Towards a minimally useful dictionary of Nuu

Bonny Sands (Northern Arizona University)
Alena Witzlack-Makarevich (University of Kiel)
Earlier distribution of the Tuu languages
(Güldemann 2014)

Tuu

!Ui

- Nǁl̂ang
- !Xam
- Vaal-Orange
- Outliers

Taa-Lower Nossob

• Nǁl̂ang (aka Nluu after the better documented variety)
• the last living member of the !Ui branch of the Tuu language family, aka ‘Southern Khoisan’ (Greenberg 1963)

The speakers

- Among linguists, Nluu was thought to be extinct for several decades (Traill 1999)
- In the late 1990s, a Nluu speaker made a radio announcement and asked other speakers to make themselves known
- Same year N. Crawhall started to work on Nǁng: “an odyssey of rediscovery” resulted in locating 28 fluent and partial speakers (Crawhall 2004:13)
- They lived hundreds of kilometers apart, mostly in townships near Upington and in small desert settlements in the Kalahari
- The elders were brought together by linguists and encouraged to speak Nǁng
Since then, most of these speakers have passed away.

As of 2016, only four known remaining speakers, all in their eighties.

The speakers are not in regular contact with one another, mainly due to restrictions in mobility as a result of old age and poor infrastructure.

All speakers use Afrikaans as their language of daily communication.
The Nǀuu speakers of the ᶬKhomani community

- The Nǀuu speakers identify themselves as belonging to the ᶬKhomani (San) community
- One the most marginalized and disadvantaged groups of South Africa, both economically and socially
- The community comprises from 400 to 2000 people, the exact number is unavailable
- Most ᶬKhomani speak Afrikaans as their mother tongue
- The symbolic use of Nǀuu is spreading among the members of the community, but this does not necessarily equate to increased competence in the language
- Some members of the community expressed interest in a dictionary with pronunciation
The documentation efforts

• An NSF project ‘Collaborative Research: Descriptive and Theoretical Studies of Nluu’ (2003–2006) at Cornell University (Amanda Miller and Chris Collins) and Northern Arizona University (Bonny Sands)

• An ELDP project A text documentation of Nluu (2006–2011), Tom Güldemann, Martina Ernszt, Sven Siegmund, and Alena Witzlack-Makarevich
→ a corpus of 100,000 words

• a number of further researchers beyond these projects
TheǂKhomani San language school

- Community language teaching efforts by the remaining speakers for about 10 years, first in two locations, now only in Upington
- The Upington school:
  - for 5–10 years old children
  - held by one of the speakers and first her granddaughter, now replaced by the community's secretary/activist
  - irregular classes (subject to funding and health issues)
  - all of the research projects produced teaching material and held several orthography workshops for schoolchildren and interested community members (the speakers themselves are illiterate)
Collected lexical material


Collected lexical material: Sands et al. 2006

- Sands et al. (2006) worked with multiple speakers
- an Excel spreadsheet with the following fields:
  - Orthography
  - Cross-ref to recordings (often of several speakers)
  - Dictionary Recording (a recording selected as representative)
  - Semantic code
  - IPA (detailed transcription with indication of dialectal variation where appropriate)
  - English, Afrikaans, KhoekhoeGowab
  - Cross-Ref to related recordings, e.g. of carrier phrases

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Cross-Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1e</td>
<td>P04_139 JK, P04_139 KE, S12_383 AO, S12_382 JK, S12_382 KE, P07_243 AO, P07_243 KE</td>
<td>nu_dict_P07_253 KE</td>
<td>N01.32.1</td>
<td>je</td>
<td>blue wildebeest, gnu (can also be used to mean 'buffalo')</td>
<td>blouwildebees</td>
<td>gaos</td>
<td>pl. [leeke, [k][k][e]], S12_383 AO, S12_383 JK, S12_383 KE; [lee he dj][ue, [ee ho dj][ue (AO)] 'gnu that walks alone' S12_384 AO, S12_384 JK, S12_384 KE</td>
</tr>
</tbody>
</table>
Collected lexical material: Güldemann et al. (2011)

- Güldemann et al. (2011) is a Toolbox dictionary of some 2000 entries
- primarily corpus-based, with some additional elicitations
- same set of fields in the identical order
  - \lx lexical item
  - \a alternative form (e.g. dialectal variation)
  - \ps part of speech
  - \ge, \ga, \gk English/Afrikaans/Khoekhoe translation
  - \pl plural form for nouns (irregular)
  - \lxid lexicon id
  - \so etymological source (Nluu, Afrikaans, English, Khoekhoe)
  - \sp speakers who use/know this lexical item
  - \audi cross-reference to a recording
  - \cfsa cross reference to the entry in Sands et al. (2006)
Towards the Dictionaria format

• The adjustment of the Toolbox dictionary (Güldemann et al. 2011) in accordance with the suggested Dictionaria uniform data model should be straightforward

• Aim for a “minimally useful dictionary” (RF)

• To-do’s:
  ✓ merge Güldemann et al. (2011) and Sands et al. (2006)
  • prepare the definitions for grammatical markers according to the guidelines to be provided
  • add the \sf field and add sound file names (available in Sands et al. 2006)
Towards the Dictionaria format

• To-do’s:
  • add definitions, instead or in addition to translations? Most entries are rather minimalistic and adding extensive definitions seems redundant for the vast majority of them
  • e.g.

        \lx !qui
        \ps n
        \ge ash
        \ga as
        \gk tsaob
        \lxid 383
        ...

  • adding “the powdery residue left after the burning of a substance (e.g. wood)” seems superfluous
“a large dark antelope with a long head, a beard and mane, and a sloping back, belongs to the family Bovidae…”
Towards the Dictionaria format

- To-do’s:
- select examples sentences from the corpus (100,000 words, annotated, Toolbox)
Towards the Dictionaria format

• A wish list
  - there is a conventional sorting procedure for click sounds (depending on click accompaniments)

  “The sorting algorithm that is used is called DUCET, which is the default for Unicode characters.”
  (Dictionaria Submission Guidelines, v. version of 03-08-2016)

  - for non-linguists a way to enter special symbols (e.g. clicks) in the search fields would be very helpful
Towards the Dictionaria format

• A wish list
  - include several orthographies?
    at least four orthographies proposed for Nluu by different teams, not substantially different, but quite a few minor disagreements e.g. [sǐǐ̯sin] sîisín (Sands et al.) vs. siinsinn (Güldemannnn et al.)
  - the so-called community orthography (Shah & Brenzinger 2016) developed in collaboration with one of the speakers and two members of theǂKhomani community, who can read some Nluu (thanks to Namaseb’s workshop), not necessarily linguistically sound (not always an adequate representation of the phoneme inventory)
Towards the Dictionaria format

• A wish list
  - can we easily recycle the examples from ValPal (we have around 300)?
  - is it possible to include sound files also for the examples? This might be of help to the few community members interested in learning Nluu and can be used in the language school
Towards the Dictionaria format

• A wish list
  
  - possibility of an off-line/download version of the dictionary
  
  - most attractive: a mobile app with offline capabilities
    (Josh’s *simple search* functionality would suffice)

  many community members possess smart phones but might be reluctant to constantly pay for the online access
Towards the Dictionaria format

• A wish list
  - link to the Tsammalex database on plants and animals
  - already contains 38 entries from Nluu
  - do we link to a lexical item or to a species/taxon?

**Tsammalex**

**lee**

- **Meaning:** blue wildebeest
- **Language:** Nllang/Nluu
- **Species:** Connochaetes taurinus
- **IPA:** lée
- **References:** Sands et al. 2006
A side-project: retro-digitized dictionaries

- A side-project
  - retro-digitalization of the Nluu part of Dorothea F. Bleek (1956) *A Bushman Dictionary* (700 pp. of lexical material on 28 languages and varieties spoken by the San groups)

- Once complete, can a edition version of a retro-digitized dictionary be published with Dictionaria? Can it be to entries form the modern dictionary?