1. Introduction

separate as separate fields of inquiry:

(1) collection, transcription, and translation of primary data
(2) analysis of data

can be viewed as

(1) \( \rightarrow \) (2) [analysis is ancillary to collection]
(2) \( \rightarrow \) (1) [collection is ancillary to analysis]

2. Distinguishing description and documentation

_description:_ - traditionally conflated with documentation, since the describer was in charge of collecting and organizing documents
- traditionally contrasted with prescriptive, historical/comparative, generative/formal

Himmelmann’s characterization of difference between _documentation_ and _description_

<table>
<thead>
<tr>
<th>DOCUMENTATION (COLLECTION)</th>
<th>ANALYSIS (DESCRIPTION)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULT</td>
<td></td>
</tr>
<tr>
<td>corpus of utterances; notes on observations and comments by speaker and compiler on a particular form or construction</td>
<td>descriptive statements, illustrated by one or two examples</td>
</tr>
<tr>
<td>PROCEDURES</td>
<td></td>
</tr>
<tr>
<td>participant observation, elicitation, recording; transcription and translation of primary data</td>
<td>phonetic, phonological, morpho-syntactic, and semantic analyses (spectrograms, distributional tests, etc.)</td>
</tr>
<tr>
<td>METHODOLOGICAL ISSUES</td>
<td></td>
</tr>
<tr>
<td>sampling, reliability, naturalness</td>
<td>definition of terms and levels, justification (adequacy) of analysis</td>
</tr>
</tbody>
</table>

Reasons for separating _description_ from _documentation_ [somewhat edited and rearranged by RGS]

(1) Description skews choice in types of data to collect; focus on data that is “relevant”
(2) Description limits corpus to use by linguists
(3) Focus on description distracts from methodological issues (naturalness, comprehensiveness, ethics, etc.); publication of primary data is not a priority
(4) Description focuses on “data as it can be written down” (ignores paralinguistic factors, social context)
Footnote 8: • In description, data is appended as selected examples
  • Skewing of document types: 10 grammars to 3 dictionaries to 1 primary
data collection, corresponding to relative prestige

3. Language Documentation and Documentary Linguistics

3.1. Basic assumptions

“The aim of a language documentation … is to provide a comprehensive record of the
linguistic practices characteristic of a given speech community.” (p. 9)

[Another comparison of description and documentation]

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>DOCUMENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>a record of the formal properties that</td>
<td>a record of the linguistic practices and</td>
</tr>
<tr>
<td>constitute the language</td>
<td>traditions of a speech community</td>
</tr>
<tr>
<td>focuses on the single goal of describing</td>
<td>unifies aspects of various approaches to</td>
</tr>
<tr>
<td>the language system</td>
<td>language, i.e. all the potential audiences for</td>
</tr>
<tr>
<td></td>
<td>the documents (synthesizing a coherent</td>
</tr>
<tr>
<td></td>
<td>framework is a major theoretical challenge)¹</td>
</tr>
<tr>
<td>analytic statements are the central</td>
<td>presents primary data in full, where</td>
</tr>
<tr>
<td>theme, with data used only for</td>
<td>analysis consists of commentary or</td>
</tr>
<tr>
<td>exemplification</td>
<td>apparatus² appended</td>
</tr>
</tbody>
</table>

3.2. The basic format of a language documentation

Three basic kinds of content:

(1) Communicative events: texts, etc. including all the naturally occurring aspects
(location, artifacts, gestures, etc.), speaker meta-linguistic commentary

(2) Lists: Taxonomies, etc. that will never appear fully manifest in communicative
events, such as paradigms, list of artifacts in various domains, flora & fauna)

(3) Analytic matters: Commentary and information supplied by speakers

Each type of content should be accompanied by:
  raw data (video, audio)
  translation (word-by-word and/or free)
  commentary (metadata about the event)

The language documentation format is uncompromisingly data-driven

(1) No primary data is excluded as “not relevant”
(2) Presentation is organized around the documents
  - Grammar points, etc. are noted within the annotation

¹ See page 10 for a bullet point list of types of audiences or pursuits that might utilize documentation.
² Apparatus is a term, apparently from traditional philology, which means something like “information
  about the document and/or help in understanding the document”. This would include such things as
  metadata (time, place, and conditions of recording, equipment used, etc.), annotations (transcriptions,
  translations, word for word glosses, explanatory notes), and general resources (cultural information, sketch
  grammar, etc.). Himmelmann, on page 14 of GHM, gives a table outlining types of apparatus.
- Points are not extracted and organized into “chapters”
- May seem inconvenient, but makes a document relevant to many users
(3) Documentation requires deep knowledge of the culture and of various disciplines

4. Issues in documentary linguistics

“Discussion and theorizing about all the considerations relevant in language documenta-
tion is the concern of documentary linguistics.” (p. 15)

- which data to collect
- recording data (“quality” of the record in terms of naturalness)
- transcription, translation, commentary
- presentation for use by others

4.1. Limits to documentation: rights of privacy and language rights

“…the interests and rights of contributors and the speech community should take
precedence over scientific interests.” (p. 16)

[Himmelmann devotes a certain amount of ink to issues of secrecy (separation of
“private” language activity from activity available for public consumption), copyright
(who “owns” documents and distribution rights), and characteristics of communities
where these are or are not issues. We will not be concerned with this much in our class,
but it does become a concern if one engages in field work.]

4.2. Parameters for the selection of communicative events

the question of “what to collect” needs “a principled, theoretically informed answer”

needed: “a systematics for communicative events”, but “a universally applicable grid of
texttypes” is probably not feasible

[Suggests crosscutting parameters of spontaneity and modality as a way to systematize
text types. The table on the next page synthesizes Himmelmann’s figure 3 and 4.]

- a principled schema helps researchers identify gaps in data
- not all speakers will be equally skilled at any or all levels
- some of the major sub-types will have more variety of types than others
- variety: as much variety should be documented as possible
- quantity: is desirable; repetition assures authenticity and non-idiomsyncrasy
<table>
<thead>
<tr>
<th>PARAMETER OF SPONTANEITY</th>
<th>PARAMETER OF MODALITY*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAJOR TYPES</td>
<td>EXAMPLES</td>
</tr>
<tr>
<td>Exclamative</td>
<td>‘ouch!’</td>
</tr>
<tr>
<td>Directive</td>
<td>‘fire!’</td>
</tr>
<tr>
<td>Conversational</td>
<td>‘scalpel!’</td>
</tr>
<tr>
<td></td>
<td>‘come here!’</td>
</tr>
<tr>
<td></td>
<td>greetings</td>
</tr>
<tr>
<td></td>
<td>small talk</td>
</tr>
<tr>
<td></td>
<td>chat</td>
</tr>
<tr>
<td></td>
<td>discussion</td>
</tr>
<tr>
<td></td>
<td>interview</td>
</tr>
<tr>
<td></td>
<td>narrative</td>
</tr>
<tr>
<td></td>
<td>description</td>
</tr>
<tr>
<td></td>
<td>speech</td>
</tr>
<tr>
<td></td>
<td>formal address</td>
</tr>
<tr>
<td></td>
<td>litany</td>
</tr>
<tr>
<td></td>
<td>memorized vocables</td>
</tr>
</tbody>
</table>

*Another possible modality might be singing.

4.3. Quality of data (gathering procedures)

[“Quality” is equated with “naturalness”, i.e. the less self-conscious and less documenter-controlled, the greater the likelihood of “good quality” = “natural” data.]

“…an important area of practical and theoretical inquiry within documentry linguistics is concerned with the evaluation and development of data-gathering procedures.” (p. 26)

observer’s paradox: The investigator becomes part of the context. “The object of research is susceptible to change because of the ongoing research process (the presence of the researchers and/or their research tools, etc.).” (p. 26)

props: (1) Props direct linguistic behavior without focusing on the behavior itself.

(2) Props allow direct comparison across genres and across languages.

BUT

(1) Props may not be culturally applicable, thus not generate natural discourse.

(2) Quality of data collected using props is not yet well-understood, e.g. they may generate new discourse types.
elicitation (the traditional manner of data collection in descriptive linguistics:

1) May introduce an new and unfamiliar type of communicative event.
2) With elicitation through another language, “one must always reckon with strong
interferences from the language in which the conversation is conducted.”

4.4. Further issues [not developed in detail in this paper]

[aspects of language documentation in addition to rights, selection, quality]

• involvement of the language community in the documentation process
• the technology used in recording and preservation—concern with obsolescence
• preparation of documentation (transcription, translation, commentary/apparatus)
• funding for documentation