

Chapter 11

Asymmetries in vowel-pair frequencies and height harmony in Bantu

Stephen Nichols

University of Manchester

In this paper, I present an exploration of vowel-pair frequencies in the nouns of six five-vowel Bantu languages and discuss their relationship with those alternations seen in verbs due to height harmony. Of particular interest is the disparity in the frequencies of two vowel pairs disagreeing in height, specifically [e.i] and [o.u], both between themselves and with the corresponding pairs which agree in height, namely [e.e] and [o.o]. The results show that while both [e.e] and [o.o] are over-represented, [e.i] and [o.u] are generally under-represented. In addition to this, [o.u] is consistently less frequent than [e.i] and the difference in representation between [o.o] and [o.u] is larger than between [e.e] and [e.i].

1 Introduction

It is well documented in the literature that vowel height harmony as instantiated in the Bantu languages is most often asymmetric with respect to rounding and/or backness and also often gives rise to alternations only in verbs (see e.g. Hyman 1999; for more see also §2). In this paper, I present an exploration of vowel-pair frequencies in nouns in six five-vowel Bantu languages – Chewa, Kalanga, Lozi, Makhuwa, Pende and Yao – and also discuss the results in the context of the system of height harmony found in each language. For reasons of space, I limit myself to the discussion of just four vowel pairs that are of particular interest, namely [e.e]–[e.i] and [o.o]–[o.u].

The results reveal that, in nouns, the two pairs that agree in height – [e.e] and [o.o] – consistently occur more frequently than their counterparts that do not agree in height – [e.i] and [o.u], with the differences between [o.o] and [o.u]

Stephen Nichols. 2021. Asymmetries in vowel-pair frequencies and height harmony in Bantu. In Akinbiyi Akinlabi, Lee Bickmore, Michael Cahill, Michael Diercks, Laura J. Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds.), *Celebrating 50 years of ACAL: Selected papers from the 50th Annual Conference on African Linguistics*, 245–259. Berlin: Language Science Press. DOI: 10.5281/zenodo.5578838



being starker than those between [e.e] and [e.i]. This is despite the fact that, in these languages, active alternations are only seen in verbs. Moreover, there is a trend for the pairs [e.i] and [o.u] to be relatively under-represented compared to [e.e] and [o.o] across the sample. The results also show that, though both occur at below-expected levels, the rounded back pair [o.u] is less frequent than [e.i], even taking into account the relative frequencies of the component vowels according to position.

Previous work such as Archangeli et al. (2012b) has also noted the under-representation in nouns of pairs repaired by alternations in verbs.¹ However, the difference between the two non-harmonic pairs [e.i] and [o.u] is not remarked upon, though it can be seen in their data (see e.g. Figure 2). In addition to this, the sample used by Archangeli et al. (2012b) looked only at canonical five-vowel languages.

In this chapter, I examine three non-canonical languages, including two in which alternations of front vowels in verbs do not occur. In both languages lacking front-vowel alternations, [e.e] and [o.o] are more represented than [e.i] and [o.u] respectively; however, [e.e] is found only marginally more than [e.i] in Lozi.

In sum, though alternations are only observed in verbs, the results show that there is nevertheless a preference in nouns for those pairs that agree in height.

2 Background

This exploration of vowel-pair frequencies considers a sample of six five-vowel² Bantu languages. The details of these languages are provided in Table 11.1 before vowel height harmony in particular is discussed further below.

As can be seen from Table 11.1, each language possesses a certain harmony system, with three different systems being represented in this sample. Vowel height harmony writ large is a widespread trait among the Bantu languages (Hyman 1999: 236, 2003: 46–47; Nurse & Philippson 2003: 7; Odden 2015: §1). In five-vowel languages, such as Shona (S.11; Beckman 1997), this often leads to active alternations only in verbal suffixes, excluding final vowels. Indeed, Beckman (1997: 38) comments that “[t]he distributional generalisations which apply to height features in Shona verbs apparently do not hold of Shona nouns” and that “vowel height in nouns is contrastive outside of the root-initial syllable”. In other cases, however, static generalisations analogous to the alternations found in verbs may

¹In addition, Harrison et al. (2002–2004) found that, in Swahili, verbs were highly height harmonic and that nouns were also generally height harmonic at an above-chance rate.

²That is, languages having the vowel phoneme inventory /i u e o a/.

11 Asymmetries in vowel-pair frequencies and height harmony in Bantu

Table 11.1: The six-language sample of five-vowel Bantu languages

Language	Guthrie code	Harmony system	Data source	Size
Chewa	M.31b	canonical	Mtenje (2001)	24,076
Kalanga	S.16	canonical	Mathangwane (1994)	8,505
Lozi	K.21	back only	Jalla (1982)	49,981
Makhuwa	P.31	back only	Kisseberth (1996)	29,802
Pende	L.11	quasi-canonical	Gusimana (1972)	38,385
Yao	P.21	canonical	Ngunga (2001)	25,954

also be seen within a majority of noun stems.³ This has previously been said, for example, of Chewa (Scullen 1992 in Downing & Mtenje 2017: 75).

The commonest variant of height harmony is the so-called “canonical” pattern (after Hyman 1999: 238). In five-vowel languages that possess canonical height harmony – such as Chewa, Kalanga and Yao – high vowels are lowered to mid vowels by preceding mid vowels whereas the low vowel /a/ neither triggers nor undergoes lowering and is opaque. In addition, as mentioned in §1, canonical height harmony is asymmetric with respect to rounding and/or backness. That is, /i/ is lowered to [e] after both /e/ and /o/ whereas /u/ is lowered to [o] only following /o/. In at least purely descriptive terms then, it is useful to talk of both front and back height harmony.

Canonical height harmony is exemplified in (1) and (2) with Chewa data taken from Downing & Mtenje (2017: 71–72).⁴

- | | |
|---|---|
| <p>(1) a. <i>-phikila</i> ‘to cook for’
 b. <i>-khutula</i> ‘to be satisfied with’
 c. <i>-tsekela</i> ‘to close for’
 d. <i>-gonela</i> ‘to sleep on’
 e. <i>-valila</i> ‘to put on’</p> | <p>(2) a. <i>-pitikula</i> ‘to overturn’
 b. <i>-funthula</i> ‘to loosen’
 c. <i>-tsekula</i> ‘to open’
 d. <i>-wonjola</i> ‘to spring a trap’
 e. <i>-sankula</i> ‘to choose out from’</p> |
|---|---|

The majority of work on height harmony in five-vowel Bantu languages has focused on this pattern. Indeed, Chewa has been an especially popular case study

³Note also that there are Bantu languages that show alternations in nouns due to vowel harmony, such as noun class prefixes in seven-vowel Koyo (C.24; see e.g. Hyman 1999: 240).

⁴For the sake of the simplicity of exposition, the infinitive prefix, tone and effects such as penultimate lengthening have been omitted in the examples provided in this section.

(see e.g. Mtenje 1985; Scullen 1992; Harris 1994; 1997; Sandstedt 2019). A selection of other well-known languages exhibiting this are Bemba (M.42; Kula & Marten 2000; Kula 2002), Luganda (E.15; Katamba 1984), Shona (S.11; Beckman 1997) and Swahili (G.42; Marten 1996; 1997).

The “quasi-canonical” variety of harmony found in Pende differs from the pattern found in Chewa in just one instance. Specifically, in front height harmony, mid [e] is found after the low vowel /a/ rather than high [i]. This is illustrated in (3) and (4) with examples from Gusimana (1972).

- | | | | |
|-----|---|-----|--|
| (3) | a. <i>-shitila</i> ‘to close for’ | (4) | a. <i>-jitulula</i> ‘to untie’ |
| | b. <i>-tungila</i> ‘to build for’ | | b. <i>-kubula</i> ‘to collapse’ |
| | c. <i>-bembela</i> ‘to leave for’ | | c. <i>-ketula</i> ‘to make a notch’ |
| | d. <i>-solela</i> ‘to clear [land] for’ | | d. <i>-logola</i> ‘to pour out’ |
| | e. <i>-talela</i> ‘to monitor for’ | | e. <i>-batula</i> ‘to cut off, detach’ |

The final two languages in the sample – Lozi and Makhuwa – lack front height harmony, as seen in (5) and (6) with Lozi data culled from Jalla (1982).

- | | | | |
|-----|--------------------------------------|-----|---|
| (5) | a. <i>-pimisa</i> ‘to help avoid’ | (6) | a. <i>-bihela</i> ‘to report to’ |
| | b. <i>-hupulisa</i> ‘to remind’ | | b. <i>-fuluhela</i> ‘to paddle towards’ |
| | c. <i>-lembisa</i> ‘to put to shame’ | | c. <i>-lemela</i> ‘to fell for’ |
| | d. <i>-longisa</i> ‘to help load’ | | d. <i>-shombotela</i> ‘to catch for’ |
| | e. <i>-tamisa</i> ‘to help tie’ | | e. <i>-shamela</i> ‘to urinate in’ |

They do, however, possess back height harmony. This is shown in (7), with data once again coming from Lozi (Jalla 1982).

- | | | |
|-----|-----------------------------------|---------------------------------|
| (7) | a. <i>-tinulula</i> ‘to undress’ | d. <i>-notolola</i> ‘to unlock’ |
| | b. <i>-lutulula</i> ‘to unthatch’ | e. <i>-pakulula</i> ‘to unbolt’ |
| | c. <i>-lekulula</i> ‘to resell’ | |

3 Methodology

3.1 Sources and pre-processing

The data used in this study are taken from the Comparative Bantu Online Dictionary (CBOLD; citations for individual sources are given in Table 11.1).⁵ Each

⁵CBOLD is accessible via the following URL: <http://www.cbold.ish-lyon.cnrs.fr/>.

file was first hand-corrected for machine-readability. This consisted mainly of corrections to the tab-separation of columns and the addition of obvious missing line breaks, as well as certain minor corrections to obvious errors in transcriptions.⁶ All subsequent processing and analysis was conducted using R (R Core Team 2019).

It should be noted here that the data sets contain entries in citation forms with both simple and complex stems (for both nouns and verbs), including some instances of nouns derived from verbs. No morphemic analysis was conducted as the data sets did not indicate boundaries between roots within compounds, or between roots and suffixes. Thus, in this respect, the results presented in §4 are based on vowel pairs that occur regardless of morphological context.

Each individual data set contained multiple columns of information, the total number of which varied slightly from language to language; however, only two columns contained in the raw data – namely the word form itself and the part of speech – were ultimately used. For each word, all content after the first space in the word column was removed to reduce contamination by non-target-language material (such as misaligned English-language definitions). This was then stripped of all non-alphabetic characters such as punctuation and numbers, and converted to lower case. Next, orthographic long vowels were shortened such that *aa* became *a* and so on. This was done as the feature of interest in this study is the quality rather than quantity of vowels. What's more, in certain cases long vowels may not be consistently transcribed, if at all. Finally, the number of part-of-speech labels for each data set was reduced to just three: verb, noun and other.⁷ In the sections that follow, I concentrate predominantly on nouns.

3.2 Processing and analysis

The final data sets having been derived, the numbers of each unique vowel phoneme and pair were then computed, with each item retaining the part-of-speech label of the word in which it occurred. This yielded the raw count of each vowel phoneme and pair within a particular language. From this, the following within-language measures were derived: observed frequency, expected frequency, and observed–expected ratio.

⁶This was, however, conducted with all due caution in order to avoid introducing any potential bias into the results.

⁷The data source for Lozi, for example, contained around 70 separate part-of-speech labels, largely distinguishing various different properties of verbs which were unnecessary for the present study. In addition to this, the Lozi dictionary was unique in that it also provided the perfective forms of verbs. In order to maintain a level of consistency across the entire sample, these were actively removed.

The observed frequency of a vowel phoneme or pair is the quotient of the raw count of that particular phoneme or pair, divided by the total number of all vowel phonemes or pairs in the data set.

The expected frequency of a given vowel pair might have been derived simply by multiplying together the observed frequencies of the two constituent vowel phonemes of that pair without regard to which position in the pair the two vowels occur.⁸ However, as noted in Archangeli et al. (2012a,b), due to, for example, morphological bias, the same vowel may not necessarily be found equiprobably in both positions and so, in order to take this into account, separate observed proportions were calculated for each vowel in both first and second position, which were then multiplied together to give the expected frequency of each pair.

The observed–expected ratio of a pair is the quotient of the observed frequency of that pair divided by its expected frequency. This is used as a gauge of the level of representation. A ratio of around 1 indicates that the vowel pair occurs approximately as frequently as expected, less than 1 that it is less frequent than expected (i.e. under-represented) and a ratio of more than 1 that it is more frequent than expected (i.e. over-represented). In addition to this, the magnitude of the ratio conveys the degree to which a pair is over- or under-represented within the language.

Lastly, note that, descriptive statistics were used rather than inferential statistics due to the small sample size of the current data set.

4 Results

Firstly, the raw counts in both nouns and verbs for each of the four vowel pairs of interest in this study – i.e. [e.e], [e.i], [o.o] and [o.u] – are shown in Figure 11.1. These show that the pairs one would expect to be infrequent given the descriptions in §2 do indeed occur at very low levels in verbs but that such low levels are not found for other pairs. Somewhat similarly, although generally found in higher numbers than in verbs, within each of the six languages in the sample [e.i] is found consistently less frequently than [e.e] in nouns. Likewise, [o.u], without exception, occurs less often in nouns than [o.o].

Note that, in Figures 11.1, 11.2 and 11.3, pale blue is used to indicate those pairs that are assumed to be harmonic in that particular language and dark red signals those pairs assumed to be non-harmonic on the basis of alternations seen in verbal suffixes. This is used to illustrate how the harmony systems differ between languages with regard to these pairs. Thus, it can be seen that the above disparities between pairs are found in nouns irrespective of the harmony system of the particular language.

⁸Of course, the two component vowels of a pair may or may not be the same as one another.

11 Asymmetries in vowel-pair frequencies and height harmony in Bantu

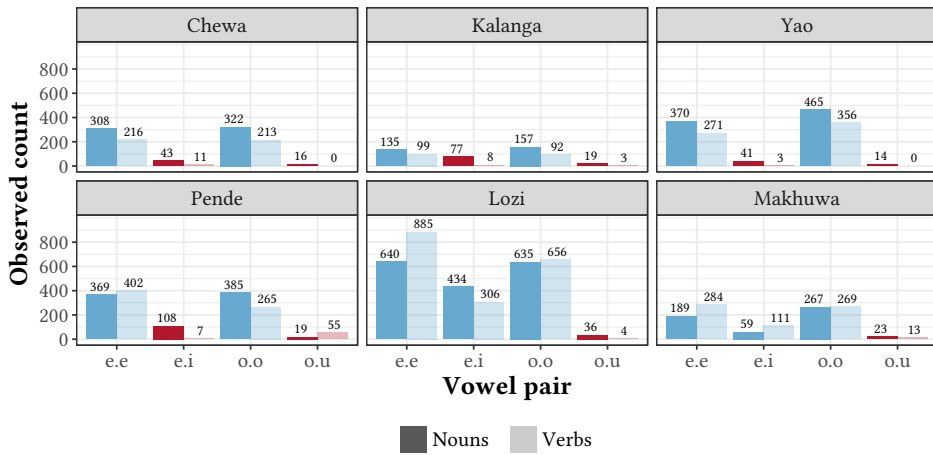


Figure 11.1: Bar chart and figures for raw counts in nouns and verbs in all six languages

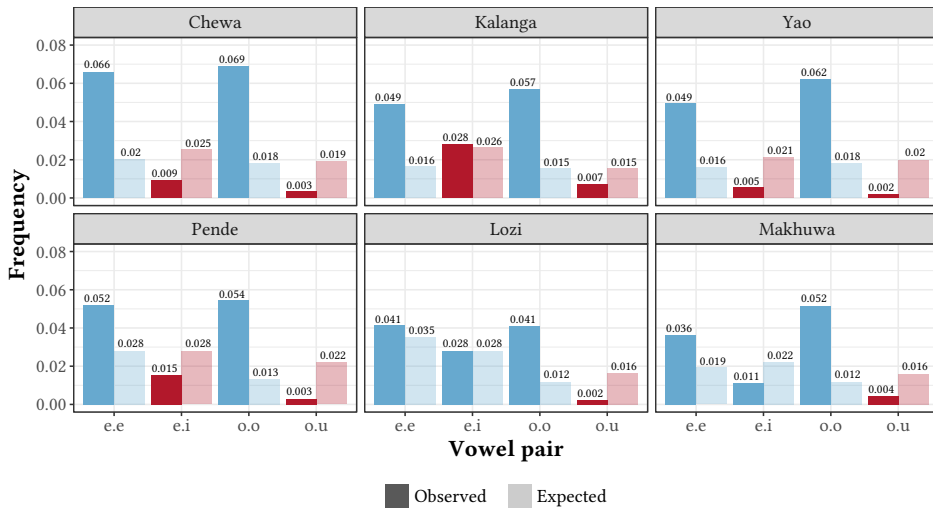


Figure 11.2: Bar chart and figures for both observed and expected frequencies in nouns in all six languages

The differences described above in nouns on the basis of the raw count data are also reflected in the observed frequencies in nouns, given in Figure 11.2. For the sake of completeness, alongside this, Figure 11.2 includes the expected frequencies for these same vowel pairs. This shows a different, much more evenly spread distribution (certain somewhat minor differences notwithstanding), indicating that these differences are not due simply to the relative frequencies in each vowel in the pair, even accounting for position.

Next, let us consider the corresponding observed–expected ratios in Figure 11.3. Firstly, this shows that the pattern seen with the raw counts and observed frequencies remains. That is, in every language, the observed–expected ratios for [e.e] and [o.o] are larger than those for [e.i] and [o.u] respectively (however, for Lozi this difference is only slight).

Secondly, those pairs that agree in height – i.e. [e.e] and [o.o] – consistently occur at expected or higher-than-expected levels whereas those pairs that disagree in height – i.e [e.i] and [o.u] – are found at much lower levels than would be anticipated in all but 2 out of 12 cases.

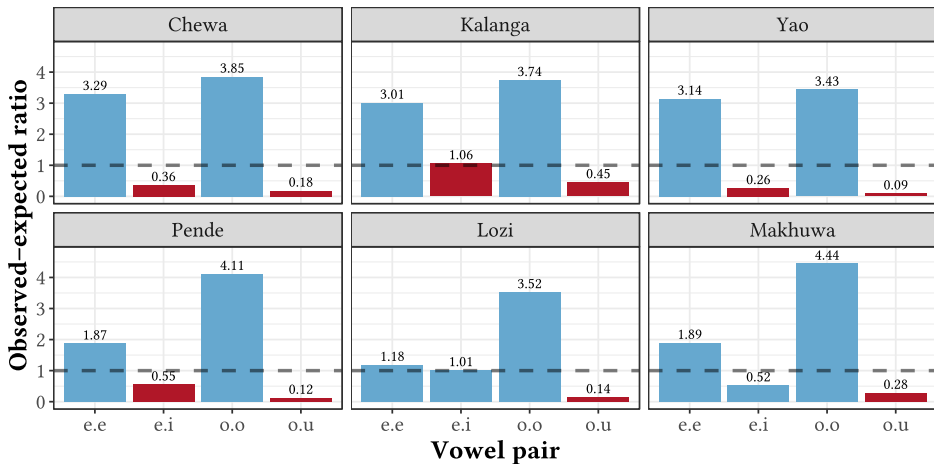


Figure 11.3: Bar chart and figures for observed–expected ratios in nouns in all six languages (the horizontal dashed line indicates a ratio of 1).

For the sake of ease of comparison, and in order to reiterate and clarify certain points made above, the observed frequencies and observed–expected ratios of relevant pairwise comparisons are pooled and re-presented in Figures 11.4, 11.5 and 11.6. Note, that, in these figures, the observed–expected ratios have been converted to log values in order to show over- and under-representation on the

same scale, with positive values indicating over-representation and negative values under-representation.

Firstly, Figure 11.4 shows the clear difference across the sample in the incidence of [e.e] and [e.i] in nouns. Likewise, Figure 11.5 demonstrates the similar but starker disparity between [o.o] and [o.u] that recurs in each language. Figure 11.6 shows that, when comparing [e.i] and [o.u] in these six languages, the latter occurs in nouns at a noticeably lower rate than the former. Lastly, comparing Figure 11.4 with Figure 11.5 shows that overall the difference in nouns between [o.o] and [o.u] is starker than between [e.e] and [e.i].

The analysis also revealed that, in nouns, back rounded [o.u] was the least common of the 25 possible vowel pairs in all six of the sample languages. This was despite the fact that [o.u] was not the vowel pair expected to be the most infrequent in any language in the sample (range 20–23; mean 21.83; median 22; standard deviation 1.17).

Front unrounded [e.i], however, varied in nouns between a rank of 13 in Kalanga and 24 in Yao (mean 20.67; median 22; standard deviation 4.03). Nonetheless, [e.i] was found to have a lower rank than expected in 5 of the 6 languages, with the sole exception being Kalanga.

5 Discussion

The first point to be acknowledged is that the under-representation of [e.i] and [o.u] and the corresponding over-representation of [e.e] and [o.o] in nouns, where alternations are absent, mirrors what is found in a great many five-vowel Bantu languages in verbs, where harmony can be seen to induce alternations.

Before discussing this any further, a related aspect of the results (especially as demonstrated by Figure 11.3) must first be addressed. This is the fact that the results do not conclusively demonstrate whether or not the harmony system of a given language, in particular the presence – as in Chewa, Kalanga, Pende and Yao – or the absence – as in Lozi and Makhuwa – of front height harmony in verbs, has an influence – direct or not – on the frequencies of [e.e] and [e.i] in nouns.

In both canonical Kalanga and non-canonical Lozi, [e.i] occurs at roughly the level we would expect given random combination. However, in Kalanga [e.e] is far commoner than [e.i]. This is not the case in Lozi, where [e.e] occurs with similar frequency to [e.i], being only marginally more represented. Nevertheless, this pattern is not replicated in Makhuwa, which also lacks front height harmony but matches more closely the observed–expected ratios of, for example, Pende (which of course does exhibit a form of front height harmony).

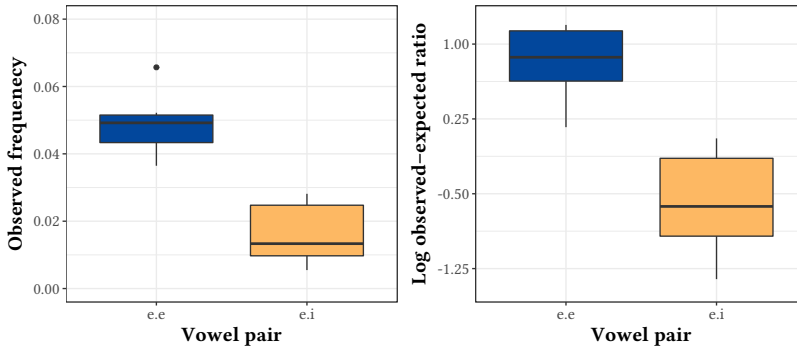


Figure 11.4: Boxplots of pooled observed frequencies (left) and log observed-expected ratios (right) in nouns for [e.e] and [e.i] only

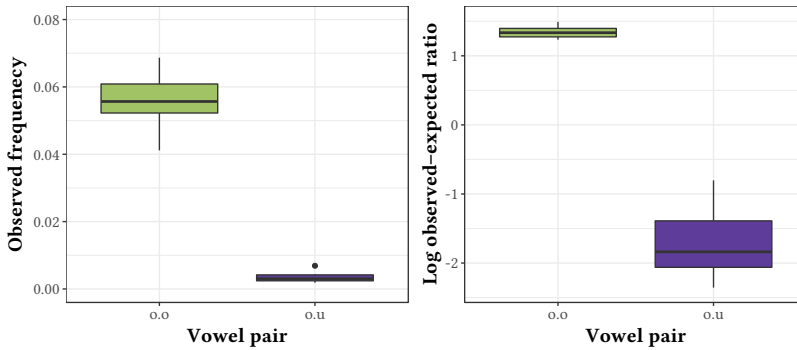


Figure 11.5: Boxplots of pooled observed frequencies (left) and log observed-expected ratios (right) in nouns for [o.o] and [o.u] only

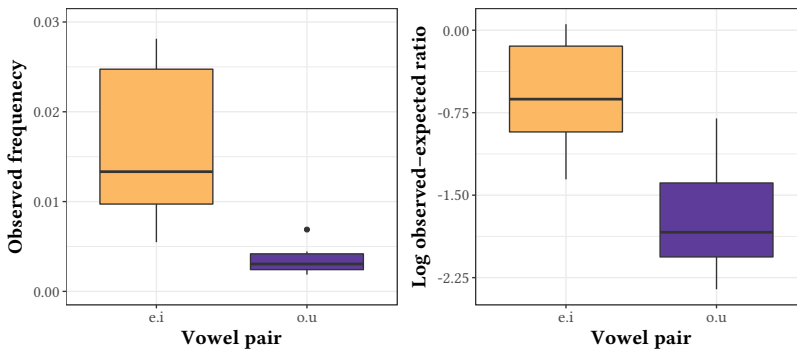


Figure 11.6: Boxplots of pooled observed frequencies (left) and log observed-expected ratios (right) in nouns for [e.i] and [o.u] only

This having been noted, the patterns seen when comparing [e.e] with [e.i] and [o.o] with [o.u] in Chewa, Kalanga, Yao and Pende and when considering the difference between the frequencies of [o.o] and [o.u] in Lozi and Makhuwa could perhaps be interpreted as a gradient counterpart in nouns to the categorical rule in verbs. This interpretation would be comparable to Martin's (2011) observations that, in English, geminates – which are prohibited tautomorphemically – are relatively under-represented in heteromorphemic items and that, in Navajo, where sibilants co-occurring within roots must agree for anteriority, compound words containing sibilants disagreeing in anteriority are likewise found less frequently than expected. On the basis of these data, Martin (2011) argues that such lexical biases arise as a compromise brought about by the competition between certain phonotactic preferences and semantic preferences.

Archangeli et al. (2012a,b) – which employed a substantially similar methodology to this paper and whose sample of six canonical five-vowel Bantu languages included Chewa, Kalanga and Yao – have also commented on the patterns of frequency found in nouns as compared to verbs. They suggest that patterns such as those observable in Chewa could be argued to be predictable simply on inductive grounds. Thus, arguing against Universal Grammar (UG) and in a favour of Emergent Grammar (EG), Archangeli et al. (2012b: 214) assert that “UG predicts an absence of an extension while EG predicts extension due to the attractor effect.”⁹

However, as Martin (2011: 757) acknowledges when considering comparable patterns in tauto- and heteromorphemic contexts in English and Navajo, such distributions “could both result from the same phonetic pressure”. This is something akin to the notion of “rule scattering” (Bermúdez-Otero 2015 after Robinson 1976). In other words, it may instead be that rather than one part of the lexicon exerting an influence over another, both share a common cause.¹⁰ In the case of English, the same pressure that gave rise to a prohibition on geminates within a morpheme also had a hand in guiding the competition in the lexicon between compounds containing geminates and those lacking geminates.

Recall also that, as noted above and as can be seen in Figure 11.3, Makhuwa, which lacks front-vowel alternations, also displays a marked difference between

⁹The “attractor effect” being the “the gradual generalization of an effect to broader classes”, such as the potential spread of vowel height harmony in verbs to nouns (Archangeli et al. 2012b: 198).

¹⁰For example, the results of the present study might be taken to suggest that the under-representation of [e.i] and [o.u] is due to the well-groundedness of their avoidance because of differences in height and that there is some sort of stronger bias against [o.u] than [e.i] (cf. however, Archangeli et al. 2012b who failed to find similar consistent effects in a sample of non-Bantu languages).

[e.e] and [e.i] in nouns. Therefore, in this particular instance at least, this cannot be due to the influence of the effect of harmony in verbs.

Moreover, it is potentially problematic for such accounts that [o.u] is much more under-represented than expected in nouns than [e.i] considering that, in four of the six sample languages, both pairs of vowels occur in extremely low levels in verbs. Thus, as [e.i] and [o.u] behave similarly in verbs in languages with both front and back height harmony, a difference in their behaviours in nouns is unexpected. This is also further reflected in a more widely cross-Bantu context by the fact that there appear to be no – or at least vanishingly few – cases of Bantu languages that possess front height harmony but lack back height harmony (Hyman 1999: 245).

6 Conclusion

In this study of six Bantu languages possessing a five-vowel inventory, I examined vowel-pair frequencies in nouns with a particular focus on [e.e]–[e.i] and [o.o]–[o.u], and considered the results of these pairs in relationship with alternations seen in verbs due to vowel height harmony.

The results show that [e.e] and [o.o], which agree in height, are generally over-represented in nouns and that [e.i] and [o.u], which correspondingly disagree in height, are generally under-represented. It was also found that the overall difference in representation of [o.o] compared with [o.u] was greater than the difference between [e.e] and [e.i]. Additionally, the back rounded pair [o.u] is without exception less frequent than front unrounded [e.i] in the current sample. This then suggests that, although there is pressure to avoid both [e.i] and [o.u], this pressure is greater regarding [o.u] than [e.i]. As for the effect of the harmony system, in three of the four languages with front height harmony [e.i] was under-represented in nouns but this was only found in one of the two languages in which front height harmony is absent.

Future work will widen the scope of this study to look at other vowel pairs that are integral to the vowel harmony systems of the Bantu language, such as [e.u] which is an important pair when considering the asymmetric manifestation of harmony in Chewa and other canonical languages. It would also be desirable not only to look at particular languages in closer detail but also to investigate a larger sample of Bantu languages and also expand out of the Bantu family, looking at a larger typologically-balanced sample, to determine whether there is any broader cross-linguistic tendency for the languages to follow patterns similar to those observed in the data presented here.

Acknowledgements

I would like to thank Wendell Kimper and Ricardo Bermúdez-Otero for supervision, Stefano Coretta for useful discussion, the audience at the 50th Annual Conference on African Linguistics in Vancouver, Canada for their questions and feedback as well as editor Laura Downing and the two reviewers of this paper for their corrections, criticisms and suggestions. Any shortcomings present in this work are my own.

References

- Archangeli, Diana, Jeff Mielke & Douglas Pulleyblank. 2012a. From sequence frequencies to conditions in Bantu vowel harmony: Building a grammar from the ground up. *McGill Working Papers in Linguistics* 22(1). 1–26.
- Archangeli, Diana, Jeff Mielke & Douglas Pulleyblank. 2012b. Greater than noise: Frequency effects in Bantu height harmony. In Bert Botma & Roland Noske (eds.), *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*, 191–222. Berlin: De Gruyter.
- Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46.
- Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.
- Downing, Laura J. & Al D. Mtenje. 2017. *The phonology of Chichewa* (The Phonology of the World's Languages). New York: Oxford University Press.
- Gusimana, Barthelemy. 1972. *Dictionnaire pende-français*. Available to download from the Comparative Bantu Online Dictionary. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Gusimana>.
- Harris, John. 1994. Monovalency and opacity: Chichewa height harmony. *UCL Working Papers in Linguistics* 6. 509–547.
- Harris, John. 1997. Licensing inheritance: An integrated theory of neutralization. *Phonology* 14. 315–370.
- Harrison, K. David, Emily Thomforde & Michael O'Keefe. 2002–2004. *The vowel harmony calculator*. http://www.swarthmore.edu/SocSci/harmony/public_html.

- Hyman, Larry M. 1999. The historical interpretation of vowel harmony in Bantu. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu historical linguistics: Theoretical and empirical perspectives*, 235–295. Stanford, CA: CSLI Publications.
- Hyman, Larry M. 2003. Segmental phonology. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 42–58. London: Routledge.
- Jalla, Adolphe. 1982. *Database of “Dictionary of the Lozi language, vol. 1: Lozi–English”*. Available to download from the Comparative Bantu Online Dictionary. Lusaka. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Lozi>.
- Katamba, Francis. 1984. A nonlinear analysis of vowel harmony in Luganda. *Journal of Linguistics* 20(2). 257–275.
- Kisseberth, Charles W. 1996. *Makhua–English lexicon*. Available to download from the Comparative Bantu Online Dictionary. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Makhua>.
- Kula, Nancy C. 2002. *The phonology of verbal derivation in Bemba*. PhD thesis. Universiteit Leiden. (Doctoral dissertation).
- Kula, Nancy C. & Lutz Marten. 2000. Constraints and processes: Evidence from Bemba, Herero and Swahili. *SOAS Working Papers in Linguistics* 10. 91–102.
- Marten, Lutz. 1996. Swahili vowel harmony. *SOAS Working Papers in Linguistics & Phonetics* 6. 61–75.
- Marten, Lutz. 1997. Licensing constraints for Swahili vowel harmony. *Papers in Linguistics from the University of Manchester* 2. 239–255.
- Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87(4). 751–770.
- Mathangwane, Joyce. 1994. *Kalanga lexical database*. Available to download from the Comparative Bantu Online Dictionary. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Kalanga>.
- Mtenje, Al D. 1985. Arguments for an autosegmental analysis of Chicheŵa vowel harmony. *Lingua* 66(1). 21–52.
- Mtenje, Al D. 2001. *Database of “Chewa dictionary”*. Available to download from the Comparative Bantu Online Dictionary. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Chewa>.
- Ngunga, Armindo. 2001. *Yao lexical database*. Available to download from the Comparative Bantu Online Dictionary. <http://www.cbold.ish-lyon.cnrs.fr/Dico.asp?Langue=Yao>.
- Nurse, Derek & Gérard Philippson. 2003. Introduction. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 1–12. London: Routledge.
- Odden, David. 2015. Bantu phonology. In *Oxford Handbooks Online*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199935345.013.59.

11 *Asymmetries in vowel-pair frequencies and height harmony in Bantu*

- R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Robinson, Orrin Warner. 1976. A 'scattered' rule in Swiss German. *Language* 52(1). 148–162.
- Sandstedt, Jade Jørgen. 2019. *A contrastivist approach to the emergence of sound inventories*. Talk given at the 27th Manchester Phonology Meeting, Manchester, UK, 23–25 May. <https://jsandstedt.hcommons.org/publications-presentations>.
- Scullen, Mary Ellen. 1992. Chicheŵa vowel harmony and underspecification theory. *Linguistic Analysis* 22. 218–245.

