Embedding dictionary development in the documentary workflow

Blaine Billings, University of Hawaiʻi at Mānoa
NINJAL Language Revitalization Synergy Symposium (LRSS) 2025   Tachikawa, Tokyo, Japan, 13–16 June 2025

1. Introduction

Dictionaries frequently seen as <u>first step</u> for language maintenance and revitalization efforts in endangered language communities (Nyangone Assam 2006; Grimm 2022)

Lexicographic work often considered *isolated* from other documentary linguistic tasks or delayed to *future projects*

However, **embedding dictionary efforts** in the documentary workflow is important for producing wholistic and effective community-centered language output

Nasal: Austronesian language spoken by ~3,000 people in four villages of coastal Bengkulu province, Sumatra, Indonesia

2. Documentation

Typical documentary workflow:
- **Goal**: Create a lasting record of language use
- Elicitation for understanding of grammatical structure
- Analysis of naturalistic conversation for broader coverage of everyday speech
- Archiving of data for future use, development, and research
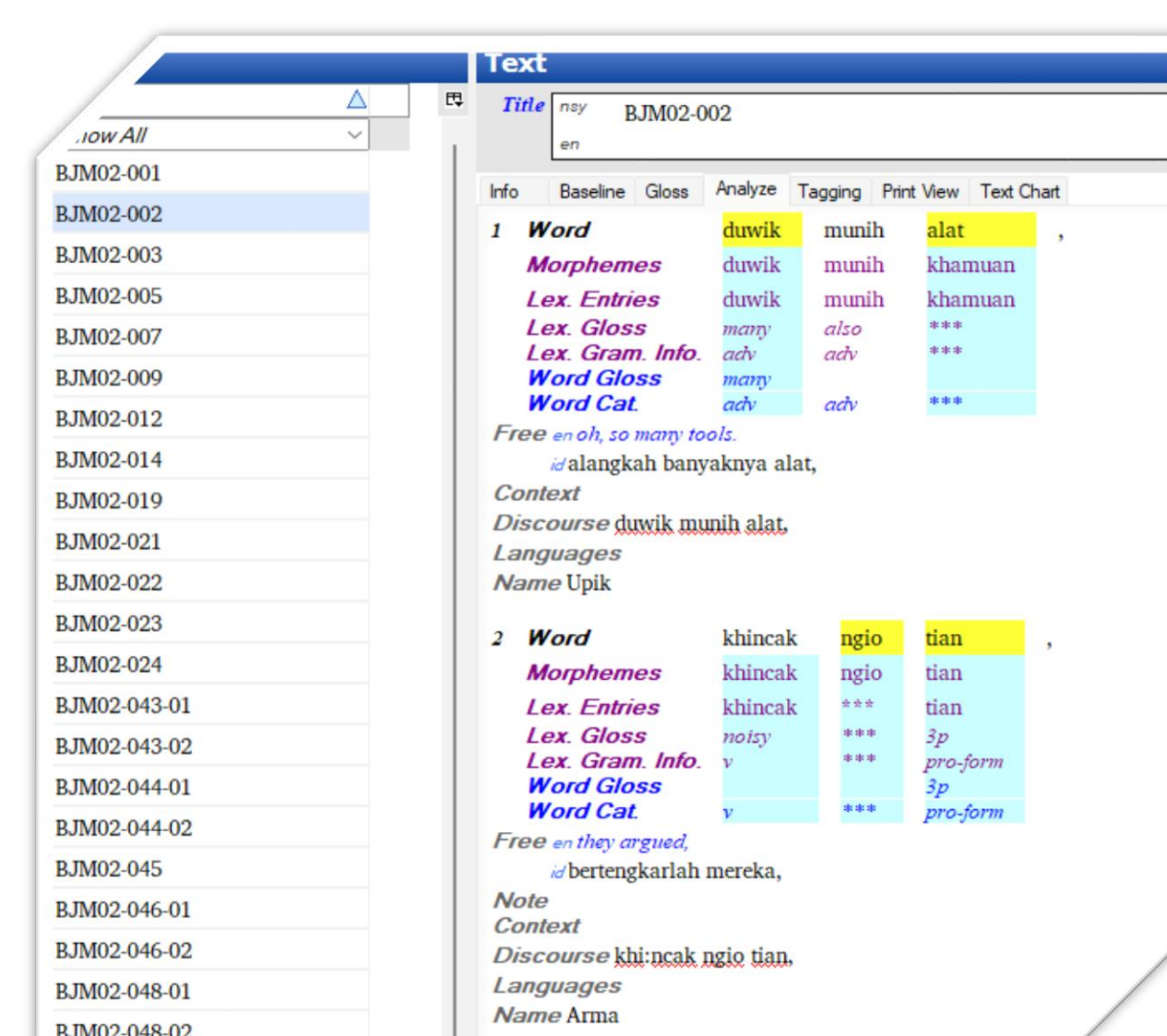
Nasal documentation:
- Ongoing project since 2017 (McDonnell 2017)
  - Little pre-existing documentation
- Collaborative effort with members of Nasal community, researchers from UH Mānoa, Atma Jaya Catholic University
- >360 hours of recorded narrative, <u>conversation</u>, elicitation
- **Planned output**: Grammar, dictionary, Nasal-language books

3. Challenges in lexicographic work

Dictionary development introduces a number of challenges:
1. Traditionally, *highly corpus-dependent* (Čermák 2003), but endangered languages often <u>lack written corpus</u>
2. Elicitation alone <u>inefficient</u> for breadth and depth of coverage
3. Dictionary efforts tend to <u>underrepresent nonstandard speech</u>
4. Grammatical and lexical analysis are <u>fundamentally interrelated</u>
5. Dictionaries, like documentary work, take <u>years of work</u>
6. <u>Funding scarce</u> for independent dictionary development projects
7. Dictionaries considered <u>inadequate for dissertation</u> projects
8. Often viewed as <u>all-or-nothing</u> for publication

4. Embedding dictionary in documentation

Many lexicographic challenges can be addressed through combining the typical documentary and lexicographic workflow:
1. **Corpus creation**: Transcriptions of narratives and conversations can be repurposed to form the basis for a language corpus
2. **Naturalistic speech**: Everyday conversation introduces words and senses that might be easily overlooked by elicitation as well as both colloquial and formal language use
3. **Grammatical analysis**: Morphological and syntactic analysis typically carried out aid in identification of collocations and inform description of word-classes
4. **Concurrence**: Integrating dictionary efforts as part of general documentary work leads to simultaneous progression of multiple aspects of language maintenance efforts

5. Extension of materials

Lexicographic materials can be extended to other resources
- **Grammar instruction**: Dictionary preface and example sentences provide foundational *grammatical* understanding
- **Mini-dictionaries**: Semantic domains can be separated out into *topical dictionaries* for education (e.g., "animal dictionaries" or "child alphabet books"; Mosel 2011)
- **Graded readers**: Integrated conversational analysis allows determination of *reading level* for linked and associated texts

6. Dictionaries for revitalization

Visibility:
- Dictionaries are often viewed as the most visible resource that demonstrate a language's inherent value

Involvement:
- Topic-oriented dictionary discussion sessions can inspire community-internal efforts to document specific cultural domains (in Nasal: wedding ceremonies, planting cycle, magic, net-making ⇒)

Education:
- Introduction of lexicographic resources in a classroom further engages language use in a formal domain
- Reference resources provide a way heritage-learners can engage with their language in everyday tasks

Future efforts:
- Archival practice common in documentary linguistics — but not yet ubiquitous for lexicographic documentation (Paterson 2023) — provides a solid foundation for future lexicographic work
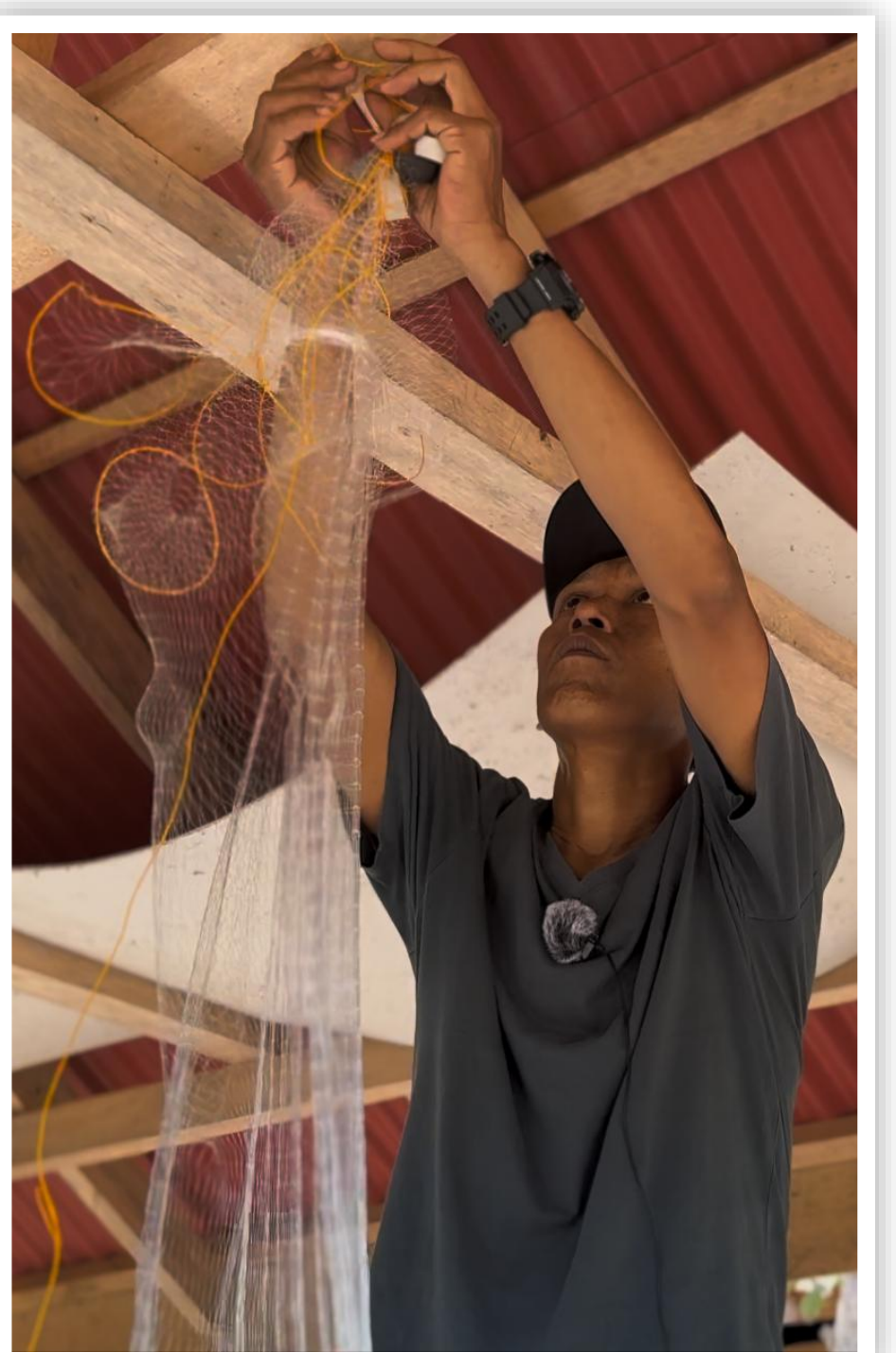
References

Čermák, František. 2003. Source materials for dictionaries. In Piet van Sterkenburg (ed.), *A practical guide to lexicography*, 18–25. Amsterdam: John Benjamins Publishing Company.

Grimm, Nadine. 2022. Documentary approaches to lexicography. In Moses Effiong Ekpenyong & Imelda Icheji Udoh (eds.), *Current issues in descriptive linguistics and digital humanities: A festschrift in honor of professor Eno-Abasi Essien Urua*, 551–568. Singapore: Springer.

Nyangone Assam, Blanche. 2006. *Dictionaries as teaching instruments for mother-tongue education: The case of Fang in Gabon*. Stellenbosch, South Africa: University of Stellenbosch (Doctoral dissertation.)

McDonnell, Bradley. 2017. Prospects of a community-based language project in the Nasal speech community. In Nicholas Ostler, Vera Ferreira, & Chris Moseley (eds.), *Communities in control: Learning tools and strategies for multilingual endangered language communities*, 64–69. London: Foundation for Endangered Languages.

Mosel, Ulrike. 2011. Lexicography in endangered language communities. In Austin, Peter K. & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 337–353. Cambridge: Cambridge University Press

Paterson, Hugh, III. 2023. Diversity and identity: Categories for OAI data-providers in the Open Language Archives Network. *NASKO* 9, 18–31.

Indonesia map: https://commons.wikimedia.org/wiki/File:Indonesia_provinces_blank.png