	(yeng)	ri	(ri)	te	(te)		

Bisyllabic

an.sam	(ansam)	ha.sam	(hasam)	ku.tan	(kutan)
ar.nan	(arnan)	hi.kan	(hikan)	lu.si	(lusi)
be.ta	(beta)	ʔhiŋ.saŋ	(hingsang)	lun.von	(lunvon)
ʔbi.ge	(bige)	i.ran	(iran)	maŋ.gaɪ	(mangay)
bi.maɪ	(bimay)	ir.ja	(irya)	maŋ.kan	(mankan)
de.lal	(delal)	ja.ni	(yani)	me.ga	(mega)
del.ri	(delri)	ʔkaŋ.peŋ	(kangpeng)	me.na	(mena)
do.mos	(domos)	ki.jaɪ	(kiyay)	mem.laŋ	(memlang)

[20] Cf. footnote 19 above.

ʔmi.mel	(mimel)	pin.da	(pinda)	si.to	(sito)
ʔmi.sen	(misen)	pra.jɔɪ	(prayoy)	so.na	(sona)
na.raɪ	(naray)	pri.leŋ	(prileng)	sun.ram	(sunram)
ʔna.res	(nares)	ra.naŋ	(ranang)	ʔtaŋ.ha	(tangha)
na.rwɔɪ	(naruoy)	re.jan	(reyan)	ti.ja	(tiya)
ʔnes.ka	(neska)	re.taɪ	(retay)	ti.jan	(tiyan)
ni.jan	(niyan)	ri.jar	(riyar)	ti.lan	(tilang)
nu.ra	(nura)	rin.ji	(rinyi)	tu.da	(tuda)
*pa.aɪ	(paay)	sa.naɪ	(sanay)	tu.jan	(tuyan)
pa.ju	(payu)	sa.ra	(sara)	u.paŋ	(upang)
pa.ma	(pama)	sam.ran	(samran)	vi.dam	(vidam)
pa.raŋ	(parang)	sar.so	(sarso)	vi.jal	(viyal)
pa.vaɪ	(pavay)	*seŋ.von	(sengvon)		
pe.ta	(peta)	si.gon	(sigon)		

Trisyllabic

a.ti.mis	(atimis)	*ji.a.saɪ	(yiasay)	sa.vi.nu	(savinu)
ʔba.ko.kas	(bakokas)	*lan.na.lor	(lannalor)	ʔse.pam.jaɪ	(sepamyay)
de.ni.maŋ	(denimang)	lil.va.jan	(lilvayan)	ta.ka.ra	(takara)
di.dim.kun	(didimkun)	ma.la.kaŋ	(malakang)	*te.maŋ.da	(temangda)
di.pan.ra	(dipanra)	ʔmaɪ.gwam.va	(mayguamva)	tu.vi.gam	(tuvigam)
gan.la.jum	(ganlayum)	pa.da.raɪ	(padaray)	ʔveɪ.gu.ben	(veyguben)
gi.ri.jin	(giriyin)	pon.dja.goŋ	(pondyagong)	vi.la.ra	(vilara)
ha.ra.jam	(harayam)	ʔri.vaŋ.has	(rivanghas)	vi.gi.raɪ	(vigiray)
i.bi.na	(ibina)	sa.teŋ.ram	(satengram)		

Trisyllabic words appear too much like compounds than single stems, so mono- and disyllabic words should be favoured in generating words. For the purpose of sensible automatic word creation, combinations with more than three syllables have been disregarded, also long vowels, which usually result from two same vowels colliding either by appending suffixes or due to compounding. Frequencies have however not been recalculated, as there already is generous rounding of percentages to full integers.

5 The most probable word

According to the statistics carried out to this point, the most probable word is *sayan*, which means ‘hole, cave’:[21]

2 syll.	→	56.67% (1168/4906)	y	→	11.80% (230/2057)
CV	→	64.00% (1319/2061)	a	→	56.33% (1001/1777)
CVC	→	34.95% (719/2057)	n	→	24.59% (437/1777)
s	→	11.45% (236/2061)	≈ 0.00003570911955		
a	→	37.80% (778/2058)	≈ 35.71 ppm		

[21] Cf. Becker: “Sayan”, 2010.

6 Conclusion

The present survey of Ayeri's word stems has shown in detail how independent lexemes are typically composed. Difficulties arose during the analysing process previous to writing this paper due to irregular handling of tense and lax vowels as well as syllable borders. Another difficulty was posed by the so-called *pseudo-infinitive* ending *-ao*, which is actually never used in the language's morphology. While there is a tendency for words to only end in nasals, liquids or /s/, this pseudo-infinitive causes the irregularity of verb stems being able to actually only consist of a C(C) syllable. Another point of criticism is that the present analysis only relies on a sample of about 2,000 words, which is a rather small sample to produce significant numbers, especially given that Ayeri is a constructed language that is still evolving – that is, there are still new word stems added from time to time, which of course changes the numbers presented here again. Whether probabilities of consonant clusters across syllable borders can be aptly programmed into the rules file used to generate the words above is yet to be determined, however bisyllabic words are already rather close to the overall aesthetics even neglecting those combinations.

References

Becker, Carsten. "Meta Information." *Tay Benung. The Ayeri Resource.* Carsten Becker. 2003-10. Web. 16 Mar. 2010. <*http://benung.nfshost.com/index.php?go=meta*>

Becker, Carsten. "The Dictionary 2.5 BETA." *Tay Benung. The Ayeri Resource.* BETA. Carsten Becker. 2003-10. Web. 16 Mar. 2010. <*http://benung.nfshost.com/index.php?go=dictionary*>

"Sayan." *Tay Benung. The Ayeri Resource.* BETA. Carsten Becker. 2003-10. Web. 29 Mar. 2010. <*http://benung.nfshost.com/index.php?go=dictionary&action=query&id=1864*>

Hoffman, Paul. *kwet.* 4 Apr. 2004. SourceForge. 4 Apr. 2004. Web. 24 Mar. 2010. <*http://kwet.sourceforge.net*>

Appendix: *kwet* rule file

```
# === KWET FILE FOR AYERI RANDOM WORD GENERATION === #
# ===                                            === #
# === BASED ON THE STEMS IN THE DICTIONARY       === #
# === AS OF 2010-03-16 (VERBS' -AO CHOPPED,      === #
# === SUFFIXES COUNTED IN, PERCENTAGES ROUNDED   === #
# === TO INTEGERS)                               === #

# ABBREVIATIONS

# *** W   Word            S   Syllable
# *** I   initial         M   medial
# *** F   final           C   consonant
# *** V   vowel           CC  consonant cluster
# *** VV  diphthong       VL  long vowel

# WORD

W : 7(S) | 57(IS . FS) | 28(IS . MS . FS)
# 8(IS . MS . MS . FS) | 1(IS . MS . MS . MS . FS)
### -> four-syllable words are more than usually not stems!

# SYLLABLE PATTERNS

# --- INITIAL SYLLABLES

IS : 3(ICCI IV) | 1(ICCI IV ICF) | 64(ICI IV) | 17(ICI IV ICF) | 3(ICI IVV) | 7(IV) |
    2(IV ICF) | 1(IVO)
# 2(ICI IVL) | 1(ICI IVL ICF)
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Other initial syllables (those < 0.5% of IS)
IVO : 8(ICCI IVV) | 8(ICI IVV ICF) | 4(IVV ICF)
# 4(ICI) | 19(ICCI IVL ICF) |  8(ICI aːɪ)
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them
### -> a syllabic initial consonant would be atypical and is due to one instance of
    initial syllabic [ŋ]


# *** Initial consonant in initial syllable
ICI : 7(p) | 6(b) | 9(m) | 5(v) | 11(t) | 5(d) | 7(n) | 11(s) | 4(r) | 4(l) | 3(j) |
    8(k) | 3(g) | 2(h)


# *** Initial consonant cluster in initial syllable
ICCI : 4(bw) | 1(nw) | 2(rw) | 1(kw) | 30(pr) | 7(br) | 10(tr) | 10(kr) | 4(gr) | 2(pj)
    | 4(mj) | 8(tj) | 2(dj) | 6(nj) | 4(sj) | 6(kj)
```

```
# *** Final consonant in initial syllable
ICF : 3(m) | 7(n) | 1(s) | 2(r) | 2(l) | 6(ŋ)


# *** Vowel in initial syllable
IV : 23(i) | 16(e) | 38(a) | 5(o) | 8(u)


# *** Vowel cluster (i.e. diphthong) in initial syllable
IVV : 27(eɪ) | 62(aɪ) | 5(ʊɪ) | 5(ɑʊ)
# 2(aːɪ) | 5(ea)
### -> "ea" in this position would normally be "ya" and is probably due to a mistake in
    analysis.
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Long vowel in initial syllable
IVL : 6(iː) | 3(eː) | 88(aː) | 3(oː)
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# --- MEDIAL SYLLABLES


MS : 2(MCCI MV) | 68(MCI MV) | 20(MCI MV MCF) | 3(MCI MVV) | 1(MV) | 1(MV MCF) | 2(MSO)
# 2(MCI MVL) | 1(MCI MVL MCF)
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Other medial syllables (those < 0.5% of MS)
MSO : 17(MCCI MV MCF) | 11(MCCI MVV) | 6(MCI MVV MCF) | 3(MVV MCF)
# 14(MCCI MVL MCF) | 6(MCI aːɪ)
### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Initial consonant in medial syllable
MCI : 6(p) | 5(b) | 7(m) | 5(v) | 10(t) | 7(d) | 7(n) | 5(s) | 11(r) | 7(l) | 6(j) |
    9(k) | 7(g) | 4(h)


# *** Initial consonant cluster in medial syllable
MCCI : 6(rw) | 6(gw) | 8(pr) | 19(tr) | 8(dr) | 6(kr) | 3(gr) | 3(mj) | 11(tj) | 8(dj) |
    8(nj) | 6(sj) | 6(kj) | 3(gj)


# *** Final consonant in medial syllable
MCF : 2(m) | 9(n) | 2(s) | 1(r) | 1(l) | 8(ŋ)


# *** Vowel in medial syllable
MV : 22(i) | 8(e) | 51(a) | 5(o) | 6(u)


# *** Vowel cluster (i.e. diphthong) in medial syllable
MVV : 12(eɪ) | 63(aɪ) | 5(ɔɪ) | 2(ʊɪ) | 2(ea) | 5(ɑʊ)
```

```
# *** Long vowel in medial syllable

MVL : 5(iː) | 92(aː) | 2(oː)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# --- FINAL SYLLABLES


FS : 1(FCCI FV) | 1(FCCI FV FCF) | 33(FCI FV) | 35(FCI FV FCF) | 10(FCI FVV) | 2(FSO)

# 3(FCI aːɪ)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Other final syllables (those < 0.5% of MS)

FSO : 13(FCCI FVV) | 5(FCI FVV FCF) | 13(FVV) | 4(FVV FCF)

# 23(FCCI FVL FCF) | 13(FCI FVL) | 5(FCI aːɪ)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# *** Initial consonant in final syllable

FCI : 5(p) | 3(b) | 5(m) | 3(v) | 9(t) | 5(d) | 10(n) | 9(s) | 10(r) | 5(l) | 12(j) |
    7(k) | 6(g) | 5(h)


# *** Initial consonant cluster in final syllable

FCCI : 1(pw) | 1(bw) | 1(mw) | 3(tw) | 3(nw) | 4(sw) | 4(rw) | 4(kw) | 3(gw) | 1(hw) |
    5(pr) | 4(br) | 8(tr) | 2(dr) | 4(kr) | 4(gr) | 3(pj) | 1(mj) | 16(tj) | 6(dj) |
    4(nj) | 11(kj) | 5(gj)


# *** Final consonant in final syllable

FCF : 5(m) | 25(n) | 3(s) | 1(r) | 2(l) | 12(ŋ)


# *** Vowel in final syllable

FV : 7(i) | 3(e) | 56(a) | 10(o) | 6(u)


# *** Vowel cluster (i.e. diphthong) in final syllable

FVV : 11(eɪ) | 67(aɪ) | 12(ɔɪ) | 5(ʊɪ) | 3(ɑʊ)


# *** Long vowel in final syllable

FVL : 1(iː) | 1(eː) | 96(aː) | 1(oː)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them


# --- MONOSYLLABLES


S : 3(SCCI SV) | 1(SCCI SV SCF) | 1(SCCI SVV) | 13(SCI SV) | 41(SCI SV SCF) | 8(SCI SVV)
    | 1(SCI SVV SCF) | 1(SV) | 5(SV SCF) | 2 (SVV) | 1(SVV SCF)

# 3(SCCI SVL SCF) | 5(SCI SVL) | 7(SCI SVL SCF) | 1(SCI aːɪ)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them
```

```
# *** Initial consonant in monosyllable

SCI : 3(p) | 5(m) | 7(v) | 13(t) | 3(d) | 6(n) | 7(s) | 5(r) | 6(l) | 13(j) | 5(k) |
    2(g) | 3(h)


# *** Initial consonant cluster in monosyllable

SCCI : 5(nw) | 9(rw) | 9(pr) | 14(tr) | 10(kr) | 5(mj) | 19(tj) | 10(dj) | 14(nj) |
    5(kj)


# *** Final consonant in monosyllable

SCF : 5(m) | 22(n) | 10(s) | 1(r) | 3(l) | 1(k) | 17(ŋ)


# *** Vowel in monosyllable

SV : 11(i) | 11(e) | 38(a) | 7(o) | 2(u) | 2(ə)


# *** Vowel cluster (i.e. diphthong) in monosyllable

SVV : 24(eɪ) | 48(aɪ) | 14(ɔɪ) | 5(ʊɪ) | 5(ea)


# *** Long vowel in monosyllable

SVL : 9(iː) | 86(aː) | 5(oː)

### -> long vowels are often due to compounding or appending other morphemes, so prime
    stems wouldn't usually have them
```