



**HAL**  
open science

## Researching endangered languages: Critical reflections on field and documentary Linguistics

Bettina M Migge

► **To cite this version:**

Bettina M Migge. Researching endangered languages: Critical reflections on field and documentary Linguistics. Colonial and Decolonial Linguistics – Knowledges and Epistemes edited by Storch, Anne; Deumert, Ana; Shepherd, Nick. Oxford, UK: Oxford University Press, 157-175., pp.157-175, 2020. hal-03085550

**HAL Id: hal-03085550**

**<https://hal.science/hal-03085550>**

Submitted on 21 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Researching endangered languages

### Critical reflections on field and documentary linguistics

*Bettina Migge*

#### <H1>10.1 Introduction<H1>

<P>The study of endangered languages,<sup>1</sup> and particularly those of the global south, has been driven by non-speakers of these languages who are of European or North American origin, and/or by researchers who were trained at such institutions. They typically pursue larger Western professional goals, or moral ideals. Christian missionaries, who have “produced the single largest body of knowledge about linguistic diversity around the world” (Errington 2008: 13), studied and ‘fixed’ non-European languages as instruments of conversion (Deumert, Chapter 11 this volume). Colonial governments often deferred to linguists for help with exercising power over colonial subjects

---

<FN><sup>1</sup> The term ‘endangered language’ is currently widely used to refer to languages whose continued use is under threat due to social, political, demographic, and other factors. The literature emphasizes that endangerment constitutes a continuum and has developed specific parameters for measuring it (e.g. *Ethnologue*, *Oxford Encyclopedia*, Austin and Sallabank 2014). Since ‘endangered’ has strong overtones of ‘emergency’ and ‘salvation’ (Hill 2002; Duchêne and Heller 2007) and is most typically used to refer to highly endangered languages (see discussion below), other organizations, such as the European Union, make use of related terms such as ‘minority’, ‘lesser-used’, or ‘marginalized’ languages. These terms are, of course, also problematic because “[t]he key criterion here is the size of the speaker population *within a specific geographic context*: an individual language may be a minority language in one region or state but a majority language in another” (Grenoble and Singerman 2014).

(Irvine 1993; Pennycook 1998). Missionaries and lay people were often driven by personal fascination with the proverbial ‘other’, as well as broader concerns such as the survival of traditional ways of life. Their representations of language practices and linguistic contexts were frequently reductive and skewed (Irvine and Gal 2000; Deumert and Storch, Chapter 1 this volume) but are often the main or only (reference) work. In some instances, missionaries’ linguistic works even led to the creation of new languages (e.g. Fabian 1986; Errington 2008; Makoni 1998b). In the twenty-first century, religious proselytizing continues to be an important reason for studying endangered languages (Stoll 1982; Errington 2008: Chapter 7) and the Summer Institute of Linguistics (SIL), “a faith-based non-profit organization”<sup>2</sup>, describes itself as a “leader in the identification and documentation of the world’s languages.” Its work overlaps with field linguistics, as many members of SIL actively contribute to knowledge production in field linguistics.<sup>3</sup>

<PI>Field or descriptive linguists engage in the description of languages to enrich existing knowledge about structural linguistic diversity in order “to make deductions about what is and what is not possible in human language.” (Crowley 2007: 12–13). In the 1990s, a new tradition called documentary linguistics emerged which has its roots in the growing concern with the loss of biodiversity (Himmelmann 2008). Practitioners advocate on behalf of endangered languages and see their documentation as “the defining project or disciplinary charter” (Woodbury 2011: 162). They have mobilized significant funds from private and public institutions such as the *Volkswagenstiftung* (DoBeS), the Rausing Fund or Arcadia, and the National Science Foundation of the United States (DEL—Documenting of Endangered Languages) for language documentation activities, including the establishment of archives. One of the hallmarks of documentary linguistics is its focus on the collection and curation of primary language data (Himmelmann 1998: 161). Finally, there is also a growing number of scholars in linguistic anthropology and postcolonial

---

<sup>2</sup> See [www.sil.org](http://www.sil.org).

<sup>3</sup> See [www.sil.org/about/discover/academic-research](http://www.sil.org/about/discover/academic-research).

studies (e.g. Irvine 1993; Silverstein 1998; Irvine and Gal 2002; Hill 2002; Duchêne and Heller 2007; Errington 2008; Kroskrity and Avineri 2014; Deumert and Storch forthcoming) and endangered language research (e.g. papers in Makoni and Pennycook 2007; Grenoble and Furbee 2010; Pérez Báez et al. 2016) who are critically assessing the goals, methods, practices, premises, and effects of research on and discourses about language endangerment and revitalization.

<PI>The different traditions—particularly missionary, field, and documentary linguistics—appear to constitute a developmental trajectory or continuum, as each positions itself somewhat differently with regard to the language, its user communities, and the nature of the enterprise (goal, audience, data, data collection, data analysis) owing to critical engagement with (aspects of) the respective other tradition(s), and engagement in somewhat different sociocultural contexts. The relationships between these traditions are also dynamic and have evolved, but there are important similarities and constants. This chapter examines some of them, focusing on their instantiation in field and documentary linguistics. As they are often at the heart of the linguistic enterprise and the professional position of the linguist, they are rarely recognized as ideological. The chapter investigates four broad themes: the research goals of these traditions, the research methodology, the research outcomes, and finally the agents of the research. Discussing each theme in turn, it explores the changes that have occurred in research on endangered languages and whether they have transformed research practice, liberating it from its colonial beginnings. I will examine the discourses in the (classic) training manuals or guides that have a continued presence in the ‘discipline’ and represent what is imparted to novices.<sup>4</sup>

## <H1>10.2 The goals of research on endangered languages<H1>

---

<FN><sup>4</sup> From interaction with practitioners, my own sense is that works from the critical tradition are not centre stage, if treated at all, in training or in the representation of the ‘discipline’. While writers such as R. M. W Dixon have also produced works that go well beyond this, curiously this is not reflected in the training guides.

<P>Introductory-level linguistics textbooks typically argue that the goal of linguistics is to understand the structure of language or how language works (Yule 2017: X). The focus is on structural linguistic aspects of language such as types of sounds or morphosyntactic phenomena and regularities, but sociolinguistic and pragmatic matters nowadays also receive some mention. The crucial aims are to refine current linguistic knowledge and to push existing frontiers of knowledge, in order to enrich our databases, to challenge the novice, and to further disciplinary theorizing about language (Deumert and Storch 2019). They are part and parcel of the objectifying and enumerative approach to language, which focuses on reducing languages into manageable and measurable objects (Makoni and Pennycook 2007), and on developing objective criteria for comparing and measuring the relative value of languages (see, for instance, the popular dimension of relative simplicity). These objectified goals leave little or no room for local or subjective motivations for studying a language, such as one's love for it, one's desire to keep it alive, to prove to doubting others that it has value, or to empower fellow speakers. This is surprising, because these 'personal' motivations play an important role in the lives of many linguists; they are often a/the deciding factor for engaging in linguistics training in the first place, and they may give us the strength to spend so much of our time beyond our regular working hours engaging with linguistics. They are erased from the centre of our academic work and are banished to the margins. Although I have discovered and also shared many such stories 'off duty', perhaps over a drink at the end of a conference day or when travelling with colleagues, when discussed in conjunction with academic work—'on duty'—personal motivations arouse suspicion and are dismissed as being too interested, irrelevant, or politicized, as they disrupt the professional position of the researcher; the hidden fear is that if researchers do not engage in this "self-alienating dynamics" (A. Storch, pers. comm. July

2018), it will cloud their objectivity and discredit them as researchers.<sup>5</sup>

<PI>A case in point is the experiences of a Native American anthropologist. He decided to study anthropology to “understand my place in the world, and help my community communicate our struggles for survival” (Ranco 2006: 61). He found, however, that this contrasted with his non-native American anthropologist peers who “seemed intent on traveling the world and experiencing new things” (ibid.). He was frequently cautioned against doing research on his own community as that would prevent him from attaining the comparative (or othering or outsider) perspective that is inherent to the anthropological endeavour (Ranco 2006: 66); his insider perspective and interested agenda were seen as hampering his academic development. He also found that delivering on the highly valued abstract themes of the discipline required employing the othering strategy and produced outcomes of little interest among his fellow community members. By contrast, the community’s (and his) interests in researching the challenging sociopolitical context of minority communities, such as how to “deal with the ways ‘local elites’ process and control our access to data” (Ranco 2006: 72), were welcomed only as applied or outreach research. They were considered inappropriate as major research topics and for furthering an academic career. He found himself being pushed to obtain new types of data on his (and similar) communities. This was, however, of little interest to his community and forced him to talk about his friends and family in an objectivizing tone.

<PI>Notions such as objectivity, empirical grounding, and the need for higher-level goals in research on (endangered) languages have also remained largely unreflected epistemes “that [are] informed by puritan, protestant codes of conduct” and show a lack of generosity and empathy (A. Storch, pers. comm. July 2018). Widely read field-linguistics guides highlight the following goals:

---

<FN><sup>5</sup> In my own experience, ‘outbursts of emotionality’ are acceptable during liminal periods: for beginning researchers, especially if they come from marginalized communities, and for accomplished, retiring ones as part of reminiscing about their career. They cause embarrassment in other periods.

1. <NL> to “provide the fodder upon which many armchair linguists depend” to answer “the Big Questions of our science—how different can languages be? and how similar must different languages be?” (Crowley 2007: 12–13).
2. to obtain reliable descriptive observations from as broad a range of languages as possible in order to make deductions about what is and what is not possible in human language” (Crowley 2007: 12–13).
3. to “want to learn linguistics and love the challenge of working on a new language” as “to learn any discipline is to get out there and do it [...]—through using them [the linguistic tools] to describe a language” (Dixon 2010a: 310–11).<NL>

<P>Documentary linguists also pursue such broader aims but phrase them in a more inclusive fashion:

1. <NL>to “consolidat[e] and enlarg[e] the empirical basis of a number of disciplines” (Himmelman 2006: 5).
2. “[m]ore and more communities have sought documentation of their languages just as they slip away [...] and] the communities themselves are generally small in terms of population (Woodbury 2011: 160).<NL>

<P>Some texts even boldly decry subjective reasons for engaging with endangered languages: “[n]o one should undertake fieldwork simply because they feel a social responsibility to document some language before it disappears” (Dixon 2010a: 310). Yet the personal pleasure of the objective linguist, which is arguably subjective and harks back to ‘the man as hunter’ image, is promoted as a

good reason for describing languages:

<EXT>*Because one enjoys the intellectual excitement of working on a 'new language'. A linguist must take pleasure in what they are doing to be able to do it well and produce worthwhile and exciting results ... and "others will draw pleasure from reading the grammar."*<EXT> <EXT-S>Dixon 2010a: 310; emphasis in the original<EXT-S>

<P>Engagement with speakers' needs and desires is generally relegated to questions of fieldwork ethics. Field linguists are rightly told to "only go into a community which welcomes them and the work they are doing" (Dixon 2010a: 311). It is not clear how this can be easily determined, as most people are not familiar with linguistics. Engagement with community interests is, however, mostly constructed as being ancillary to descriptive work and is presented mainly as a way of thanking people for allowing access to their language.

<PI>Documentary linguists are less explicit. However, the fact that Woodbury (2011: 160), for instance, feels a need to justify the desire to save endangered languages by invoking linguistic reasons—"the languages for which documentation is sought show ever more genetic and typological diversity from one to the next" and "many linguists and scholars in related fields have been inspired by the human and scientific dimension of the issue"—suggests that subjective reasons can be entertained only when they coincide with higher-level ones. Thus, both traditions position their research as a service to science, and more broadly to mankind (see Hill 2002 and Duchêne and Heller 2007 for critical assessments of such discourses).

<PI>Documentary linguistics also differs from field linguistics in that it is interested in uniting and integrating the aims of different 'stakeholders' (community members, linguists and other practitioners, and academics). Following Boas' tradition of ethnology, documentary linguists



advocate in favour of the importance of interdisciplinarity (cf. Austin and Sallabank 2014; Grenoble and Furbee 2010). One way in which this is facilitated is by focusing on the collection and curation of primary data and postponing in-depth analysis. Odé (2015: 156), however, argues that linguistically untrained speakers of endangered languages find texts with interlinear glosses and translations as opaque as descriptive materials. Ranco (2006) also maintains that members of endangered communities are more interested in information about the dominant culture, the operation of its (oppressing) institutions, and pointed analysis of how their power can be undermined, rather than raw data about their own community. Such politically engaged studies, however, fall outside the purview of the typical documentary enterprise, which is concerned with linguistic and community-internal phenomena (but see critical research by e.g. Kroskrity and Avineri 2014). According to Ranco (2006: 66), these data often feel as if they are providing “linguistic cages for the experiences” of fellow community members, that can and have been used against the community. The emphasis on amassing primary data, now of a more diverse and sophisticated type, however, fits perfectly with the colonial agenda and does not really change practices of objectification (Deumert and Storch 2019).

<PI>Finally, both traditions determine the need for linguistic research ‘intervention’, based on linguistic typologies that rank languages using fuzzy linguistic notions such as linguistic complexity and degree of cross-linguistic markedness (Dixon 2010a: 312). Crowley (2007: 58) also rightly invokes researcher-based (language competences, personal contacts, personal lifestyle issues, personal language preferences), ‘practical’ (e.g. accessibility, “stepping on other people’s [researchers’] toes” (p. 60)), and discipline-based (e.g. information about the language family; rarity of linguistic features) considerations that guide the selection of languages for descriptive work. Again, the needs, desires, and interests of the speakers of endangered languages do not figure in these lists, but see discussion fora such as Indigenous Languages and Technology (ILAT)

(Deumert, pers. comm. 2017). These considerations mean that a wide range of languages, and particularly those which emerged through contact, which are still vibrant, but whose speakers need support to ensure their survival, are not eligible for funding from language documentation funds (Garrett 2006). It would indeed be interesting to research the projects submitted to funding agencies since the 1990s to see how different factors have impacted funding decisions.

<PI>To follow Ranco's (2006) argument, research on endangered languages should give equal weight to linguistic and sociocultural research. The documentation and/or description of a language should also include investigation of the larger social, political, and cultural context and the dialectic relationship between the global, national, and community levels to identify the factors that promote and inhibit endangerment, to enable the development of effective actions to reverse it, and to critically assess these actions from both emic and etic perspectives (see Meek 2010).

### <H1>10.3 Researching endangered languages<H1>

<P>Every academic discipline has a toolbox of principles and procedures that licensed practitioners apply in order to obtain, handle, and analyse the data that are at the heart of their academic theorizing and model-building. It is not static over time or unitary, because different research contexts present different challenges and researchers have different perspectives or ideologies. Other important impetuses for variability and change are interdisciplinarity and dialogue with the people whose societies, languages, and cultures are being researched. The latter is, however, a much more recent and contested path of change because "indigenous knowledge systems ... include experience and culture as a means of defining what is known" (Grenoble and Whitecloud 2014: 340) rather than "require that results be testable and replicable" (ibid.: 341). Traditionally, the

former have no place in the northern scientific endeavour and are “considered to be an impediment to scientific enquiry” (ibid.). However, this has been changing in recent years.

<PI>Research methods in structural linguistics that are positivistic, extractive, and descriptive, and draw mostly on a language-internal logic and cognitivist notions of language (Grenoble and Whitecloud 2014: 344). They used to dominate linguistic practice and theorizing. In this approach, language can be accessed with minimal reference to its context of occurrence or its users, and can be described using a concise set of rules and principles that reference cross-linguistic regularities. The guided production by one speaker, or a small number of speakers, provides sufficient data for theorizing. Research in historical linguistics by Herzog, Weinreich, and Labov, and in the sociology of language by Hymes and Gumperz eventually demonstrated the socially contingent nature of language development, the diversity of linguistic productions, and the need for greater engagement with society to obtain more representative linguistic data. As a result, large(r)-scale databases consisting of language from different contexts and socially stratified sets of speakers, and their exploitation using quantitative methods, became popular. With the discursive turn in the Humanities and Social Science, a greater focus on the meaning-making capacity of language and the role of non-objectifying factors that condition people’s (language) behaviour—such as Hymes’ (1972) communicative competence—has (re)-emerged. This includes micro-level perspectives on language use, and their relationship to micro- and macro-level overt and covert social structures and ideologies (Eckert 2012), including issues of positionality. Relying on ethnographic data and detailed analysis of recorded or observed situated events, research investigates the interactionally based socially constructive functionality of linguistic practices within and across contexts. Linkages with culture and belief are explored and the mechanisms that bring about indexicality, including the methods of data collection and analysis and their constructive power, are critically investigated (Silverstein 1992; Irvine and Gal 2000; Errington

2008). This approach provides insights into aspects of language that play a role in people's lives, the use of little-understood linguistic structures and elements such as pragmatic markers, and allows us to (re)examine long-standing theoretical issues. It is in many ways closer to the ways in which lay people conceptualize language, recognizing the integrated relationship between language and culture as a lived experience (Grenoble and Whitecloud 2014).

<PI>In the remainder of this section I explore whether a process of deconstruction has taken place in research on endangered languages and how this has transformed practices in the field. Field linguists propose that language documentation is a culturally and ideologically neutral practice that is distinct from modern generative linguistics. An important manual argues that “[a] linguist should work in terms of basic linguistic theory, the cumulative theoretical framework which underlies almost all grammar writing and typological generalization” and “should be a good all-round linguist” because “every part of the language description is equally important and each part interrelates with the others” (Dixon 2010a: 310). The notion of culture figures only in relation to accessing the community, with researchers being alerted to different kinds of reactions that communities may have to linguists' work, and advised on how to ‘overcome’ them and engage the community to support linguists' work (Dixon 2010a: 315–18). The impression that the reader gets is that there is one immutable framework that the researcher has to simply work through in a somewhat mechanical manner.

<PI>Documentary linguistics traditionally takes a greater interest in culture. Tracing its origin to Boas and his charter for ethnography, language is seen as intricately linked to culture and thus documentation of language entails documentation of culture because “much of the content of culture, e.g. rituals, oratory, narrative, verbal art and onomastics, was linguistic in nature” (Woodbury 2011: 163). Documentary linguistics does not prescribe a particular method of data collection, as “[t]he data-gathering techniques used for any individual project were established

through conferencing among members of the group and through consultation with the [community]" (2011: 175). This inclusive approach presumably puts linguistic skills to practical use, allows greater community engagement and thereby facilitates deeper insights into the culture and its relationship to language within a community. The impression one gets though is that there are few models available. A number of programmatic papers (e.g. Childs et al. 2014), the existence of critical research on language endangerment and revitalization, and more ethnographically oriented works by some researchers attest to (growing) awareness among researchers, but this is not reflected in the (more?) widely read guides to both field and documentary linguistics. They suggest that training in these areas is not of central concern.

<PI>Change is evident in documentary linguistics with respect to the kinds of things being studied. Unlike field linguistics, which appears to equate language with grammatical structures, defined as the phonology, morphology, syntax, and, usually to a lesser extent, the lexicon of a language (Crowley 2007: 128), documentary linguistics also recognizes "a number of different, though interrelated phenomena," including "registers and varieties, social and local," and "language as a social practice as well as a cognitive faculty" (Himmelmann 2006: 2). One aspect that has recently come to the fore is language ideologies (see, for example, the papers in Austin and Sallabank 2014), that is, speakers' (and non-speakers') beliefs about a language, including its relationship to people, places, contexts, and other related languages (see Di Carlo and Good 2014), and its role in representing people and structuring communities (Kroskrity and Avineri 2014). Research into language ideologies offers insights into the processes that support or undermine the continued use of a language and is crucial for devising revitalization strategies, for instance. It is not clear though whether this would be part of the main endeavour or whether such research would be undertaken by anthropological and educational collaborators as suggested by Gippert et al. (2006).

<PI>For lay people, languages are emotionally charged means of self-expression that are instrumental for negotiating belonging and identity. These pragmatic aspects of language, including the social dynamics in which they are embedded, are central to the functioning, the choices, and the passion that language(s) inspire. However, such phenomena, including mediated language phenomena (writing, electronic media, signage) and linguistic politeness, communicative repertoires etc. do not generally figure in mainstream descriptive and documentary discussions (see Deumert and Storch 2019; Deumert 2014). There is also little discussion about local conceptualizations of language and local linguistic practices, including their fuzziness, and methodological implications that arise for studying and representing languages or linguistic practices in such contexts (Migge and Léglise 2013).

<PI>Documentary linguistic traditions also privilege a closed system approach to language as there is often a focus on monolingual practices (Childs et al. 2014). This contrasts with the non-monolingual social reality of most communities. If we want to know why languages are being maintained or not, we need to understand how people use them and the emic and ethic ideologies, as they most certainly impact practice.

<PI>For the field linguist, the task of fieldwork appears to be mostly mechanical. Contrasting anthropological with linguistic field research, Crowley argues that linguists have a precise goal and object of study:

<EXT>We know, for example, that whatever language we are going to study, it must have phonemes. We won't know how many phonemes there will be, or what shape they might have, but we can be absolutely certain that there will be phonemes.<EXT> <EXT-S>Crowley 2007: 20<EXT-S>

<P>The main task of field research is fine-tuning the known facts, as there is “some kind of broad framework upon which we can hang the various facts that we are about to discover” (2007: 20). This is indicative of the largely behaviourist conceptualization of research, where the researcher attempts to explain observable facts solely on the basis of objective, discipline-internal reasoning. Research parameters are largely based on knowledge generated by other (northern) researchers and the researcher essentially induces speaker(s) to ‘fill in the blanks’ and to confirm existing knowledge. Documentary linguists also aim to describe a language and to develop a broader knowledge base of the language to integrate the needs of speakers and other interested parties. There appears to be no fixed method for doing this, though. This is presumably negotiated locally with other interested parties, including speakers. Little is known about this process, but the finalized descriptions suggest that the structurally oriented linguist dominates the process and outcome.

<PI>Traditionally, the dominant method for obtaining information about a language was elicitation, where the researcher and a speaker of the language engage in a socially asymmetrical one-on-one encounter. Its aim was to isolate the language system from the messiness of social practice and to reduce it to an orderly, rule-governed object that could easily be categorized, ordered, and dominated (Errington 2008; Irvine 2008; Deumert and Storch, Chapter 1 this volume). The researcher presented words and constructions in a shared language, usually an ex-colonial one, to the speaker of the language; the latter then rendered these in the language being researched and the researcher then examined their structure and makeup to devise new, related, and increasingly more complex constructions for acceptability assessment by speakers. Although this interaction was typically performed bilingually, only monolingual phrases and sentences were accepted. This artificial back and forth, which is akin to formal language learning and often felt like an interrogation or disciplinary action from the point of view of the speaker, was often the main context in which information about the language and actual samples of language were generated.

These researcher-driven artificially constructed language samples were then reproduced in grammars and dictionaries to illustrate the language, and ultimately came to represent the language and the people who spoke it (Deumert and Storch, Chapter 1 this volume).

<PI>The desire to impose (a new) order and to remove a language from the realm of its local mundane contingencies to a laboratory-like context is further underlined by the liminal physical context in which the elicitation exercise is often conducted (Irvine 2008; Deumert and Storch, Chapter 1 this volume). In one textbook, novices are advised to set up their working environment as follows: “For [your own] personal comfort, you will need a table with a smooth surface of the right working height, a stable chair with a comfortable back, good light, and good protection from the sun and rain” (Crowley 2007: 94). These kinds of idealized laboratory and socially marginal conditions further emphasize the distinctiveness of the descriptive exercise from community practice. They are likely to have an alienating effect on speakers, often leading to language samples that bear little resemblance to community practice. They also further emphasize the asymmetrical and unidirectional nature of the relationship between the Northern researcher and the people from the South, who are essentially relegated to the position of informants.

<PI>Although documentary linguistics manuals promote the use of diverse data-collection methods, a fair amount of the information that goes into grammar sketches still relies on elicitation, as does initial contact with the language. Curiously, nothing is said about how researchers’ descriptive practices should be adapted to reflect a collaborative context. Existing guides say little or nothing about the methods and critical assessment of the very nature or design of existing data-collection methods, and there is no critical reflection of their own positionality (e.g. Gippert et al. 2006: vi) beyond the need for collaboration. It is not clear whether, or how, speakers’ views of the process are integrated, nor whether, or how, they get to tell their own story (Deumert and Storch, Chapter 1 this volume). Discussions on data-collection methods instead focus a great deal on legal



and technological details (see e.g. the articles in Gippert et al. 2006; Grenoble and Furbee 2010), which also functions to confirm the scientific and professional status of the documentation enterprise in Northern scientific perspectives.

<PI>Documentary and descriptive approaches also rely on “a corpus of continuous spontaneous speech” (Crowley 2007: 121). The language description tradition has a utilitarian approach. Texts are essential for verifying the insights derived from elicitation work and for fine-tuning the researcher’s “understanding of the grammar of a language,” “how sentences are linked together in discourse,” and for providing “very valuable lexical information” (Crowley 2007: 108–9). This information feeds into elicitation work, allowing the generation of new constructions and vocabulary. Researchers are advised, however, not to spend too much effort on recording, as “by the time that you have recorded a certain amount of speech, each individual text will provide fewer and fewer new pieces of lexical [and structural] information” (Crowley 2007: 109).

<PI>Documentary linguists, in contrast to field linguists, argue “that the collection, transcription and translation of primary data” (Himmelman 1998: 161) is distinct from descriptive activities, namely the analysis of language: “[T]he interrelation between the two activities is no longer seen as one of unilateral dependency, with the documentary activity being ancillary to the descriptive activity (i.e. primary data are collected IN ORDER TO make a descriptive statement of the language)” (Himmelman 1998: 165). Documentary linguists promote a bottom-up, empirical approach in which the collection and preparation of texts from different social contexts is the focus of the research, rather than the provision of data for (or against) specific linguistic theories (Woodbury 2011: 163). For some documentary linguists, such as Himmelman, analysis is vital only insofar as it supports the consumption of the texts by the various stakeholders. Others argue that the analysis, collection, and preparation of texts should happen concurrently and feed into one another, as envisaged in Boas’ approach (Woodbury 2011). However, both descriptive and

documentary linguists agree that grammar writing should not be the primary focus of the documentation enterprise.

<PI>Both approaches have an extractive approach to texts (Ranco 2006); data is gathered to satisfy professional research needs. Crowley (2007: 124ff.), for instance, advises researchers to start out with very simple short texts and to move to more complex ones when the researcher's knowledge of the language has increased. The content does not receive much attention, as researchers are told to record what people want to talk about and to document aspects of the community's oral tradition to give back to the community. Texts are to be obtained in a laboratory-type, and socially and aurally sanitized, environment where the narrator narrates only for the researcher, rather than in habitual context.

<PI>In the documentary linguistics tradition, where the creation of a multipurpose record of a language that "leaves nothing to be desired by later generations wanting to explore whatever aspect of the language they are interested in" (Himmelman 2006: 3) is at the heart of the enterprise, the researcher is advised to:

<EXT>[S]trive to include as many and as varied records as practically feasible, covering all aspects of the set of interrelated phenomena commonly called a language. Ideally, then, a language documentation would cover all registers and varieties, social or local; it would contain evidence for language as a social practice as well as a cognitive faculty; it would include specimens of spoken and written language; and so on.<EXT> <EXT-S>Himmelman 2006: 2; but see also Berge 2010<EXT-S>

<P>This highlights a concern for an inclusive and rich representation of language. However, it has a rather enumerative sense to it: the wider range of entities is important in order to obtain as

complete a data set as possible, with the promise of new phenomena for theorizing. This change in focus does not, however, appear to have had a transformative effect on research methods as there is no discussion about the functions of these texts, what constitutes ‘appropriate’ texts, or how to collect them. This work is delegated to socially oriented researchers such as anthropologists (Franchetto 2006), and to collaborations with the community. Although this is not overtly discussed, the data appear to come mostly from community-internal interactions, leaving cross-community interactions and practices aside.

## <H1>10.4 Representing endangered languages<H1>

<P>Field linguists produce written documents that are produced professionally and for profit by international commercial publishing houses, which guarantee their easy availability in a high-quality format for consumption by as large an academic audience as possible. For field linguists this is a prerequisite of their work, as the ‘generation’ of knowledge makes sense only if it is shared with others and used to advance the current typological knowledge base; it is also the only way that they can advance their careers, and justify their time investment to the institutions that pay their salaries (Crowley 2007). Community members, including those who provide the language data, are, however, often not aware of this logic and in fact, might object to it on various grounds when they become aware of it. They might be reluctant to share community knowledge with outsiders and see it as a ‘sell-out’ or as a money-making ploy by Northern researchers and their institutions (Hill 2006). The skewing towards the Northern researcher community is further supported by the very nature of field linguistics publications. Very often the outcomes of field-linguistic research are not available where the language is spoken because such places often provide insufficient commercial opportunities for Northern publishers, who generally market their publications at prohibitive prices.

Apart from requiring high levels of literacy, these publications are often also inaccessible for speakers of the language (and other local stakeholders) because they are written in a major European language, such as English. This makes them easily accessible to a large financially able audience, regardless of whether that language is spoken locally. Moreover, they are written in a style that is unlikely to be accessible to people who have no training in linguistics and might have low levels of literacy (Crowley 2007; Deumert and Storch, Chapter 1 this volume).<sup>6</sup>

<PI>Among the three main products ('the trilogy') that result from field-linguistic research, grammars are accorded the highest prestige because they provide the basis for answering the big (structuralist) questions in linguistics. They are a recognized publication and can be exploited for further publications in academic journals. Their content also easily feeds into the production of new linguistic knowledge via comparative studies and databases (such as *The World Atlas of Language Structures*; <https://wals.info/>). They are, however, of little interest to community members because "most grammars are by and large not comprehensible to non-specialists" (Crowley 2007: 34). Non-linguists are generally not familiar with how linguists reduce language to separate levels of description, and the simple and generic sample sentences contained in grammars often have an alienating effect on speakers (Deumert and Storch, Chapter 1 this volume).

<PI>Field linguists produce dictionaries and text collections to engage the speaker community. These texts correspond much more closely to speakers' concern for meaning-making. However, they have largely become a marginalized activity in academic circles and are considered to be ancillary to the main work, as they are difficult to publish or to exploit for academic journals and attract little academic credit. They are also rarely written in locally accessible languages because, originating from North America or Europe, linguists usually lack sufficient expertise and North American or European publishers and funders publish only works in widely spoken

---

<FN><sup>6</sup> Grammars that adopt a more accessible format for use by the wider public are labelled 'pedagogical grammars' and are assigned a lower academic value by professional linguists.

languages unless they receive substantial subsidies. A compromise solution is to collaborate with community members who are literate in both local and ex-colonial languages (see Grenoble and Whitecloud 2014). Text collections were traditionally seldom published in their entirety; only a selection appears in the appendix of a grammar to illustrate the spontaneous use of the rules and principles discussed in the grammar (Goury and Migge 2003/2017). The texts are selected for linguistic rather than cultural reasons and come with little or no metadata about their origin (e.g. author, narrator, location, purpose). Despite efforts at being inclusive, it is clear that the outputs are geared to the needs and interests of professional and usually Northern audiences.

<PI>Documentary linguistics aims to mitigate the researcher-driven agenda by creating “a lasting, multifunctional record of the linguistic practices attested at a given time in a given speech community and the knowledge speakers have about these practices” (Himmelmann 2008: 346). While there is still debate about what this lasting, multifunctional record should consist of (Himmelmann 2006: 7–14), there is agreement that it must involve an annotated and fully curated, openly accessible electronic corpus of texts of a language (via a recognized language archive). The corpus should contain both evidence of representative “events that are likely to take place even in the absence of any person documenting it,” and evidence of “the structural possibilities of the language in question” (Schultze-Berndt 2006: 214). Its function is to represent “observable linguistic behavior, i.e. examples of how the people actually communicate with each other” (Himmelmann 2006: 7).

<PI>A curated electronic corpus of texts corresponds much more closely to non-linguists’ or speakers’ understandings of language as segments of language are not artificially detached from continuous speech but are retained in their entirety. However, the focus is still largely utilitarian. First, the sociocultural relevance of the texts or their aesthetic value receives little systematic attention because the texts are treated as spontaneous speech samples. Second, speech events are

rarely fully contextualized, as corpora rarely contain a full ethnographic description of the community, and there is usually no explanation of why particular recordings are part of the corpus and why others were omitted. So we do not have a clear understanding of the social functions of texts, the wider circumstances of their performance, or community members' assessments of these texts (Keenan 1974). They often also only include one instance of a particular type of situated language use, rather than multiple examples, which helps to perpetuate the widely held view that southern societies, including their speech habits, are static (Keenan 1974; Deumert, Chapter 11 this volume). More recent and critical literature has arguably paid closer attention to meaning making (Grenoble and Furbee 2010; Austin and Sallabank 2014; Kroskrity and Avineri 2014), studying how social practices are an integral part of and contribute to the maintenance of an endangered language. However, it appears to me that this kind of research is seen as additional to the core of the actual description/documentation enterprise and has not yet been mainstreamed in research and teaching to the point that we can speak of a paradigm shift. The main goal is the production of an annotated corpus of audio recordings and related materials.

<PI>An important feature of documentary linguistics is that the corpus of recordings is made available as sound and/or video files rather than just in written form. This makes the language samples livelier, as they preserve a greater number of aspects of their spontaneous performance, and their consumption does not rely on literacy, allowing the samples to be revisited, reinterpreted, and used by different sets of other professionals and lay people. However, there are also significant drawbacks. Access depends on electricity, and on technology which is still not readily available in many parts of the world (Deumert 2014). Accessing language archives also requires familiarity with the use of technology (security measures, permissions, data-mining tools) and a detailed knowledge of major languages. Given these issues and the need for linguistic segmentation and translation into global languages, the dissemination of the outcomes of language documentation projects is still

heavily skewed in favour of northern researchers' needs.

<PI>Documentary linguistics projects also produce a grammar sketch and a basic dictionary. According to Mosel (2006), the former should provide information about the overall structure of the language, the main building blocks of the language, such as its sounds and its main phonological processes, a discussion of its word classes, functional morphemes, inflectional paradigms, derivational morphemes and morphological processes, and all the word order rules. Himmelmann (2006: 22), among others, argues vehemently against very detailed grammatical descriptions, as they attract too little interest from non-linguistic audiences. However, there is the danger that such materials may turn out to be too basic, producing a reductive representation of the language.

## <H1>10.5 Agents in language documentation<H1>

<P>A crucial aspect of any language project is the people who participate in it. Depending on who initiates, designs, and runs a project, and who obtains and processes the data, very different outcomes will emerge because people pursue different agendas and have different perspectives on what the nature of the project is. The field and documentary linguistics literature does not discuss all of these roles in detail but focuses on some more than others. One role that conspicuously receives little explicit discussion is that of the initiator of projects. Given the criteria for selecting a language for description (see Section 10.2), it is clear that only professional linguists can be in charge of 'selecting' languages, as the categories and typologies are unknown to non-linguists. While communities may express an interest in having their language described, realization depends on the professional linguist's and funder's assessment that a sufficient case for endangerment and uniqueness can be made. Professional (usually Northern) linguists have an important gatekeeping role in this process because they are the people have to assess languages, write proposals, and

evaluate funding applications (Deumert and Storch, Chapter 1 this volume). The professional linguist assures funders that the delivery of the project and its outcomes are compliant with current linguistic reasoning and will produce the necessary impact in terms of publications. This puts the linguists and their knowledge in a very powerful position. In the field-linguistics tradition, there does not appear to be any discussion of this issue beyond the notion of fieldwork ethics; it is accepted that these projects are largely lone-ranger undertakings in which the linguist initiates, designs, and runs the project (Dwyer 2010). The documentary linguistics approach, in contrast, proposes to mitigate this powerful position of the linguist by proposing that projects should be multidisciplinary and that their aims and design should be determined conjointly by all interested parties, including community members, anthropologists, and educators. However, in reality there are still only a rather small number of fully collaborative projects that include a range of stakeholders. This is often for financial reasons and because of the absence of easily available local and international partners.

<PI>The field linguistics literature is quite explicit about the fact that practitioners usually work with one principal speaker or a small number of people to explore the language and generate the samples of language to be analysed and archived (Crowley 2007: 89). This person was traditionally identified as the informant or language helper and is currently increasingly referred to as the language consultant. Either the community assigns a person to the researcher, or the researcher identifies someone during the initial stages of their stay; over time this is often expanded to a small group of people.<sup>7</sup> Initially, the linguist meets regularly with the speaker(s) to do elicitation work. The focus then gradually shifts to the collection and interpretation of spontaneous speech samples. The community member provides the narratives and/or helps the researcher to find

---

<FN><sup>7</sup> Note that service personnel and other people in the community who have frequent contact with outsiders have often figured quite prominently in such roles (McLaughlin and Sall 2001; see also Price and Price (2003) in relation to anthropological research).



suitable people who can produce them. They then aid the researcher in analysing the recorded sample, for instance by assigning glosses, finding translations, and interpreting constructions. While this work does not require specific qualifications, there is nevertheless an expected profile. The language consultant should ideally have a good speaking competence in the language and be considered to be a native or primary speaker of the language,<sup>8</sup> be interested in language, be reliable, be fluent in a global language, have a good grasp of local traditional and popular culture, including language, as well as being quick to understand and engage with the needs of the descriptive linguist and his/her research methods (Crowley 2007: 91). They must also understand that linguists require “exact repetition ... rather than some kind of a paraphrase” (Crowley 2007: 92), i.e. linguists require an exact transcription and translation of recorded texts and texts representing both modern and traditional speech forms. They must be able to reliably distinguish between and be able to produce both conservative and modern forms of speech. Although consultants are expected to be bilingual, the speech that they habitually present to the researcher should not be tainted by their bilingual status. In short, the onus of making the language intelligible to Northern research falls primarily on the chosen speaker of the language, who must not only be competent in both a global language and the language being studied, but also learn to break the latter down according to linguists’ categories. Given this asymmetrical relationship and the fact that this is the linguist’s project rather than a collaboratively conceived one, linguists like to construct the encounter as a service relationship, usually paying the speaker by the hour. As discussed in McLaughlin and Sall (2001), the perception of its being a service relationship is not always shared by the speaker, who might also have other reasons for engaging in it, such as to pay the linguist back for other services, to learn something new, or simply because they like the intellectual challenge. While this kind of service relationship often seems to reduce the speakers of the language to mere information

---

<FN><sup>8</sup> For a short discussion of ways to refer to the chosen speakers of a language, see Crowley (2007: 85–6).

providers, who are supposed to go along with the researcher's 'game' rather than to provide their own insights, over time linguists and consultants often develop a cordial relationship and engage in a mutually rewarding intellectual exchange.

<PI>Interestingly, the speakers of the language receive little specific treatment in the documentary linguistic literature. Since there is an agenda of inclusiveness, all community members are invited to contribute. The burden of interpretation and cultural mediation is to be shared among parties, such as speakers, linguists, and anthropologists. However, looking at projects archived in TLA or ELDP, it appears that few documentation projects were created by such interdisciplinary teams. Speakers are not conceptualized as service providers for the researcher. Instead, they are invited to actively contribute to designing and running the project, and to determining its outcomes. Speakers' actual impacts on these aspects of documentary projects are, however, not clear and the collaborative ethos probably makes them invisible. For Grinevald and Sinha (2016: 33–4) the distinction between language documentation as “value-free scientific activity” and revitalization as “linguistic social work” indicates this asymmetrical relationship and a language as object ideology. Another indicator is the training ethos of such projects to empower communities (Rice 2010) which can also be read as a Northern paternalistic discourse which sees southern communities and their members as powerless, needy, and in need of betterment (Ziai 2013). Conceptualizing research as a ‘betterment strategy’ may also conceal a power hierarchy where speakers are positioned in a subordinate place—as learners—and receive instruction from the superior—the teacher—whose instructions they have to follow.<sup>9</sup> It also implies that community members will not be remunerated for their, often substantial, efforts. Note also that although empowering research is defined as “research on, for and with” the community (Rice 2010: 27), there is a fair amount of discussion in the literature about what community members, communities, and other stakeholders can gain from

---

<FN><sup>9</sup> Unless a collaborative understanding of teaching and learning is assumed.

the process, and what the ethical obligations of the researcher are, suggesting that the participants in a documentation project are, in fact, not yet conceived of as equals with common goals and rights (Deumert and Storch, Chapter 1 this volume).

<PI>The literature on both field and documentary linguistics says very little about who carries out the analysis of the data, including preparing documents for publication. In many cases, the initial part of this process is carried out by researchers with help from speakers of the language in the location where the language is spoken. However, the final preparation of files and texts is typically done in the linguists' workplace in a European or North American city, together with student research assistants.

## <H1>10.6 Conclusions<H1>

<P>In this chapter I argue that research on endangered languages remains preoccupied with the amassing of data to feed Northern scientific activities, and that longstanding asymmetries between speakers' and researchers' interests have not been resolved. Despite critical voices and changes in (mainstream) linguistic research methods towards a greater acceptance and inclusion of hybridity, these changes are only partially reflected in documentary and particularly field linguistics research. Concerns with notions of objectivity, representativeness, replicability, and effective cataloguing and academic theorizing of language still dominate research practice. There is growing awareness of the need for greater community participation in the analysis of textual rather than decontextualized data, and for the investigation of (multilingual) communicative practices and meaning-making, including how these relate to the wider sociocultural context, and how the latter relates to endangerment and marginalization. However, these concerns are not well reflected in training manuals and are only slowly being integrated into actual projects. The products of research and

access to them also remain heavily skewed towards the Northern researcher community, who remain the main gatekeepers and decision-makers in research on endangered languages. In short, research on endangered languages still does not serve to address the current status quo, but is largely concerned with feeding the Northern scientific endeavour.

<PI>Comparing descriptive linguistic and documentary linguistic approaches, it is clear that both have changed and that the latter is attempting to address some of the shortcomings of the former. This concern has not translated into a fundamental transformation, or a paradigm change, in research on endangered languages. The impression is that a larger vision has not yet emerged, or perhaps that researchers are still reluctant to venture into a new reality and are therefore applying patches to address the most problematic aspects. It seems to me that real transformation is dependent on a critical assessment of the 'big issues' and aims of linguistics as a whole. For instance, what constitutes language and can it be studied in a non-bounded and objectivizing manner? Do we need a concept of language? What are the aims and the purposes of linguistic research? Can theoretical, empirical, and applied approaches be integrated into a common endeavour? Are there non-extractive research methods?