Problems in Bantawa Phonology and a Statistically Driven Approach to Vowels

Rachel Vogel

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts in Linguistics

Swarthmore College
December 2015

Abstract

This thesis examines several aspects of the phonology of Bantawa, an endangered and fairly understudied Tibeto-Burman language of Nepal. I provide a brief review of the major literature on Bantawa to date and discuss two particular phonological controversies: one concerning the presence of retroflex consonants, and one concerning the vowel inventory, specifically whether there are six or seven vowel phonemes. I draw on data I recorded from a native speaker to address each of these issues. With regard to the latter, I also provide an in-depth acoustic analysis of my consultant’s 477 vowels and consider several types of statistical models to help address the issue of the number of vowel contrasts. My main conclusions, based on the data from my consultant, are first, that there is evidence based on minimal pairs for a contrast between retroflex and alveolar stops, and second, that there is no clear evidence for a seven-vowel system in Bantawa. With regard to the latter point, additional avenues of research would still be needed to explore the possibility of allophonic variation and/or individual speaker differences.

1. Introduction

There are over 120 languages spoken throughout Nepal, representing four different language families (Lewis et al. 2014; Ghimire 2013). Due to the lasting effects of a long history of social, political, and economic pressures, however, many of these languages are now endangered (Toba et al. 2005). Many languages have also been subject of little to no linguistic scholarship. Bantawa, a Tibeto-Burman language spoken in Eastern Nepal, is one such

---

1 I would like to thank Professor Nathan Sanders for his support and
example. In particular, there has been very little original phonological work on Bantawa and no phonetic analysis that I could find. Moreover, there are a number of gaps and unresolved issues within the phonological literature that does exist, which are compounded by inconsistencies both within a single author’s work and between analyses of different authors. This thesis attempts to clarify some of the major areas of confusion, using new data collected from a native Bantawa speaker (referred to here by her initials CRB) in summer 2015.

First, I provide an overview of Bantawa in its geographic and social context and of the important linguistic work on the language to date in Section 2. I then introduce my consultant in Section 3 and discuss the methods used in our elicitation sessions. In Section 4, I narrow the focus to Bantawa phonology, presenting a synthesis of the key descriptions available and an introduction to two particularly critical points of controversy in the existing literature: the number and quality of coronal plosives and the number and quality of the non-low unrounded vowels, specifically in the central/back region of the vowel space. With regard to the former, I present data from CRB that provide clear evidence for a set of contrastive dental and retroflex stops. The latter issue, however, is much more complex, and I address it in greater detail in Sections 5-7 below.

Because previous descriptions of the Bantawa vowels based on loose phonological transcriptions are at times contradictory, they have led to a fundamental confusion about the overall vowel system. I develop here an
acoustic alternative approach, mapping CRB’s vowels in terms of the general vowel space. On the basis of this acoustic information, I then focus on the question of the number and nature of vowels in the back/central area. There appears to be limited work in phonetics focusing on statistically dividing up clusters of vowel data into phonemes and, as a result, a lack of clear conventions for how to do this. Therefore, in Sections 5 and 6, I explore four possible ways to do so specifically in relation to CRB’s vowels.

In Section 7, I review and discuss the findings related to both controversies. Finally, in Section 8, on the basis of my overview of previous work and my present analysis, I draw two main conclusions. With regard to the coronal consonants, I argue that an inventory reflecting at least the speech of my consultant must include contrastive sets of dentals and retroflexes. With regard to the vowel inventory, I ultimately conclude that the most sensible analysis includes a total of six phonemes, with only a single non-low central vowel. In this section, I also propose avenues of further research into both controversies.

2. Background on Bantawa

2.1. Geographic and Genetic Context

Bantawa (also referred to as Rai Bantawa) is one of Nepal’s approximately 120 languages. It is spoken primarily in the Eastern districts of Dhankuta,
Bhojpur, and Sunsari by an estimated 133,000 people, according to the 2011 Nepali census. (See the boxed in region in Figure 1 below).

Figure 1. Map of Nepal (from Doornenbal 2009: 6)

Bantawa is a member of the Kiranti group of roughly 30 known languages within the larger Tibeto-Burman family. These languages are widely spoken across Eastern Nepal and part of India and generally are considered to fall within the Himalayish branch of Tibeto-Burman languages (Ebert 1994). The exact position of Bantawa within this group, however, is unclear. Several different models have been proposed for the genetic relationships among Kiranti languages. Specifically, Kiranti language scholar Mikhailovsky (1994) proposed a split between Eastern and Western Kiranti, in which Bantawa falls on the Eastern branch. A more recent classification was proposed by Van Driem (2001), however, that splits Kiranti into three branches: Western, Central, and Eastern, with Bantawa falling within Central Kiranti’s Southern subgrouping. Ethnologue nevertheless still follows the
earlier classification of Mickhailovsky (1994). Doornenbal (2009) says that while the 2001 classification is overall an improvement, the three-way split on a single tier fails to indicate that Bantawa, and likely other Central Kiranti languages, share more similarities with the Eastern than the Western languages, so work remains to be done in developing an accurate model. Van Driem’s (2001) classification, as the most recent and up-to-date model, is reproduced in Figure 2 below:

![Kiranti Family Tree](image)

*Figure 2. Kiranti Family Tree (reproduced from Doornenbal 2009:11)*

Bantawa speakers are classified as members of a subgroup of Kiranti peoples in the area, known as Rai. For this reason the Bantawa language is often called Rai Bantawa, and in fact, as the largest of the languages spoken by the Rai, it is also sometimes just called Rai. The term *Rai* is a Nepali exonym, however, and the grouping is based on geography rather than genetic relationship among the languages (Ebert 1994), so I will refer to the language as Bantawa in this thesis.
2.3. Social Context

Like many of Nepal’s languages, Bantawa is significantly challenged by the country’s official language, Nepali, and like almost all Kiranti languages, it is endangered (Ebert 1994, Doornenbal 2009). In fact, Ethnologue classifies Bantawa as 6b “threatened” on the EGIDS scale, according to which languages with numbers beyond 6a are termed endangered. At this point, few Bantawa speakers are monolingual, and most also regularly use Nepali (Lewis et al. 2014). Nevertheless, as the largest Rai language, Bantawa is often used as a lingua franca among other Rai groups in Eastern Nepal. Moreover, it is used for some local radio and television broadcasting as well as in a Bantawa-medium monthly literary magazine called *Bungwakha* (CRB p.c.; Doornenbal 2009). It has also been taught in schools up to class five (the equivalent of fifth grade in the American education system), although school programs are inconsistent due to limited and erratic support from the government (Lewis, et al. 2014; CRB p.c.; Doornenbal 2009). Thus, the status and future of Bantawa in the region are currently in flux.

2.4. Previous Linguistic Work on Bantawa

While linguistic work on Bantawa dates back at least to B. H. Hodgson’s 1857 *Comparative Vocabulary of the Languages of the Broken Tribes of Nepal* (Doornenbal 2009), there has been fairly little more recent work on the language. The most recent substantial contributions based on original fieldwork are Rai’s (1985) PhD Dissertation, *A Descriptive Study of Bantawa,* and Doornenbal’s (2009) MA Thesis, *A Grammar of Bantawa: Grammar,*

There have also been several articles on Bantawa, and larger publications on Kiranti linguistics more generally, primarily based on Rai’s data. These include Ebert’s (1994) *The Structure of Kiranti Languages: Comparative Grammar and Texts*, and Winter’s (2003/2011) Bantawa dictionary. Of the recent publications, it should be noted that Doornenbal’s thesis is the most comprehensive source, and it is this one that I focus on in greater detail below, examining both consistencies and contradictions with other studies as well as with my own data.

3. Data Collection Methodology

Since there is conflicting and somewhat incomplete information on the Bantawa language, as I will discuss in Section 4, I use additional native speaker data to address several specific phonological issues.

Originally I intended to conduct research on a language related to Bantawa in Nepal as the final part of my study abroad program in Spring 2015. Since the program was cut short by the April 2015 earthquake, I developed an alternative approach, working locally in Philadelphia with a native speaker of Bantawa, more common than the language in my original plan. All original data in this thesis come from our elicitation sessions together over the summer.

3.1. Bantawa Consultant
My consultant, Chandrakala Rai Bantawa, who I refer to from now on as CRB, is 58 years old. Like Doornenbal's consultants, she originally comes from Hatuva in the Bhojpur district, however, for the last 30 years, she has been living in Kathmandu. While still in Bhojpur, she worked in Bantawa radio and was involved in the development of a Bantawa language textbook, but, since leaving her hometown, she has primarily used Nepali. In fact, none of her children, now all adults, learned to speak Bantawa growing up. CRB was only temporarily in Philadelphia to visit her son for the summer.

3.2. Data Elicitation

I met with my consultant, CRB, seven times, in her son's apartment. The first meeting was primarily a social visit, during which we got to know one another, and I explained my research. The meetings that followed focused on the recording of both natural speech and specifically elicited words, phrases, and passages in Bantawa. All of the elicitation sessions were conducted in a mixture of Nepali and English and were recorded on my computer using Praat (Boorsma and Weekick 2015). I then transcribed the Bantawa parts using standard IPA notation. An acoustic analysis was also carried out using Praat, the details of which are presented later, in Section 5.

3.2.1. Wordlists

The data used for all of my analyses presented in this thesis come from wordlists, which I elicited with several different approaches. The three
wordlists that I used for elicitation are included in the Appendix at the back of this thesis.

The first wordlist approach was based on Doornenbal’s (2009) grammar. At the end of the grammar, he provides a glossary, and I constructed a wordlist with some of these words as well as some words from examples in his phonology and morphology sections. Since the dialect Doornenbal describes is essentially the same as that spoken by CRB, I expected that she would know most of the words. It turned out, however, that she was not familiar with much of this wordlist. It was not clear if she had forgotten most of the words or had simply never used them.

The second attempt used a 210-word Swadesh list provided by Professor Gasser at Swarthmore College that had been adapted for Austronesian languages by the University of Auckland for the Austronesian Basic Vocabulary Database (http://language.psy.auckland.ac.nz/austronesian/). I was familiar with this list already, having also used it in Prof. Gasser’s Field Methods course in Fall 2014. While there were still many items that CRB did not know, this wordlist worked better than the first approach.

The final approach for collecting individual words used another version of the Swadesh list, which obtained from a researcher at Tribuvan University in Kathmandu, Professor Dan Raj Regmi. This list, also 210 words, was specifically adapted for use with languages of Nepal. As could be
expected, this list, with the most relevant regional and cultural adaptations, was particularly effective for the elicitation of basic data.

3.2.2. Challenges

Throughout the elicitation sessions, I encountered several unexpected challenges that, in the future, could perhaps be avoided or handled differently. We worked in CRB’s son’s apartment, which was next to a busy street and several hospitals, so sounds from cars and sirens came in through the window, especially on hot days when CRB wanted to leave the window open for ventilation. Were I to repeat this research or conduct further research under similar circumstances, it would be important to ensure that the window be properly closed and, if outside noise continued to filter in, to meet in a quieter location.

Additional challenges arose with commotion inside the apartment, when CRB’s son’s friends came in and out, and with sounds from cooking and cell phones. Although I asked CRB to turn off her phone during recordings, she would often forget and then talk to people when they called during elicitions. Another issue that came up was with food. At several meetings, CRB offered me food in the middle of recording and occasionally ate snacks herself while she spoke, including crunchy foods like crackers and cookies. I hesitated to be seen as critical or rejecting her hospitality, so I did not ask her to stop. I later had to discard recordings that were distorted by her eating. Ultimately many of these issues seem to come down to misunderstandings surrounding the research process and the importance of clear recordings. Of
course, because CRB was not familiar with linguistic research, especially phonetics, these issues were not immediately obvious to her. Thus, it would be important in the future to discuss the research process more explicitly than I did. In particular, I think there should be greater focus on how to create a controlled recording environment. A further strategy for eliminating distractions during elicitation sessions and for maintaining a friendly, relaxed atmosphere overall would be to alternate between more strictly defined recording periods with no phones or food on the one hand and break periods with tea and snacks on the other, also allowing time to check phones and to chat.

The last challenge was that since moving to Kathmandu, CRB stopped using Bantawa consistently and as a result forgot some vocabulary. At times, she would get frustrated when we had to skip over words and phrases that she could not remember, but she was eager to reconnect with and share her mother tongue. In fact, it seems that the project helped her start to remember more, and it made her more interested in reconnecting with Bantawa upon returning to Nepal. One strategy that we developed was creating time for CRB to prepare before recordings. We worked through each set of words or sentences twice, first to take notes and practice and then to record the list. CRB also worked hard between meetings, often making notes on what she remembered since the last session and even calling friends and relatives back home to check words with them. Then, she would bring any new information to the start of the next meeting. This strategy proved to
be quite effective, sometimes almost doubling what CRB remembered the first time as well as encouraging her and balancing out the frustration of initially forgetting words.

3.2.3. Transcription

Because Bantawa is an understudied and unstandardized language, it was not immediately clear how to most effectively transcribe the words. The approach developed for this research followed four main steps. First, I looked for a recognizable and productive affix in the word. For example, /wa/ morphemes are quite common to denote relatedness to water or fowl, and to mark male nouns (Doorenbal 2009).

Second, I checked CRB’s notes, which she wrote in the same Devanagari that she uses for Nepali. This approach was definitely problematic in cases where Bantawa phonemes do not correspond one-to-one with Devanagari symbols or with Nepali, and the writing should not be the basis of phonological transcription on its own. Nevertheless, it was valuable in certain contexts. In particular, I gave special attention to CRB’s transcriptions when I knew that my intuitions as a native English speaker were likely to interfere, for example, in identifying aspiration contrasts in stops and affricates.

Third, I consulted Doornenbal’s (2009) grammar and Winter’s (2011) dictionary; even if the full forms of words in these sources differed from mine, they were often helpful for identifying roots. While Doornenbal’s and Winter’s transcriptions also sometimes differed from mine and were not
consistent with CRB’s speech, when the first two strategies failed or were not relevant, this approach could be useful, at least as a point of comparison with my own intuitions.

Finally, if these three strategies were not applicable, I based decisions only on my auditory judgments. Of course, my judgments were also always considered alongside the information from the other three steps.

4. Overview of Bantawa Phonology

The focus of the data analysis in the present thesis is the phonological system of Bantawa. As mentioned, the most thorough recent source on the topic to date is Doornenbal (2009). Therefore, I briefly summarize his description of the core properties of Bantawa phonology that are relevant for my investigation. I then discuss two major areas of controversy, specifically with regard to the consonant and vowel inventories. In this context, I also introduce some observations based on the data I elicited from CRB.

4.1. Syllable Structure

Doornenbal (2009) describes Bantawa’s syllable structure as \( C_1(G)V(C_2) \). \( C_1 \) can be any consonant in the language in syllables with simple onsets, or a fricative in syllables with complex onsets. \( G \) represents the Bantawa glides (/j/ and /w/), and \( C_2 \) is restricted to oral and nasal stops. For example, the first syllable of the word \text{hwa}t\text{ni} \ ‘over there’ is maximally complex, with a cluster in the onset and with a coda.

4.2. Consonants
There is considerably less certainty surrounding the phonemic inventory of Bantawa. With regard to the consonants, Doornenbal’s inventory of native phonemes is shown in Table 1. It should be noted that some non-standard symbols that Doornenbal uses have been changed to match the IPA notation used elsewhere in this thesis.

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Plosive</em></td>
<td>p, b</td>
<td>t, d</td>
<td>k, g</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pʰ, bʰ</td>
<td>tʰ, dʰ</td>
<td>kʰ, gʰ</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Nasal Stop</em></td>
<td>m</td>
<td>n</td>
<td></td>
<td>η</td>
<td></td>
</tr>
<tr>
<td><em>Fricative</em></td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td>h²</td>
</tr>
<tr>
<td><em>Affricate</em></td>
<td>ts, dz</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tsʰ, dzʰ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Glide</em></td>
<td>w</td>
<td></td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Liquid (lateral)</em></td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Trill</em></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Bantawa Consonants (based on Doornenbal 2009).

The major area of uncertainty regarding Bantawa consonants, which Ebert (1994) refers to in her overview of Kiranti phonology, concerns the status of retroflex consonants. Doornenbal claims that, with the exception of Nepali loans, Bantawa shows no contrast between dental and retroflex stops. Instead, he lists one set of four alveolar oral stops: /t, d, tʰ, dʰ/ for native Bantawa, but this issue is not agreed upon by all researchers. Specifically, Rai (1985) observes a contrast between two sets of coronals, which he calls

---

2 Doornenbal (2009) originally lists this as pharyngeal, however, according to both Rai (1985) and my consultant CRB, it is actually glottal. I have chosen to reflect this second classification in my chart (Table 1).
apico-dental and apico-alveolar, the latter of which Doornenbal (2009)
describes as retroflex. Rai does not however list minimal pairs for the
contrast.

The issue of the retroflex consonants is also the most striking
difference between Doornenbal’s inventory and the data I elicited from my
consultant CRB. Contrary to Doornenbal’s analysis, CRB has maintained
that there is in fact a dental/retroflex contrast for oral stops, and her choice
of coronal remained consistent for any given word both within a single
elicitation session and from one day to another. She also regularly corrected
me when I replaced one coronal with the corresponding sound in the other
place of articulation (i.e. replacing /t/ for /t/ or /d/ for /d/).

Furthermore, I have identified a clear minimal pair in CRB’s
translation of the wordlist from Tribhuvan University illustrating the
contrast between dental and retroflex voiceless stops: /t/ and /t/ [tan]
meaning ‘head’ and [tan] meaning ‘hair’. In the wordlist in the appendix to
his grammar, Doornenbal lists both with an alveolar ([tan]), however, I found
no instances at all of alveolar plosives in my data.

4.3. Vowels

Doornenbal’s inventory of vowel phonemes is shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td></td>
<td>o</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Bantawa Vowels (based on Doornenbal 2009).³

As with the consonants, there are some discrepancies in the analyses of the vowels in the previous literature. This seems to be due to two main issues. First, because there has been little work on patterns of allophony, it is not always clear which differences are phonemic or predictable. Second, since there is no available phonetic analysis, all approximations of the vowel space to date have been based solely on impressionistic transcriptions of previous authors.

Doornenbal's grammar appears to offer the most thorough description of vowel alternations. These include the optional backing of vowels after velar consonants as well as a high degree of variability in the realization of the central vowel /i/, the patterns of which he says are not easy to define but are related to features of adjacent syllables. Nevertheless, the domain of these patterns, in terms of when they occur and whether they are obligatory, is often not clearly defined.

Regarding the basic vowel space itself, while the available phonological descriptions seem to agree on five of the vowels (/i e u o a/), the number and quality of the additional phoneme(s) remain unclear. Specifically, this issue is seen with the non-low unrounded vowel(s) in the central/back region of the vowel space. There seem to be several conflicting

³ In addition to the six monophthongs shown in Table 2, Doornenbal identifies a diphthong /eɪ/, which he says occurs only rarely in native Kiranti words.
views on the matter, and the issue is further confused by differences in the
notations used by each author. Rai (1985) lists a total of seven vowels, with a
distinction between a central vowel that he writes as /A/ and an unrounded
back vowel that he writes as /U/. In her overview of existing work on
Bantawa phonology, Ebert (1994) chooses to re-transcribe these as /a/ and
/i/ respectively. Doornenbal (2009) and Winter (2003/2011), on the other
hand, both only list six vowels, with a single vowel in the region in question:
Winter’s/ü/ and Doornenbal’s /i/. Both authors describe these as
phonemically high and central, although Doornenbal also allows for a varied
realization of this vowel. According to Ebert, several other authors, including
Hansson (1991), have also listed back unrounded vowels, transcribed as
/u/. Table 3 below presents an overview of these different authors’
notations with corresponding descriptions when they are available.

<table>
<thead>
<tr>
<th>Source</th>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doornenbal 2009</td>
<td>i</td>
<td>High central but can be realized anywhere in central region</td>
</tr>
<tr>
<td>Winter 2011</td>
<td>ü</td>
<td>High central; note that Ebert (1994) describes it as back unrounded</td>
</tr>
<tr>
<td>Rai 1985</td>
<td>A</td>
<td>Mid central unrounded</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>Back unrounded</td>
</tr>
<tr>
<td>Ebert 1994</td>
<td>(a)</td>
<td>Central mid with marginal phonemic status (only in borrowings)</td>
</tr>
<tr>
<td></td>
<td>i/u</td>
<td>Back unrounded</td>
</tr>
<tr>
<td>Hansson 1991</td>
<td>i</td>
<td>Back unrounded</td>
</tr>
</tbody>
</table>

Table 3. Non-low back/central unrounded vowels by author and notation.

While it is possible that the inconsistencies in the descriptions of the
evowel system are attributable in part to differences in notation, such stylistic
differences cannot alone account for the discrepancies in the number of
reported vowels. That such a controversy persists indicates that the previous methods of elicitation and the intuitions of speakers and researchers used for transcription may not have been adequate to precisely determine the organization of the vowel space.

In the rest of this thesis, I therefore take an alternative approach, based on acoustic analysis of the vowels in CRB’s speech to not only map the location of Bantawa vowels in the vowel space but also investigate the controversy surrounding the status of the non-low unrounded central/back vowels.

5. Acoustic Analysis Methodology

Given the absence of phonetic work on Bantawa in general, and the crucial need for such information in order to understand the vowels of the language, I have conducted an acoustic analysis of the vowels in CRB’s speech. This allows me to propose a more confident map of the vowel space, and it will make room for more detailed analysis of allophonic variation moving forward.

The vowel analysis I present uses only data from the wordlist elicitations to avoid possible articulation changes that might arise in connected speech. In developing my acoustic analysis of the Bantawa vowel system, as described below, I first made measurements using Praat. I then plotted the vowels in the vowel space using the first and second formants. Finally, as I explain in detail in Section 6, I conducted a series of statistical
analyses, also including the third formant, to arrive at my final proposal for
the vowel phonemes of Bantawa, at least as found in CRB’s speech.

5.1. Data Measurement

In order to make all the necessary vowel measurements, I first manually
created textgrids in Praat for the relevant sound files. I then added several
interval tiers. On the first tier, I labeled each Bantawa monophthong vowel; I
used /a/ for any non-low mid/back unrounded vowel. I then labeled the
entire Bantawa word and the English translations on the second and third
tiers, respectively. The format of these textgrids is illustrated in Figure 3
for the Bantawa word [lad̩ɪtma] meaning ‘moon.’

Figure 3. Segmented and annotated textgrid for the word [lad̩ɪtma] ‘moon.’

After discarding vowels that had been too reduced to reliably measure
or had been distorted by background noise, 477 vowels remained.
Measurements of these vowels were then extracted from the textgrids using a script in Praat. The script first divided the vowels (identifiable in Praat as any segment with a label on the vowel tier) into five equal length segments. Then, it made the following measurements for each portion and for the entire vowel: duration and averages of the first three formants (F1, F2, and F3).

5.2. Data Analysis

The vowel measurements extracted in the steps above were then used to model and compare the feasibility of two competing conceptions of the vowel space – one with six phonemes and one with seven. For each model, I created a scatterplot with the R programming language (R Development Core Team 2015) mapping the distribution of vowels according to their F1 and F2. The mean F1 and F2 for each phoneme were also calculated and marked on the same graph. I used measurements only from the middle fifth portion of the vowels for the scatterplots in order to avoid formant transitions due to adjacent sounds.

The scatterplot for the six-vowel system is provided here in Figure 4. As can be seen, each phoneme is plotted in a different color, with blue representing /i/, red representing /e/, green /a/, black /u/, purple /o/, and orange /o/. Individual points for each realization of a phoneme are marked by the IPA symbol for that sound. Large dots mark the mean F1 and F2 for each phoneme in the corresponding color.
For the seven-vowel system, I divided the cluster of /æ/ vowels into two using a K-Means Clustering algorithm, as defined by MacQueen (1967) and previously adopted for phonetics by Hall and Smith (2006) as a method of identifying vowel phonemes Greek. This algorithm divides a cluster of points into a specified number of smaller clusters – in this case two – in such a way that it minimizes the distance between each point and the center of its respective cluster. In this system, vowels previously considered to be /æ/ are categorized as either high unrounded /u/ or mid central /ʌ/, following the classification of Rai’s seven-vowel system.

6. Results: Comparison of 6 and 7 Vowel Systems

In this section, I present the scatterplots I created for the models of the Bantawa vowel space. Specifically, I compare the vowel distribution in a six vs. seven-vowel system, focusing on the status of the non-low central/back unrounded vowels.

6.1. Six-Vowel System
Figure 5 provides another scatterplot for the six-vowel system. While Figure 4 showed the distribution of all 477 vowels in the data, this plot only includes individual points for those classified as /a/ to facilitate the examination of this vowel in particular.

![Figure 5. Distribution of /a/ within a six-phoneme model.](image)

As seen in the six-phoneme system modeled in Figure 5, the vowels labeled as /a/ are widely dispersed. That is, the F2 for /a/ ranges from less than the average values for back vowels to greater than the average values for front vowels. In addition, the F1 ranges from less than the average values for high vowels to greater than the average for /a/. Given such a wide range in the vowel's realization, we could question whether it really is a single phoneme. Next, I therefore consider whether a seven-phoneme analysis can better account for the data.

6.2. Seven-Vowel System
In this section, to create a set of seven total phonemes, the vowels previously categorized as a single /a/ are divided into two K-Means clusters. I label these clusters as /u/ and /ʌ/, following Rai’s description of one high unrounded back vowel and one mid central vowel. I also assess the potential effects that different types of compression of the formant values have on the analysis. I conduct t-tests in R to determine whether the new clusters are statistically different from surrounding vowels in any of their first three formants. Specifically, these tests compare the following vowel pairs: /u/ ~ /u/, /u/ ~ /o/, /ʌ/ ~ /o/, and /ʌ/ ~ /a/. While the first two formants are used to plot the vowels in the vowel space, the third formant, typically associated with rounding, is included in the t-tests as a possible additional distinguishing characteristic. It should be noted that the K-Means clustering was run with and without F3, and the results were essentially the same.

6.2.1. Seven-Vowel Model Using Raw Hz

The scatterplot in Figure 6 shows the distribution of vowels after the original /a/ has been divided into two K-Means clusters corresponding to /u/ and /ʌ/. For this first seven-vowel model, the first and second formant values used are in raw Hz. As above, the phoneme symbols represent individual realizations of a vowel, and solid colored circles represent the mean F1 and F2 for each phoneme. The orange circle previously corresponding to /a/ in the six-vowel model has been replaced by a tan circle representing /u/ and an orange circle representing /ʌ/.
Table 4 provides the t-test results for each vowel comparison, showing the probability that the vowels are in fact statistically different from one another in each of their first three formants. The standard confidence interval for statistical difference is a p-value of ≤ 0.05, however, it should be noted that the error rate is compounded for analyses with multiple points of comparison. With four pairs of phonemes tested in terms of their F1, F2, and F3, there are 12 points of comparison. I therefore use a Bonferroni correction here, and in the following analyses, to more conservatively assess the p-values. In this correction, a new critical p-value is calculated for the overall comparison by dividing the conventional 0.05 by the number of comparisons, in this case 0.05/12, which is roughly 0.004. In Table 4, and the tables to follow, cells with p-values less than or equal to the confidence interval after a Bonferroni correction are shaded in green.
As Table 4 shows, even with the Bonferroni correction, both /u/ and /ʌ/ are statistically different from their neighboring phonemes at least in terms of their F1.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/</td>
<td>3.000e-11</td>
<td>0.164</td>
<td>0.152</td>
</tr>
<tr>
<td>/u/</td>
<td>2.389e-05</td>
<td>0.698</td>
<td>0.223</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>5.573e-13</td>
<td>1.968e-08</td>
<td>0.035</td>
</tr>
<tr>
<td>/ʌ/</td>
<td>0.001</td>
<td>0.792</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 4. P-values for raw K-Means clusters.

It is important to note here that although /u/ and /ʌ/ are statistically different from the neighboring more established phonemes, the positions of these new clusters in Figure 6 do not correspond to a high ~ mid split as described by Rai (1985). Instead, they differ more in F2 (on the x-axis of the plot), indicating a central ~ back division. The division on this dimension is particularly large in this case because F2 spans a much larger range in Hz than does F1. Hz are not, however, reflective of auditory perception, so a raw change in Hz for F2 is not as perceptible as the same change for F1. As a result, clusters that appear distinct from one another in F2 using Hz do not necessarily reflect a perceptible contrast (Hall and Smith 2006; Sanders and Padgett 2008).

One way to compensate for this issue is to compress F2 in some way before generating new divisions in the data. I therefore next apply several common strategies of compression in conjunction with creating a seven-phoneme model.
6.2.2. Seven-Vowel Model Using Log Compression

The first approach to compression that I consider is taking the log of both F1 and F2, which brings the scale of F2 much closer to that of F1. The scatterplot in Figure 7 shows a plot of the vowel space when /a/ is divided into two K-means clusters using these log values instead of the original formant measurements in raw Hz. Table 5 shows the results of the t-tests on the newly defined clusters.

![Figure 7. Seven-vowel system with log-compressed K-Means clustering](image)

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/ = /u/?</td>
<td>0.885</td>
<td>0.241</td>
<td>0.534</td>
</tr>
<tr>
<td>/u/ = /o/?</td>
<td>1.650e-05</td>
<td>0.725</td>
<td>0.051</td>
</tr>
<tr>
<td>/ʌ/ = /o/?</td>
<td>2.526e-32</td>
<td>2.495e-05</td>
<td>0.904</td>
</tr>
<tr>
<td>/ʌ/ = /a/?</td>
<td>0.002</td>
<td>0.002</td>
<td>4.187e-07</td>
</tr>
</tbody>
</table>

Table 5. P-values for log-compressed K-Means clusters

As can be seen in Figure 7, using log-compressed formant values for the K-Means clusters gives us a high ~ mid division between /u/ and /ʌ/ much more inline with the system described by Rai (1985). As the t-test results show in Table 5, however, in this model, the K-Means cluster...
classified as /u/ is not statistically different from /u/ in any of its first three formants, and if anything, on the scatterplot it appears slightly farther back than /u/, which seems unlikely. This suggests that using log-compressed data does not yield more insight into a possible seven-phoneme system in Bantawa.

### 6.2.3. Seven-Vowel Model with Bark Compression

The next approach to F2 compression that I consider involves the conversion of Hz into a perception-based scale proposed by Zwicker (1961) and referred to as Bark. I used the following algorithm to convert my data:

\[
\text{Bark} = 13 \arctan(0.00076f) + 3.5 \arctan \left( \frac{f}{7500} \right)^2 \]  

(Traunmüller 1990), in which \( f \) is in Hz. I then ran the K-Means clustering algorithm on these values to again divide /a/ into two clusters. A new scatterplot was created, and t-tests were run again. The results are shown in Figure 8 and Table 6.

![Figure 8. Seven-vowel system with Bark scale K-Means clustering.](image)

Figure 8. Seven-vowel system with Bark scale K-Means clustering.
Table 6. P-values for Bark scale K-means clusters.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/ = /u/?</td>
<td>2.884e-06</td>
<td>0.101</td>
<td>0.745</td>
</tr>
<tr>
<td>/o/ = /o/?</td>
<td>0.085</td>
<td>0.908</td>
<td>0.150</td>
</tr>
<tr>
<td>/e/ = /o/?</td>
<td>1.337e-29</td>
<td>5.228e-08</td>
<td>0.804</td>
</tr>
<tr>
<td>/A/ = /a/?</td>
<td>0.085</td>
<td>0.530</td>
<td>2.117e-06</td>
</tr>
</tbody>
</table>

It can be seen in Figure 8 that upon running the K-Means algorithm on the data once it was converted to the Bark scale, the division between /u/ and /A/ is in between the central ~ back division seen using raw Hz and the high ~ mid division with the log compression. According to the t-test results, /A/ is statistically different from both /o/ and /a/ in at least one formant, and /u/ is statistically different from /u/ in its first formant, but /u/ is no longer different from /o/ in any of the first three formants. In fact, it can also be seen in Figure 8, that there is a great deal of overlap both between /u/ and /o/ and the /A/ and /a/. Therefore, as was the case with the log compression, this model does not yield a clear cut seven-vowel system.

6.2.4. Seven-Vowel Model with Mel compression

The final approach I consider uses the mel scale, another commonly used perception-based scale, developed by Stevens (1937). I converted the data using the following algorithm: mel = 2595log10(1+f/700) (O'Shaughnessy 1987), in which f is in Hz. Next, I divided /a/ again into two K-Means clusters corresponding to /u/ and /A/. As before, a scatterplot was created and t-tests were run on the relevant pairs of phonemes. The new plot and results for this model are provided in Figure 9 and Table 7.
The t-test results here indicate that both /u/ and /ʌ/ are again statistically different from their neighboring phonemes in at least one formant. As can be seen in the scatterplot, the split between /u/ and /ʌ/ has once again become between central and back and thus unlike the high ~ mid division originally expected.

7. Discussion: Controversies in the Bantawa Consonant and Vowel Inventories

As discussed in the previous sections, there are controversies regarding the inventories of both the consonants and vowels of Bantawa. The data I recorded of CRB’s speech allows me to address both of these issues.
7.1. **Consonant Controversy**

As explained previously, the primary challenge in existing literature with regard to the consonant phonemes of Bantawa has to do with the coronal consonants. While my main resource from the existing literature (Doornenbal 2009) describes the system as having solely alveolar stops, another source, Rai (1985), who Doornenbal challenges, includes a contrastive series of retroflex stops. The wordlist data I elicited from my CRB contains both retroflex and dental stops, including a clear minimal pair [taŋ] and [tan] and thus offers support for the system proposed by Rai.

7.2. **Vowel Controversy**

The main challenge with regard to the vowel system in Bantawa concerns the number of phonemes, in particular the number of non-low unrounded central/back vowels. In the previous literature, Doornenbal's (2009) analysis presents six phonemes with a single non-low central vowel, while Rai's (1985) presents seven, with two contrasting non-low unrounded central/back vowels. In order to investigate the contrasts among the vowels, I turned to acoustic analysis, and specifically to the measurement of the vowel formants. In particular, I considered the first two formants, which provide information about the height and frontness of the vowels, as well as the third formant, which correlates with rounding.

In the two previous sections, I examined the nature of the vowel system by plotting and comparing competing models of the vowel space based on CRB's data. First, the six-vowel system with a single non-low
central phoneme /a/ was plotted. A potential problem was seen for this distribution, in the broad dispersion of the central vowels.

I then considered the possibility suggested by Rai that in fact there are two phonemes in this area— one high and one mid – written here as /u/ and /ʌ/. To this end, I used K-Means clustering to predict where two distinct clusters corresponding to /u/ and /ʌ/ might lie. In addition to modeling this system in raw Hz, I considered models using three compressions of F2, a log compression and conversions to the Bark and mel scales, considered to correspond more closely to human perception, to see if any of these options proved more informative. Thus, I created plots for four seven-vowel models in total.

While the seven-vowel systems plotted using different compressions differed in terms of where the centers of /u/ and /ʌ/ fell and the degree of overlap between these and the surrounding vowels, no clear patterns arose to favor any particular seven-vowel system along the lines proposed by Rai (1985). In addition to the visual indication of the various distinctions and overlaps seen in the scatterplots in Figures 6 – 9, t-tests showed variation in which of the three formants were significantly different according to the different compressions.

These results are summarized in Table 7 for the four models of clustering examined. For each comparison, = means that the values were not found to be different (i.e., the p value was insignificant), and ≠ means that the values were statistically different.
Table 7. Summary of T-Test Results for Vowel Contrasts with Different Models.

As can be seen in this summary table, no clear patterns of statistical significance emerge to favor any particular model. That is, none of the models has a clear concentration of statistically significant p values to offer unambiguous support for seven distinct phonemes similar to Rai’s described system.

Of the three models using compressed F2 values, only the version using the mel scale offers a distribution in which the K-Means clusters corresponding to both /u/ and /ʌ/ are statistically different from neighboring phonemes in at least one of their formants. The distribution in this version, however, includes a rather central ~ back division between /u/ and /ʌ/ instead of the high ~ mid division expected according to Rai (1985). While the other two compressed seven-vowel models offer the expected high
~ mid split, this comes at the cost of /u/’s statistical difference from one of its neighbors (either /o/ or /u/), reducing the number of distinct phonemes back to six. Moreover, in terms of the clusters’ physical locations as reflected in the scatterplots, all three models using F2 compression appear to include a good deal of crowding and overlap. The overlap is especially noteworthy between /ʌ/ and /a/ in the plots for Bark and mel, in which the two vowels differ only in F3. In fact, appropriate interpretation of a statistical difference in F3 between /ʌ/ and /a/ is itself questionable, as the phonemes are generally understood to both be central and unrounded.

Although there was a problem with the six-vowel model with the central vowel /a/ appearing so widely dispersed, in the absence of clear evidence in support of a seven-vowel system, the best solution seems to be to retain the simpler inventory as described by Doornbal (2009). Further research must be done with additional data, however, to better understand the large dispersion of the /a/. One possibility is that there is allophonic variation that leads to slightly different positioning of the central vowel or to its merging with another phoneme under certain conditions. Another possibility is that my consultant used more variability than other speakers might use because she has not consistently spoken Bantawa in recent years. Thus, a larger corpus with speech from several speakers would be needed to examine these issues further.
8. Conclusions

In this thesis, I discussed the phonology of the endangered language, Bantawa, spoken in Nepal. Because there is minimal literature on this language, especially literature based on original data, I collected new data from a native speaker, CRB, and used this to further investigate several key issues in Bantawa phonology. After discussing how I collected a corpus of speech from my consultant, I presented a brief summary of the phonological literature on the language to date and drew attention to two controversies in the phoneme inventory: one with consonants and one with vowels.

With regard to the consonants, I focused on the question of the total number of coronal stops in Bantawa and the controversy between Rai (1985) and Doornenbal (2009) concerning the status of retroflex stops in particular. Data from my work with CRB, including a clear minimal pair between [ṭan] ‘hair’ and [ṭan] ‘head’ found in the wordlist elicitations, strongly supports a dental/retroflex contrast. In light of these results, however, Doornenbal’s explicit argument against such a contrast following his work with speakers from the same area is particularly surprising. It is unlikely, since Doornenbal’s thesis is so recent, that a generational shift could account for the difference, but it is perhaps possible that there are dialectal differences even within the region of Hatuva. A second possibility is that CRB’s speech has changed since moving to Kathmandu, and her dental/retroflex contrast is a result of Nepali influence. Work with more speakers, either from CRB’s community or elsewhere, would be valuable in exploring the issue of coronal
stops further to see whether the contrast extends beyond CRB’s own idiolect, and in particular, to speakers who continue to use Bantawa more regularly.

With respect to vowels, the transcription methods used by previous authors were not precise or consistent enough to create a clear sense of the phoneme inventory in Bantawa. The limitations of these previous methods led me to explore more phonetic and statistically driven alternative approaches to investigate contrasts within CRB’s vowels. In particular, I focused on addressing the controversy and uncertainty surrounding the number and quality of non-low central and/or unrounded back vowels. Specifically, I compared Doornenbal’s (2009) six-vowel analysis including a single central phoneme written as /a/ in this thesis and Rai (1985)’s seven-vowel analysis with two contrasting phonemes, written here as /u/ and /ʌ/ in place of Doornenbal’s /a/.

In sections 5 and 6, I presented scatterplots based on measurements of the first two formants of 477 vowels from CRB’s wordlist elicitations, to visualize the overall distribution of vowels in CRB’s vowel space and to model the competing analyses of the phoneme inventory. For the seven-phoneme analysis, I divided the non-low central vowels, all transcribed as /ə/ in the six-phoneme analysis, into two K-Means clusters. I created several versions of this model and tested the degree of overlap between the K-Means clusters and neighboring vowels using a series of t-tests.

It was seen that while the six-phoneme analysis resulted in the wide dispersion of /ə/, none of the seven-phoneme models offered a particularly
desirable outcome either. Specifically, there was no seven-phoneme model in which the two K-Means clusters were statistically different from their neighbors and also that reflected Rai's (1985) high ~ mid division between /u/ and /a/. Moreover, the overlap and crowding in the seven-phoneme scatterplots explored in this thesis raises the question of whether such a model is in fact characteristic of meaningful contrasts. In sum, we are led to prefer the simpler analysis of CRB's Bantawa in which there is only a single non-low central vowel and six vowel phonemes in total.

A logical next stage to this research would be to look further into a scenario with conditioned variation of the central vowel, perhaps leading to two allophones resembling [u] and [a], as this might account for some of the dispersion of /a/ observed with the six-phoneme model. It would also be important in future work to reconsider the treatment of clear outliers in the data, for instance vowels transcribed here as /a/ that appear as far forward as the front vowel /e/. One possibility for the future would be to look only at points within a certain number of standard deviations from the mean.

Given the degree of apparent crowding and overlap in the seven-phoneme models, the relationship between statistical findings and their relevance in the actual perception and production of Bantawa speakers who have ongoing contact with other speakers should also be examined in greater detail. Again, like with the consonants, work with additional speakers would be valuable, especially for comparison with Bantawa speakers who have had
somewhat less exposure to Nepali or at least who continue to primarily use Bantawa in daily life.
APPENDIX

• Wordlist 1. Targeted words from Doornenbal’s (1991) grammar
  snow
day after tomorrow
witch
darkness
tomato
rebel
basket
raspberry
sweet potato
cardamom
agitation
thunderstorm
explode
nephew
niece
storage basket
paternal aunt
knowledgeable
people
tabacco
how many
donot kill
donot place it
to walk
to forget
prayer
he sent (it)

• Wordlist 2. 210-Word Swadesh List from the Austronesian Basic Vocabulary Database
  Available online at http://language.psy.auckland.ac.nz/austronesian/

• Wordlist 3. 210-Word Swadesh List from Tribhuvan University’s Linguistic Survey of Nepal
  1 body 71 rice (husked) 141 far
  2 head 72 potato 142 big
  3 hair 73 eggplant 143 small
  4 face 74 groundnut 144 heavy
  5 eye 75 chili 145 light
  6 ear 76 turmeric 146 above
  7 nose 77 garlic 147 below
<table>
<thead>
<tr>
<th>8</th>
<th>mouth</th>
<th>78</th>
<th>onion</th>
<th>148</th>
<th>white</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>teeth</td>
<td>79</td>
<td>cauliflower</td>
<td>149</td>
<td>black</td>
</tr>
<tr>
<td>10</td>
<td>tongue</td>
<td>80</td>
<td>tomato</td>
<td>150</td>
<td>red</td>
</tr>
<tr>
<td>11</td>
<td>breast</td>
<td>81</td>
<td>cabbage</td>
<td>151</td>
<td>one</td>
</tr>
<tr>
<td>12</td>
<td>belly</td>
<td>82</td>
<td>oil</td>
<td>152</td>
<td>two</td>
</tr>
<tr>
<td>13</td>
<td>arm/hand</td>
<td>83</td>
<td>salt</td>
<td>153</td>
<td>three</td>
</tr>
<tr>
<td>14</td>
<td>elbow</td>
<td>84</td>
<td>meat</td>
<td>154</td>
<td>four</td>
</tr>
<tr>
<td>15</td>
<td>palm</td>
<td>85</td>
<td>fat (of meat)</td>
<td>155</td>
<td>five</td>
</tr>
<tr>
<td>16</td>
<td>finger</td>
<td>86</td>
<td>fish</td>
<td>156</td>
<td>six</td>
</tr>
<tr>
<td>17</td>
<td>fingernail</td>
<td>87</td>
<td>chicken</td>
<td>157</td>
<td>seven</td>
</tr>
<tr>
<td>18</td>
<td>leg</td>
<td>88</td>
<td>egg</td>
<td>158</td>
<td>eight</td>
</tr>
<tr>
<td>19</td>
<td>skin</td>
<td>89</td>
<td>cow</td>
<td>159</td>
<td>nine</td>
</tr>
<tr>
<td>20</td>
<td>bone</td>
<td>90</td>
<td>buffalo</td>
<td>160</td>
<td>ten</td>
</tr>
<tr>
<td>21</td>
<td>heart</td>
<td>91</td>
<td>milk</td>
<td>161</td>
<td>eleven</td>
</tr>
<tr>
<td>22</td>
<td>blood</td>
<td>92</td>
<td>horns</td>
<td>162</td>
<td>twelve</td>
</tr>
<tr>
<td>23</td>
<td>urine</td>
<td>93</td>
<td>tail</td>
<td>163</td>
<td>twenty</td>
</tr>
<tr>
<td>24</td>
<td>feces</td>
<td>94</td>
<td>goat</td>
<td>164</td>
<td>one hundred</td>
</tr>
<tr>
<td>25</td>
<td>village</td>
<td>95</td>
<td>dog</td>
<td>165</td>
<td>who</td>
</tr>
<tr>
<td>26</td>
<td>house</td>
<td>96</td>
<td>snake</td>
<td>166</td>
<td>what</td>
</tr>
<tr>
<td>27</td>
<td>roof</td>
<td>97</td>
<td>monkey</td>
<td>167</td>
<td>where</td>
</tr>
<tr>
<td>28</td>
<td>door</td>
<td>98</td>
<td>mosquito</td>
<td>168</td>
<td>when</td>
</tr>
<tr>
<td>29</td>
<td>firewood</td>
<td>99</td>
<td>ant</td>
<td>169</td>
<td>how many</td>
</tr>
<tr>
<td>30</td>
<td>broom</td>
<td>100</td>
<td>spider</td>
<td>170</td>
<td>which</td>
</tr>
<tr>
<td>31</td>
<td>mortar</td>
<td>101</td>
<td>name</td>
<td>171</td>
<td>this</td>
</tr>
<tr>
<td>32</td>
<td>pestle</td>
<td>102</td>
<td>man</td>
<td>172</td>
<td>that</td>
</tr>
<tr>
<td>33</td>
<td>hammer</td>
<td>103</td>
<td>woman</td>
<td>173</td>
<td>these</td>
</tr>
<tr>
<td>34</td>
<td>knife</td>
<td>104</td>
<td>child</td>
<td>174</td>
<td>those</td>
</tr>
<tr>
<td>35</td>
<td>axe</td>
<td>105</td>
<td>father</td>
<td>175</td>
<td>same</td>
</tr>
<tr>
<td>36</td>
<td>rope</td>
<td>106</td>
<td>mother</td>
<td>176</td>
<td>different</td>
</tr>
<tr>
<td>37</td>
<td>thread</td>
<td>107</td>
<td>older brother</td>
<td>177</td>
<td>whole</td>
</tr>
<tr>
<td>38</td>
<td>needle</td>
<td>108</td>
<td>younger brother</td>
<td>178</td>
<td>broken</td>
</tr>
<tr>
<td>39</td>
<td>cloth</td>
<td>109</td>
<td>older sister</td>
<td>179</td>
<td>few</td>
</tr>
<tr>
<td>40</td>
<td>ring</td>
<td>110</td>
<td>younger sister</td>
<td>180</td>
<td>many</td>
</tr>
<tr>
<td>41</td>
<td>sun</td>
<td>111</td>
<td>son</td>
<td>181</td>
<td>all</td>
</tr>
<tr>
<td>42</td>
<td>moon</td>
<td>112</td>
<td>daughter</td>
<td>182</td>
<td>to eat</td>
</tr>
<tr>
<td>43</td>
<td>sky</td>
<td>113</td>
<td>husband</td>
<td>183</td>
<td>to bite</td>
</tr>
<tr>
<td>44</td>
<td>star</td>
<td>114</td>
<td>wife</td>
<td>184</td>
<td>to be hungry</td>
</tr>
<tr>
<td>45</td>
<td>rain</td>
<td>115</td>
<td>boy</td>
<td>185</td>
<td>to drink</td>
</tr>
<tr>
<td>46</td>
<td>water</td>
<td>116</td>
<td>girl</td>
<td>186</td>
<td>to be thirsty</td>
</tr>
<tr>
<td>47</td>
<td>river</td>
<td>117</td>
<td>day</td>
<td>187</td>
<td>to sleep</td>
</tr>
<tr>
<td>48</td>
<td>cloud</td>
<td>118</td>
<td>night</td>
<td>188</td>
<td>to lie</td>
</tr>
<tr>
<td>49</td>
<td>lightening</td>
<td>119</td>
<td>morning</td>
<td>189</td>
<td>to sit</td>
</tr>
<tr>
<td>50</td>
<td>rainbow</td>
<td>120</td>
<td>noon</td>
<td>190</td>
<td>to give</td>
</tr>
<tr>
<td>51</td>
<td>wind</td>
<td>121</td>
<td>evening</td>
<td>191</td>
<td>to burn</td>
</tr>
<tr>
<td>52</td>
<td>stone</td>
<td>122</td>
<td>yesterday</td>
<td>192</td>
<td>to die</td>
</tr>
<tr>
<td>53</td>
<td>path</td>
<td>123</td>
<td>today</td>
<td>193</td>
<td>to kill</td>
</tr>
<tr>
<td>54</td>
<td>sand</td>
<td>124</td>
<td>tomorrow</td>
<td>194</td>
<td>to fly</td>
</tr>
<tr>
<td>55</td>
<td>fire</td>
<td>125</td>
<td>week</td>
<td>195</td>
<td>to walk</td>
</tr>
<tr>
<td>56</td>
<td>smoke</td>
<td>126</td>
<td>month</td>
<td>196</td>
<td>to run</td>
</tr>
<tr>
<td>57</td>
<td>ask</td>
<td>127</td>
<td>year</td>
<td>197</td>
<td>to go</td>
</tr>
<tr>
<td>58</td>
<td>mud</td>
<td>128</td>
<td>old</td>
<td>198</td>
<td>to come</td>
</tr>
<tr>
<td>59</td>
<td>dust</td>
<td>129</td>
<td>new</td>
<td>199</td>
<td>to speak</td>
</tr>
<tr>
<td>60</td>
<td>gold</td>
<td>130</td>
<td>good</td>
<td>200</td>
<td>to hear/listen</td>
</tr>
<tr>
<td>61</td>
<td>tree</td>
<td>131</td>
<td>bad</td>
<td>201</td>
<td>to look</td>
</tr>
<tr>
<td>62</td>
<td>leaf</td>
<td>132</td>
<td>wet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>root</td>
<td>133</td>
<td>dry</td>
<td>202</td>
<td>I</td>
</tr>
<tr>
<td>64</td>
<td>thorn</td>
<td>134</td>
<td>long</td>
<td>203</td>
<td>You (informal)</td>
</tr>
<tr>
<td>65</td>
<td>flower</td>
<td>135</td>
<td>short</td>
<td>204</td>
<td>you (formal)</td>
</tr>
<tr>
<td>66</td>
<td>fruit</td>
<td>136</td>
<td>hot</td>
<td>205</td>
<td>he</td>
</tr>
<tr>
<td>67</td>
<td>mango</td>
<td>137</td>
<td>cold</td>
<td>206</td>
<td>she</td>
</tr>
<tr>
<td>68</td>
<td>banana</td>
<td>138</td>
<td>right</td>
<td>207</td>
<td>we (inclusive)</td>
</tr>
<tr>
<td>69</td>
<td>wheat (husked)</td>
<td>139</td>
<td>left</td>
<td>208</td>
<td>we (exclusive)</td>
</tr>
<tr>
<td>70</td>
<td>barley</td>
<td>140</td>
<td>near</td>
<td>209</td>
<td>you (plural)</td>
</tr>
</tbody>
</table>

40
REFERENCES


MacQueen, James. 1967. Some Methods for classification and Analysis of


