The Nata Documentation Project: An Overview


This is a pre-print of an article accepted for publication in Jason Kandybowicz and Harold Torrence (eds.) Africa's Endangered Languages: Documentary and Theoretical Approaches (Oxford University Press) to appear in 2017. Please do not distribute.

This chapter introduces the research context of three papers in this volume (Anghelescu et al., Déchaine et al., Gambarage & Pulleyblank), all of which discuss an aspect of the grammar of Nata (Guthrie E45), an under-described Eastern Bantu language spoken in Tanzania. This overview also presents issues that situate the theme of these papers within the larger context of the documentation of endangered languages.

1. Goals of the project

The Nata documentation project started as a year-long Field Methods course at the Department of Linguistics, University of British Columbia (UBC) in 2012-2013 (September 2012 to April 2013). A collaborative research group — which includes faculty members, graduate and undergraduate students, as well as off-campus volunteers — called the Nata Working Group (NaWoG) was established in June of 2013. NaWoG has two goals. First, to produce and maintain a record of the Nata language via descriptive grammatical sketches. Second, to analyze language phenomena in Nata that challenge or shed light on current linguistic theories. Integration of data collection with theoretical analysis is the cornerstone of this project.

2. The Nata language: an introduction

Nata, called Ekinata by speakers, is an under-described Eastern Bantu language mainly spoken in Nata Mbiso, Nata Motukeri, and Makondose (Serengeti District, Mara Region, northwestern Tanzania), with some speakers in Mugeta and Kyandege (Bunda District), as indicated by the star in Figure 1. In Guthrie’s (1948) classification, Nata is in Zone E, Group 40, Index 5. The approximate number of speakers is 7,000 (Muzale and Rugemalira 2008). Table 1 gives the number of speakers for other languages in the E40 cluster. The number of speakers given in this chart represents speakers across geo-political borders; for instance, Gusii has around 1,000

---

1 The Nata Grammatical Sketch (GS) is planned as an online resource with descriptive “sketches” of phonetic, phonological, morpho-syntactic (nominal, verbal, deverbal), semantic, as well as information-structure features. In addition, the GS will also include information relative to demography and language context, as well as provide sample narratives.
speakers in Tanzania, but many more in Kenya. Values taken from SIL 2009 are marked by an asterisk; all other values are from Muzale and Rugemalira (2009), who only survey speakers in Tanzania. These numbers detail the number of speakers for dialects which other sources (SIL for example) report under one language name. We call attention to the fact that, in terms of numbers of speakers, Nata lies at the lower bound of the E40 cluster. (We return to this below when we discuss the language endangerment status of Nata.)

Figure 1. The geographical location of the Nata speech community

<table>
<thead>
<tr>
<th>Language</th>
<th>Speakers</th>
<th>Language</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gusii</td>
<td>*2,205,300</td>
<td>Sonjo</td>
<td>24,618</td>
</tr>
<tr>
<td>Kuria (Kurya)</td>
<td>*620,000</td>
<td>Ikoma</td>
<td>19,939</td>
</tr>
<tr>
<td>Logooli</td>
<td>*690,000</td>
<td>Kabwa (Kabhwa)</td>
<td>17,692</td>
</tr>
<tr>
<td>Tiriki</td>
<td>*210,000</td>
<td>Sweta</td>
<td>10,735</td>
</tr>
<tr>
<td>Idaxo</td>
<td>*171,000</td>
<td>Isenye</td>
<td>8,238</td>
</tr>
<tr>
<td>Isuxa</td>
<td>*171,000</td>
<td><strong>Nata</strong></td>
<td>7,050</td>
</tr>
<tr>
<td>Suba</td>
<td>*139,000</td>
<td>Hacha</td>
<td>7,008</td>
</tr>
<tr>
<td>Zanaki</td>
<td>97,429</td>
<td>Shashi</td>
<td>4,449</td>
</tr>
<tr>
<td>Ngurimi (Ngoreme)</td>
<td>52,360</td>
<td>Surwa</td>
<td>4,394</td>
</tr>
<tr>
<td>Ikizu</td>
<td>48,456</td>
<td>Singa</td>
<td>extinct</td>
</tr>
<tr>
<td>Simbiti</td>
<td>38,086</td>
<td>Ware</td>
<td>extinct</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------</td>
<td>---------</td>
</tr>
</tbody>
</table>

Table 1: Numbers of speakers for languages in the Bantu E40 zone

The phonemic inventory of Nata (Gambarage [Johannes] 2007) includes seven vowels, as in (1), and fifteen consonants, as in (2). Consonants in parentheses occur only in certain phonological environments and are not phonemic.

(1) Nata vowel inventory

\[
\begin{array}{cccc}
i & u & e & o \\
\varepsilon & \varepsilon & \circ & a \\
\end{array}
\]

(2) Nata consonant inventory

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Alveopalatal</th>
<th>Postalveolar</th>
<th>Palatal</th>
<th>Vela</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>(b)</td>
<td>t</td>
<td>(d)</td>
<td></td>
<td></td>
<td>k</td>
<td>(g)</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>n</td>
<td>η</td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>β</td>
<td>s</td>
<td></td>
<td>f</td>
<td></td>
<td>θ</td>
<td>h</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td>tʃ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approx.</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

3. Language endangerment status of Nata

For the nine factors identified by UNESCO, Nata is considered to be a severely or critically endangered language, with a cumulative ranking of 1.55/5; see Table 2.2 Relative to Factor 1, intergenerational language transmission, Nata is definitely endangered: it is mostly used by the parental generation, and children no longer learn the language as a first language. Intergenerational transmission is in the process of being broken, but the child-bearing generation can still use the language. Relative to Factor 2 (absolute number of speakers), Nata is definitely vulnerable: it is a small speech community with approximately 7,000 speakers. Relative to

---

2 For a review of the various scales used to assess language endangerment, see Tsunoda (2005). In addition to the UNESCO scale used here, another widely used scale is EGIDS (Extended Graded Intergenerational Disruption Scale): https://www.ethnologue.com/about/language-status (Lewis and Simons 2010).
Factor 3, proportion of speakers with total population, Nata is extremely vulnerable: as of 2013, Tanzania has an estimated population of approximately 49 million, so the proportion of Nata speakers is very small. To put this in context, the language with the largest proportion of speakers in Tanzania is Sukuma, which has 5 million, i.e., roughly 10% of the population. Relative to Factor 4, domain of language use, Nata has a highly limited domain of usage. Nata is indigenous to an area of Tanzania that is part of the Serengeti resort economics, and attracts workers from many different places. Consequently, the lingua franca is Swahili, and opportunities to use Nata are restricted to the private sphere; i.e., at home and with extended family members. Relative to Factor 5, response to new domains and media, Nata is inactive in that it not used for broadcast, in print media, or in the educational system. Relative to Factor 6, availability of materials for language education and literacy, no orthography is currently available to the community. (The Nata Working Group has developed a practical orthography, which needs to be tested and developed with community members.) Relative to Factor 7, language attitudes and policies, Nata is in a situation of forced assimilation. Swahili is the official language of Tanzania, English is the administrative language, and non-dominant languages are neither recognized nor protected. Relative to Factor 8, community members’ attitudes towards language, the Nata speech community has an extremely positive view of the language as a cultural resource. Relative to Factor 9, type and quality of documentation, the documentation of Nata is fragmentary: there are some grammatical sketches, word-lists and texts are available but inadequate, audio and video recordings of varying quality exist but without annotation. In this context, the documentation activities of this project are timely because they provide a solid base for the eventual re-stabilization of the Nata speech community.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Criteria</th>
<th>Status</th>
<th>Rank (/5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>language transmission</td>
<td>definitely endangered</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>number of speakers</td>
<td>small; definitely vulnerable</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>proportion of speakers</td>
<td>very small; extremely vulnerable</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>usage</td>
<td>highly limited domain</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>new domains &amp; media</td>
<td>inactive</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>literacy</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>language policies</td>
<td>forced assimilation</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>community attitudes</td>
<td>positive attitude</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>documentation</td>
<td>fragmentary</td>
<td>2</td>
</tr>
</tbody>
</table>
3. Methodology

There is very little linguistic description and analysis of Nata (see Factor 9 in Table 2), so the research group has adopted an exploratory approach to data collection. The methods we use reflect the convergence of three factors. First, we are fortunate to have a Nata speaker-linguist as a member of our research group in the person of Joash Gambarage. Second, the intellectual culture of our research unit predisposes us to focus on the interplay between theory and data. Third, we use elicitation techniques that allow in-depth exploration of language. We discuss each of these in turn.

3.1 Involvement of a speaker-linguist

All data are elicited from Joash Gambarage, a male consultant in his mid thirties, who is a native Nata speaker born and raised in Mugeta, Mara Region, Tanzania. Mr. Gambarage is currently a doctoral student in the UBC graduate linguistics program, is a balanced Nata-Swahili bilingual, and has received tertiary-level education in English. Because he is both a fluent speaker and a trained linguist, he is in a privileged position to guide us through the language in a way that maximizes the significance and impact of the research that we conduct. Concretely, the three papers in this section are an example of the kind of synergy that can arise when significant input is provided by a speaker-linguist. The three topics investigated — vowel harmony in the nominal domain, tone in the verbal and nominal domain, and deverbal nouns — all involve a complex interplay between different parts of the grammar. Such work requires a deep knowledge of the language, which is only possible if one is a native speaker. This native speaker knowledge — called I-language (Chomsky 1986) by generativists — is what a formal grammar is a model of, and is akin to the Saussurean notion of langue.

Our focus on I-language by no means excludes the relevance of E-language, akin to the Saussurean notion of parole. The latter focuses on data related to variation, as well as to the broader social and communicative use of language. In an ideal world, both types of research should be conducted in parallel. We recognize that language is defined by how speakers use it amongst themselves. As such, a single speaker can only offer so much data and only data of a specific kind; accordingly, there is information which we do not have access to related to register, style, gender, and age. For instance, because we have not worked with any very old or

Table 2: Nata Language endangerment Status: 14/45 (cumulative score); 1.55/5 (av. score)
very young speakers of Nata, we do not have data points that bear on how usage changes with age, either in real or apparent time. We likewise do not have data points relating to possible gender-based differences, or to regional variation.

Despite these potential drawbacks, it nevertheless remains the case that close and sustained collaboration with a single speaker-linguist yields enormous benefits. It results in systematic and internally consistent paradigmatic data, and allows in-depth analysis of specific morphological, syntactic, semantic, and pragmatic features.

3.2 The interplay between documentation and theory

Our research on Nata is informed by the interplay between documentation and theory. In principle, language documentation can inform theory in two ways: i) by confirming predictions made by a given formal model; ii) by challenging the predictions made by that model. Conversely, theory likewise guides data collection in two ways: i) it predicts the existence of certain data sets (positive data); ii) it predicts the non-existence of certain data sets (negative data). For related discussion of the interplay between data and theory, see Rice (2001, 2006). As we discuss in detail below (§5), for the three contributions to this volume, both documentation and theory converge on a robust partition between nouns and verbs in Nata.

3.3 Elicitation techniques

We use three primary elicitation techniques: English-to-Nata, Nata-to-English, and storyboard elicitation.

*English to Nata elicitation.* Because the research group members are not fluent in Nata, English-to-Nata is the most common form of elicitation. In this mode the elicitor prepares a list of target forms or contexts, such as morphological paradigms, syntactic frames, or discourse contexts. The target forms, frames or contexts are then presented to the consultant in English, who then provides a Nata translation. While this technique allows rapid progress regarding elicitation of forms, it suffers from a lack of ecological validity, as it runs the risk of giving undue weight to those aspects of Nata grammar that happen to converge with English-based forms and usage. It does however provide a good stepping stone to begin investigating the nuances of Nata.

*Nata to English elicitation.* In this second mode, the elicitor creates a phonological, morphological, syntactic, semantic or pragmatic frame and pairs it with an utterance of Nata in order to determine if the utterance is permissible in that frame. This form of elicitation provides a way to test hypotheses about both form and usage. It is a crucial step in providing a
comprehensive description of Nata phonology, morphology, syntax, semantics, and pragmatics based on Nata-internal evidence. Because the Nata language consultant is a speaker-linguist, he can assist elicitors in preparing target forms, which greatly increases both the breadth and depth of empirical coverage.

*Storyboard elicitation.* This technique, which was pioneered at UBC (Burton & Matthewson 2011), provides a way of collecting utterances in well-defined contexts. In this mode of elicitation, the elicitor creates a series of images arranged in linear order, usually with six to nine panels, in a comic-strip like fashion. The images give a visual representation of a particular situation, and can be used to target lexical items, syntactic constructions, semantic contrasts, or discourse contexts. There is no text in the comic; rather it is presented to the consultant, who then narrates the depicted situation in the target language. The research group has so far used this technique once, in the context of testing question-answer congruence and its sensitivity to person contrasts (Burton et al. 2013).

4. The Nata Online Linguistic Database (OLD)

The Nata Online Linguistic Database is a collaborative web application used to document Nata. It was built using the Online Linguistic Database, an open source software for creating multi-contributor databases for storing and curating linguistic data. For a description of OLD see Dunham et al. (2014); for a more detailed account of the software see Dunham (2014).

*Nata OLD* helps NaWoG members to share data and create an online corpus. This is especially useful in the context of an under-documented minority language like Nata where access to speakers is limited and where there are no written or digital resources. (To our knowledge, the only written resource is the SIL Bible translation in Ikoma, a language closely related to Nata.)

In addition to facilitating collaboration and data-sharing, *Nata OLD* helps to maintain a consistent data set, in particular with respect to orthographic transcriptions. Using the OLD’s orthography-based input validation feature, *Nata OLD* prevents contributors from entering orthographic transcriptions that cannot be constructed via concatenation of graphemes in the specified orthography.

---


The orthography specified in Nata OLD\(^5\) is given in (1’) and (2’) below, which correspond to (1) and (2), respectively.\(^6\) That is, <é> is the orthographic representation of phonetic /é/, <mb> represents [b], <ny> represents /ɲ/, etc. The orthographic inventory contains only ASCII characters,\(^7\) with the exception of the COMBINING DOT BELOW (U+0323), and the COMBINING ACUTE ACCENT (U+0301). This design decision makes it easy to write Nata without the need for special-purpose software keyboards with IPA-derived Unicode characters. Note finally that vowels with acute accent (to represent high tone) are included in the orthography in order for orthography-based input validation to work.

(1’) Orthographic vowel inventory specified in Nata OLD

\[
\begin{align*}
\text{i} & \quad \text{i} \\
\text{e} & \quad \text{é} \\
\text{e} & \quad \text{é} \\
\text{o} & \quad \text{ó} \\
\text{ọ} & \quad \text{ọ́} \\
\text{a} & \quad \text{á} \\
\end{align*}
\]

(2’) Orthographic consonant inventory specified in Nata OLD

<table>
<thead>
<tr>
<th></th>
<th>Bi-labial</th>
<th>Alveolar</th>
<th>Alveo-palatal</th>
<th>Post-alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>mb</td>
<td>t</td>
<td>nd</td>
<td></td>
<td>k</td>
<td>ng</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td>ny</td>
<td>ng’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>bh</td>
<td>s</td>
<td>sh</td>
<td></td>
<td>gh</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td>ch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approx.</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>y</td>
</tr>
</tbody>
</table>

Note that IPA transcription is used in Anghelescu et al. and Gambarage & Pulleyblank, while orthography is used in Déchaine et al. The decision to use an IPA or an orthographic

\(^5\) The Nata OLD orthography is based on previous descriptions of Nata phonology (Gambarage [Johannes] 2007) and was designed by NaWoG members in consultation with the developer of the OLD. The Nata orthography has so far been used as an instructional tool for UBC Fieldwork classes, and as a research tool by NaWoG. It remains to be seen if it will be adopted as a practical orthography by the Nata speech community.

\(^6\) In OLD, the Nata orthography is underlyingly represented as a simple string of comma-delimited character sequences: <mb, nd, ng, t, ch, h, k, m, n, ny, ng’, r, bh, s, sh, gh, w, y, i, u, é, o, ó, é, o, ó, a, ã>.

\(^7\) ASCII = American Standard Code for Information Interchange.
representation is determined by the analysts’ goals: the first two papers are focused on morpho-phonology, the third paper is focused on morpho-syntax. We also draw attention to the fact that orthographic mb and bh are allophones of the same phoneme /b/, which is realized as a prenalized stop when it follows another nasal stop [nb] (orthographic mb), and is elsewhere realized as a voiced bilabial fricative [β] (orthographic bh).

Within Nata OLD are also specified an inventory of phonemes and of morpheme delimiters. Together these inventories are used to restrict which character sequences can be entered in the morpheme break fields of Nata forms. This morpheme break phonemic inventory corresponds almost exactly to the phonemic inventory in (1, 2), except for special provisions related to language-specific morphotactics.

Input validation is also enabled for broad phonetic transcriptions in Nata OLD. In this case, all of the phonemes of (1, 2) are permitted, as are the allophones [b], [d], [g], and [r]. (The flap is the surface realization of intervocalic /r/.)

The input validation as configured in the Nata OLD for the orthographic transcription, morpheme break, and broad phonetic transcription fields was implemented to encourage the creation of a consistent data set that can be effectively searched and used by a number of researchers. However, we also did not want to discourage the contribution of data from fieldworkers operating under different analytical assumptions. Our strategy is to begin with validation based on a particular set of phonological assumptions and then remove restrictions based on feedback from Nata OLD users and Nata Working Group members. This approach is effective at balancing consistency with inclusivity, and we believe that the use of web-based collaborative tools like OLD can assist in the development of orthographic conventions.

5. The prevalence of the noun/verb partition in Nata

The papers in this volume (Angelescu et al., Déchaine et al., Gambarage & Pulleyblank) all treat the noun/verb partition. The Nata verb complex consists of a root in combination with proclitics and suffixes. The proclitics include Tense/Aspect, as well as pronominal subject and object marking (SM, OM). The suffix slots includes optional extensional suffixes and an obligatory final suffix in the form of a final vowel (-a, -i, -u -e, -o) or an aspectual suffix (-ire). Our present understanding of the syntax of the Nata verb complex appears in (3). The root raises via head-movement to attach to the extensional suffixes (Baker 1988; Zeller 2013); we label this extended verb stem ExtP. The rightmost suffixes (the final vowel and aspectual -ire) are external to ExtP, and in Nata have the properties of Voice. To the left lie the Tense/Aspect clitics, as well

---

8 The phonemic inventory is specified in the system as a comma-delimited list:
\[t,c,d,h,k,m,n,f,\beta,β,r,s,ʃ,y,\epsilon,j,i,u,\dot{u},e,\ddot{e},o,\ddot{o},ɛ,\ddot{ɛ},\ddot{a},\ddot{a}.\]
as the pronominal clitics; the latter are introduced into the Specifier position of TP (for Subject Marking) and VoiceP (for Object Marking).

(3) Syntactic constituency of the Nata verb complex

Phonologically, we adopt a version of Downing’s (1999) morpho-phonological structure, which distinguishes the Derivational stem (D-Stem) from the Inflectional stem (I-Stem), as in (4). The D-Stem groups together the root and extensional suffixes, and is equivalent to ExtP in the syntactic representation above. The I-stem groups the D-stem with the rightmost suffixes, i.e. the final vowel and aspectual -ire. The proclitics subdivide into two phonological classes according to whether they fall inside or outside of the MacroStem. A crucial property of both the morpho-syntactic and morpho-phonological analysis is the claim that the OM proclitic has a privileged relation with the VerbStem, either by virtue of being sister to the extended vP or by being contained in the MacroStem; As we shall see below, in Nata, this is consistent with the alignment of tone.

(4) Phonological constituency of the Nata verb complex

The representations in (3) and (4) raise the question, not resolved here, about the nature of the phonology-syntax mapping. Relevant to this issue is the behavior of suffix-controlled vowel
harmony, discussed below, which in Nata is triggered by the rightmost final vowel, and affects the vowel melody of the entire MacroStem: this is expected under the syntactic analysis (where the final vowel and OM/Class Prefix form a constituent), but not under the phonological analysis.

For the purposes of the present discussion we use the morphoogical templates templates in (5), which combine morpho-phonological labelling (Word, Macrostem, Stem) with syntactic constituency: the verb template is given in (5a), and the template for underived and derived nouns is given in (5b-c). For nouns, the two prefix slots are filled by (optional) pre-prefixes and obligatory noun-class marking. In addition, deverbal nouns predictably have the same set of suffixes that verbs do, namely extensional suffixes and a final suffixal vowel.

(5) Morphological Templates
   a. \[
   \text{[Word SM (TAM) [MStem (OM) [Stem Root -(Ext)] -fv/Asp ]] Verb}
   \]
   b. \[
   \text{[Word (ppf) [MStem CM[Stem Root ] ] ]} \text{ Noun}
   \]
   c. \[
   \text{[Word (ppf) [MStem CM[Stem Root -(Ext)] -fv ]] Deverbal Noun}
   \]

A novel feature of all three papers is the emphasis on analyzing the patterns of both verbal and nominal forms. The three domains investigated are tone melody, vowel harmony, and nominalization. In different ways, for each of these domains, the noun/verb partition plays a pivotal role.

5.1 Tone

The distribution of high tones follows from boundary effects, where tones align to well-defined morpho-phonological boundaries, in particular, the left and right edge of the MacroStem. Syllables are the tone bearing unit in Nata. Each syllable may bear a high tone or a low tone; high tone associated with long vowels will be realized as a level high tone unless they are in a word penultimate syllable. For nouns, tone is lexically specified, and defines three tone classes, namely: Macrostem-initial H (6a), Macrostem-initial LH (6b), and Macrostem-final H (6c). If high and low tones were freely distributed we would predict more than these three tone patterns.\(^9\)

(6) a. \[
\text{[MStem H … ]}
\]
   \[
\text{[o- [MStem rú- [ yanɔ ] ] ]}
\]
   \[
\text{ppf c9 intestine.fat}
\]
   ‘intestinal fat, Class 9’

\(^9\) Tonal minimal pairs, though rare, do exist in Nata; e.g. [[tʃa-[kɑ]]] ‘house, Class 10’, [[tʃa-[ká]]] ‘lion, Class 10’. The monosyllabic root makes this specific pair exceptional: -kɑ ‘house’ versus -ká ‘lion’.
For verbs, H tone is syntactically conditioned by tense, aspect and mood. The attested patterns are the same as for nouns, namely Macrostem-initial H (7a), Macrostem-initial LH (7b), and Macrostem-final H (7c). In Nata, negation is part of the irrealis mood paradigm; this is motivated by the fact that irrealis forms (including the affirmative past conditional and the negative near past), exhibit the same tonal pattern, namely Macrostem-final H.

\[
\begin{align*}
(7) \quad & \text{a.} \quad \begin{bmatrix} \text{MStem} & H & \ldots \end{bmatrix} \\
& \begin{bmatrix} \text{u-ka} & \text{yí} & [\text{som-a}] \end{bmatrix} \\
& \text{2sg-past OM7 read-fv} \\
& \text{‘You read’} \\
\end{align*}
\]

\[
\begin{align*}
\text{b.} \quad & \begin{bmatrix} \text{MStem} & L & H & \ldots \end{bmatrix} \\
& \begin{bmatrix} \text{n-oo} & \text{yí} & [\text{som-ɛ}] \end{bmatrix} \\
& \text{COMP-2sg OM7 read-subjunctive} \\
& \text{‘Please read it’} \\
\end{align*}
\]

\[
\begin{align*}
\text{c.} \quad & \begin{bmatrix} \text{MStem} \ldots & H \end{bmatrix} \\
& \begin{bmatrix} \text{ŋ-aaŋga} & [[\text{som-irɛ}] \end{bmatrix} \\
& \text{COMP.1sg-past-cond read-PFV} \\
& \text{‘I would read’} \\
\end{align*}
\]

5.2 Vowel harmony

Sequences of mid vowels in Nata exhibit tongue root harmony (Gambarage 2013), and are conditioned by three contexts: (i) root-internal; (ii) root-controlled; (iii) suffix-controlled.

*Root-internal harmony.* All roots are harmonic, for example -heende ‘lentil Class 11’ and -yɛrɛ thing, Class 7’. While all mid vowel sequences agree in their tongue root value (8a), high and low vowels are neutral, occurring with both advanced and retracted mid vowels (8b-c).
Root-controlled harmony. The tongue root value of a prefix is conditioned by the vowel quality of the initial root vowel. (This is completely general, and holds of both verbs and nouns, though Gambarage & Pulleyblank 2014 focus on nouns.) If the first syllable of the root contains an ATR vowel, then the vowel of the noun-class prefix is ATR, as seen in o-ro-[héende] ‘lentil, Class 11’. If the first syllable of an underived root contains an RTR vowel, the vowel of the noun-class prefix is a high ATR vowel, as seen in o-rú-[saro] ‘bead, Class 11’. Had the mid vowel prefix ro- appeared before the root saro, the [o…a] sequence would constitute a harmony violation: *o-ro-[saro]. As a result, the prefix appears as a high vowel, thereby achieving harmony by removing a potentially disharmonic vowel in the prefix. Note that a low vowel is neutral within a stem (hence [saro] is wellformed) but requires harmony in prefixes (hence rú-[saro], not *ró-[saro]).

Suffix-controlled harmony. In two morphological conditions, RTR mid vowels follow root vowels that are otherwise ATR. In such cases, root vowels retract, resulting in a sequence of RTR vowels. Of the two suffixes that trigger harmony, one is verbal and one is nominal. In verbal contexts, the subjunctive final vowel -e triggers harmony, as in (9a). In nominal contexts, the deverbal instrumental (“inanimate agent”) triggers harmony, as in (9b). Note that when harmony is triggered by an RTR suffix, the retraction persists into the prefixes. That is, in this class of cases, prefixes are mid retracted: they do not raise to high. Only the subjunctive -e and nominalizer -a suffixes condition this kind of retraction; extensional -VC suffixes undergo harmony, (9c-d) and the final vowel -a has no effect on a preceding mid vowel (9e).

(8) a. o- ro- [rt héende ]
    ppf- c11- lentil
    ‘lentil, Class 11’
    e- ki- [rt ýe̞ro ]
    ppf- c7- thing
    ‘thing, Class 7’

b. o- mo- [rt súβe ]
    ppf- c1- man
    ‘man, Class 1’
a- ma- [rt kééŋeeti ]
    ppf- c6- ant
    ‘big ants, Class 6’

c. e- me- [rt kéra ]
    ppf- c4- tail
    ‘tails, Class 4’
o- ýu- [rt taaró ]
    ppf- c20- river
    ‘river (augmentative), Class 20’

(9) a. [tɔ-ʊʊr-ɛ ]
    /toor-ɛ /
    1pl-put-subj
b. [ɔ-mo-tʃʊr-ɔ ]
    /o-mo-toor-ɔ /
    ppf-C3-put-instrument
‘let us put (subjunctive)’  ‘load, Class 3, i.e. thing that is put’

c. ko-[réh-er-a]  d. ku-[yéγ-er-a]
c15-pay-appl-fv  c15-carry-appl-fv
‘to pay for’  ‘to carry for’

e. ko-[mér-a]
c15-swallow-fv
‘to swallow’

5.3 Deverbal nouns

Deverbal nouns have two tone classes. Macrostem-initial H is found with animate and inanimate agent nouns (inanimate agents are construed as instruments), as in (10). Macrostem-initial LH includes animate and inanimate patient nouns, as well as event nouns and verbo-nominal infinitives, as in (11).

(10)  \[ \begin{array}{c}
\text{MStem} \\
[H \ldots]
\end{array} \]

a. [o \text{MStem} \text{mó} \text{[sóm-i]}]
ppf c1 read-agent
‘person who reads, Class 1’

b. [e- \text{MStem} \text{yí} \text{[sóm-er-ɔ]}]
ppf c7 read-appl-instr
‘thing that you read with, Class 7’

(11)  \[ \begin{array}{c}
\text{MStem} \\
[L \quad H \ldots]
\end{array} \]

a. [o \text{MStem} \text{mu} \text{[séeγ-u]}]
ppf c1 like-patient
‘person who is liked, Class 1’

b. [o \text{MStem} \text{yí} \text{[sóm-u]}]
ppf c7 read-patient
‘thing that is read, Class 7’

c. [o \text{MStem} \text{yu} \text{[sóm-a]}]
ppf c15 read-fv
‘event of reading, Class 15’
While event nouns (11c) and infinitives (11d) pattern together tonally (both have Macrostem-initial LH) and both are formed with the prefix from noun class 15, they differ relative to whether or not they take the pre-prefix: the deverbal event noun appears with a pre-prefix, while the infinitive lacks the pre-prefix.

The fact that deverbal nouns divide into two tone classes sheds light on the three-way tone partition found with underived nouns and verb stems. Consider Table 3. First is the fact that while tone melody for verbs and deverbal nouns is morpho-syntactically conditioned, for noun roots it is lexically specified; we call this “syntactic tone” versus “lexical tone”. Second is the fact that, for syntactically conditioned tone, Macrostem-initial H is the elsewhere case, and arises if a root does not bear an underlying tone. (This is always the case for verbs in Nata, as they are toneless by default, i.e., they are not lexically specified for tone.) Third is the fact that, with verbs and deverbal nouns, the LH tone melody is morpho-syntactically conditioned (Anghelăescu et al., this volume). This means LH verbal and de-verbal forms constitute an inflectional class. As for Macrostem-final H, this arises only with right-edge suffixes that are specified for H-tone: for reasons that remain unclear at present, this is found only in verbal contexts. Once again, this points to a noun/verb partition: while some rightmost suffixes in the verb stem are specified for H, as far as we know this does not occur with nominal final vowels.

<table>
<thead>
<tr>
<th>syntactic tone</th>
<th>deverbal noun</th>
<th>verb stem</th>
<th>lexical tone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓ anim. agent noun inanim. agent noun</td>
<td>✓ narrative past habitual progressive</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓ patient noun event noun</td>
<td>✓ infinitive subjunctive hortatory</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓ ✓</td>
<td>✓ (negative) past conditional negative habitual negative progressive</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 3: deployment of tone in Nata

Semantically, deverbal nouns divide into two classes according to whether they are event-denoting or entity-denoting. Event-denoting deverbal nouns include manner, event, and
stative nouns, (12). Entity-denoting deverbal Ns include agent, patient, and instrument nouns, (13).

(12) a. [rii-hutát-a] c5-turn.around-fv ‘manner of turning around, Class 5’

b. [o-γu-sóm-a] ppf-c15-read-fv ‘event of reading, Class 15’

c. [o-βú-sóm-i] ppf-c14-read-fv ‘state of being educated, Class 14’

(13) a. [o-mú-sóm-i] ppf-c1-read-agent ‘person who reads, Class 1’

b. [o-mw-iβ-u] ppf-c1-steal-patient ‘person who is stolen, Class 1’

c. [e-γí-sóm-ɛr-ɔ] ppf-c1-read-appl-instr ‘thing that you read with, Class 7’

A notable feature of Nata is the fact that the passive and causative suffixes have “drifted” away from the extensional suffix class, and are partially integrated into the final suffix slot; this is what motivates treating the final suffix position as a Voice head in (3) above. The re-positioning of these erstwhile extensional suffixes into the final suffix position can be seen in (13a) and (13b), where the -i causative and the -u passive occur in the final vowel slot; with deverbal nouns, these finals vowels derive agent and patient nouns respectively. (See Déchaine et al., this volume, for further discussion.)

References


