

Documentary lexicography

The need for a field

Blaine Billings (blainetb@hawaii.edu)

University of Hawai'i at Mānoa

International Conference on Language Documentation and Conservation (ICLDC9)

7 March 2025



Summary

1. The current state of lexicographic practice in documentary linguistics is insufficient

Summary

1. The current state of lexicographic practice in documentary linguistics is insufficient
2. Documentary linguistics presents unique challenges for lexicographic work

Summary

1. The current state of lexicographic practice in documentary linguistics is insufficient
2. Documentary linguistics presents unique challenges for lexicographic work
3. Changes in our practice need to be made in order to reflect the goals of our field

Why care about documenting the lexicon?

Lexicography in documentary linguistics

Two Halves: Grammar & lexicon

Lexicography in documentary linguistics

Two Halves: Grammar & lexicon

Boasian trilogy: Grammar, dictionary, text collection

Lexicography in documentary linguistics

Two Halves: Grammar & lexicon

Boasian trilogy: Grammar, dictionary, text collection

Lexicon forms vital component of documentary record

Lexicography in documentary linguistics

Two Halves: Grammar & lexicon

Boasian trilogy: Grammar, dictionary, text collection

Lexicon forms vital component of documentary record

Dictionaries often the first output desired by language communities

Lexicography in documentary linguistics

Two Halves: Grammar & lexicon

Boasian trilogy: Grammar, dictionary, text collection

Lexicon forms vital component of documentary record

Dictionaries often the first output desired by language communities

How are these facts reflected in practice and training?

State of lexicographic practice

Grammar-dictionary ratio 10:3 output (Himmelmann 1998: 191)

State of lexicographic practice

Grammar-dictionary ratio 10:3 output (Himmelman 1998: 191)

Typical lexicographic output

“about 3,000 entries; by lexicographical standards, this is too short to be considered a full dictionary” — Chelliah & de Reuse (2011: 228)

“The chances are very slender that this generation or the next will produce more great monuments for any of these regional[=Indonesia] languages.” — Wolff (1991: 2567)

- There are notable exceptions (e.g., parts of North America)

State of lexicographic practice

Grammar-dictionary ratio 10:3 output (Himmelman 1998: 191)

Typical lexicographic output

“about 3,000 entries; by lexicographical standards, this is too short to be considered a full dictionary” — Chelliah & de Reuse (2011: 228)

“The chances are very slender that this generation or the next will produce more great monuments for any of these regional[=Indonesia] languages.” — Wolff (1991: 2567)

- There are notable exceptions (e.g., parts of North America)

< 15% lexical materials archived (Paterson III 2015)

State of lexicographic training

“It is regrettable that so few universities in North America offer training in lexicography as part of their graduate programs in linguistics, especially given the central role that dictionaries play in both the documentation and revitalization of Indigenous languages.”

— Lachler & Pankratz (2017: 112)

State of lexicographic training

“It is regrettable that so few universities in North America offer training in lexicography as part of their graduate programs in linguistics, especially given the central role that dictionaries play in both the documentation and revitalization of Indigenous languages.”

— Lachler & Pankratz (2017: 112)

Methodological description for lexicographic projects from start to finish in documentary linguistics is severely lacking (Nabirye 2011: 179–181; P. M. Anderson 2020: 9)

State of lexicographic training

“It is regrettable that so few universities in North America offer training in lexicography as part of their graduate programs in linguistics, especially given the central role that dictionaries play in both the documentation and revitalization of Indigenous languages.”

— Lachler & Pankratz (2017: 112)

Methodological description for lexicographic projects from start to finish in documentary linguistics is severely lacking (Nabirye 2011: 179–181; P. M. Anderson 2020: 9)

Handbooks addressing lexicography for documentary linguists are few and far between

Lexicography in documentary linguistics???

Two halves: Lexicographic work underrepresented in academic output

Lexicography in documentary linguistics???

Two halves: Lexicographic work underrepresented in academic output

Boasian trilogy: And yet training rarely addresses lexicography

Lexicography in documentary linguistics???

Two halves: Lexicographic work underrepresented in academic output

Boasian trilogy: And yet training rarely addresses lexicography

Lexicographic work (in some form) does not meet standards (and often comes much later)

Lexicography in documentary linguistics???

Two halves: Lexicographic work underrepresented in academic output

Boasian trilogy: And yet training rarely addresses lexicography

Lexicographic work (in some form) does not meet standards (and often comes much later)

Little is discussed about the complexities of the process

Lexicography in documentary linguistics???

Two halves: Lexicographic work underrepresented in academic output

Boasian trilogy: And yet training rarely addresses lexicography

Lexicographic work (in some form) does not meet standards (and often comes much later)

Little is discussed about the complexities of the process

Our conception for documentary lexicographic output does not match the reality

Why not work on dictionaries?

We don't have official training!

Nor do many lexicographers, and we need to start somehow.

Why not work on dictionaries?

We don't have official training!

Nor do many lexicographers, and we need to start somehow.

Dictionaries are time consuming!

So are grammatical analysis.

Why not work on dictionaries?

We don't have official training!

Nor do many lexicographers, and we need to start somehow.

Dictionaries are time consuming!

So are grammatical analysis.

You can publish brief grammatical analyses/sketches!

As well as mini-dictionaries.

Why not work on dictionaries?

We don't have official training!

Nor do many lexicographers, and we need to start somehow.

Dictionaries are time consuming!

So are grammatical analysis.

You can publish brief grammatical analyses/sketches!

As well as mini-dictionaries.

There is little academic (linguistic) value for publishing lexicographic work!

There is immense value for the community.

How might we reenvision lexicography?

Reenvisioning lexicography... (goals)

Similar goals?

“an accurate and adequate record of a language for posterity” — Rhodes & Campbell (2018: 107)

“The goal of lexicography itself is the creation of dictionaries” — Rundell (2012: 63)

Reenvisioning lexicography... (goals)

Similar goals?

“an accurate and adequate record of a language for posterity” — Rhodes & Campbell (2018: 107)

“The goal of lexicography itself is the creation of dictionaries” — Rundell (2012: 63)

Lexicography:

- Which, then, is the goal for the documentary linguist?

“to present a full account of the words of a language, in all their meanings and patterns of use”

— Kilgariff (2007)

- Lexicography is about lexicon **not** dictionary

Reenvisioning lexicography... (outputs)

Himmelmann (2012):

- Documentary work inherently concerned with raw and primary data

Reenvisioning lexicography... (outputs)

Himmelmann (2012):

- Documentary work inherently concerned with raw and primary data

Documentary output:

- Raw data (recordings)
- Primary data (transcriptions, translations)
- Notes on meaning
- Lexical database

Reenvisioning lexicography... (outputs)

Himmelmann (2012):

- Documentary work inherently concerned with raw and primary data

Documentary output:

- Raw data (recordings)
- Primary data (transcriptions, translations)
- Notes on meaning
- Lexical database

Lexicographic output:

- Dictionary/ies

“this does not mean that as a rule documentation projects should not strive to produce some published product.” — Himmelmann (2012: 205)

Reenvisioning lexicography... (methods)

Corpus:

- Typical part of documentation project
- NOT on the scale of modern lexicographic work
- How to leverage the data available?

Reenvisioning lexicography... (methods)

Corpus:

- Typical part of documentation project
- NOT on the scale of modern lexicographic work
- How to leverage the data available?

Data management:

- Ethical concerns (e.g., data ownership)
- Archiving and reusability
- Collaboration

Reenvisioning lexicography... (methods)

Documentary methods:

- Rapid word collection (Boerger & Stutzman 2018)
- Semantic domain elicitation (Moe 2003)
- Informal “word searches”

Reenvisioning lexicography... (methods)

Documentary methods:

- Rapid word collection (Boerger & Stutzman 2018)
- Semantic domain elicitation (Moe 2003)
- Informal “word searches”

Elsewhere in lexicography:

- Database development
- User-based dictionary design
- Frameworks for lexical semantics

Reenvisioning lexicography... (values)

Academic value

“Higher level analysis, rather than documentation itself, has been the coin of the realm: in nearly all graduate departments of linguistics it has been grammars—or better still, grammars of grammars—that have been suitable as doctoral dissertations, whereas dictionaries, or, heaven forbid, text collections—the low end of the hierarchy—have not. The usual justification for this [is] that text collection and even dictionary making are only clerical activities”

— Woodbury (2003)

Reenvisioning lexicography... (values)

Documentary value

“It is not impossible for a linguist to compile a dictionary on the basis of a grammar, text collection, and unpublished lexical field notes, but it is impossible to compile a comprehensive grammar on the basis of a dictionary and text collection”

— Chelliah & de Reuse (2011: 227)

So again... why?

Lexicographic goals in the documentary context are unique

So again... why?

Lexicographic goals in the documentary context are unique

Lexicographic outputs need to better address documentary concerns

So again... why?

Lexicographic goals in the documentary context are unique

Lexicographic outputs need to better address documentary concerns

Documentary context introduces methodological challenges

So again... why?

Lexicographic goals in the documentary context are unique

Lexicographic outputs need to better address documentary concerns

Documentary context introduces methodological challenges

Value accorded to lexicographic work needs to be readdressed

An idea... so what do we do?

In current projects

Start working on a dictionary!

Archive lexical materials

Integrate targeted lexical elicitation

Develop mini-dictionaries

In future work

Integrate lexicography into your curriculum (general, field works, ...)

Encourage students to work on lexicographic projects

Support the academic value of lexicographic work

Conclusion: What this talk *IS* about

- ✗ Stop working on grammatical description
 - ✓ **Start working on more lexicographic work**
- ✗ Only truly final outputs should be published
 - ✓ **Incrementally output/publish/archive your work**
- ✗ Leave lexicographic work to communities
 - ✓ **Engage with communities on their desired outcomes**
- ✗ We have insufficient knowledge to carry out lexicographic work
 - ✓ **Teach lexicographic methods in your courses**
 - Develop literature on lexicographic methods for documentary linguists**

Documentary lexicography

“It would be useful to better bridge general lexicography and dictionary making for smallscale communities in the future. It would be helpful for documentary/descriptive dictionary makers to take note of the insights of general lexicographers, who have a wealth of practical experience. In fact, many issues that documentary dictionary makers struggle with also preoccupy general lexicographers...”

— Grimm (2022: 555)

Questions & concerns:

1. What about corpora in the Boasian trilogy?
2. What about time?
3. What about archiving?
4. What about incomplete data?

References

- Anderson, Patricia M. 2020. *Revitalization lexicography: The making of the new Tunica dictionary*. Tucson: University of Arizona Press.
- Boerger, Brenda H. & Verna Stutzman. 2018. Single-event rapid word collection workshops: Efficient, effective, empowering. *Language Documentation & Conservation* 12, 147–193.
- Chelliah, Shobhana L. & Willem J. de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. London: Springer.
- 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.
- Himmelman, Nikolaus P. 2012. Linguistic data types and the interface between language documentation and description. *Language Documentation & Conservation* 6, 187–207.
- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1), 161–196.
- Kilgariff, Adam. 2007. Word sense. In Eneko Agirre & Philip Edmonds (eds.), *Word sense disambiguation: Algorithms and applications* (Text, Speech and Language Technology vol. 33), 29–46. London: Springer.
- Lachler, Jordan & Elizabeth Pankratz. 2017. Moving towards value-added digital repatriation in lexicography for indigenous languages in Canada. In Nicholas Ostler, Vera Ferreira & Chris Moseley (eds.), *Proceedings of the 21st FEL Conference*, 107–114. Hungerford: Foundation for Endangered Languages.
- Moe, Ronald. 2003. Compiling dictionaries using semantic domains. *Lexikos* 13, 215–223.
- Nabirye, Minah. 2011. Compiling the first monolingual Lusoga dictionary. *Lexikos* 19, 177–196.
- Paterson III, Hugh. 2015. Lexical dataset archiving: An assessment of practice. Poster presented at the *4th International Conference on Language Documentation and Conservation*, Honolulu, HI, 26 February – 1 March.
- Rhodes, Richard A. & Lyle Campbell. 2018. The goals of language documentation. In Kenneth L. Rehg & Lyle Campbell (eds.), *The Oxford handbook of endangered languages*, 107–122. Oxford: Oxford University Press.
- Rundell, Michael. 2012. ‘It works in practice but will it work in theory?’ The uneasy relationship between lexicography and matters theoretical. In Ruth Vatvedt Fjeld & Julie Matilde Torjusén (eds.), *Proceedings of the 15th EURALEX International Congress*, 47–92. Oslo: Department of Linguistics & Scandinavian Studies, University of Oslo.
- Wolff, John U. 1991. Lexicography of the languages of Indonesia aside from Indonesian and Javanese. In Franz Josef Hausmann et al. (eds.), *Wörterbücher* [Dictionaries], vol. 3, 2566–2568. Berlin: Walter de Gruyter.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. *Language Documentation and Description* 1, 35–51.