

John Benjamins Publishing Company



This is a contribution from *Endangered Languages and Languages in Danger. Issues of documentation, policy, and language rights*.

Edited by Luna Filipović and Martin Pütz.

© 2016. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Language documentation 20 years on

Peter K. Austin

School of Oriental and African Studies (SOAS), University of London

In the last decade of the 20th century a new field of language research emerged that has come to be known as ‘language documentation’ or ‘documentary linguistics’ (Himmelman 1998, 2002, 2006; Lehmann 2001; Austin 2010; Grenoble 2010; Woodbury 2003, 2011). In this paper we explore how it was defined in the seminal work of Himmelman (1998) and others, including what were presented as significant characteristics that distinguished language documentation from language description, and how the field has changed and evolved over the past 20 years. A focus on best practices, standards, tools and models for documentary corpora appeared in the early years, which led later to more critical discussions of the goals and methods of language documentation. The paper examines some current developments, including new approaches to language archiving, and suggests that there are opportunities for language documentation to adopt a more socially-engaged approach to languages and linguistic research, including better engagement with language revitalisation. There are also opportunities to work towards addressing what is currently a language documentation output gap through experimentation with new genres and innovations in writing and publication.

Keywords: archiving, documentary corpora, documentary linguistics, language archiving, language documentation, language revitalisation

1. Defining language documentation¹

Language documentation (also known by the term ‘documentary linguistics’) aims, according to the seminal definition in Himmelman (1998:161), ‘to provide a comprehensive record of the linguistic practices characteristic of a given

1. This is a revised and extended version of Austin 2014 (published in the student publication *JournalLIPP*). For detailed comments on an earlier draft I am grateful to Christine Beier, Aaron Broadwell, Shobhana L. Chelliah, Lise Dobrin, Lauren Gawne, Anthony Grant, Lenore Grenoble, Guillaume Jacques, Friederike Luepke, Waruno Mahdi, David Nathan, Willem de

speech community... This... differs fundamentally from... language description [which] aims at the record of a language... as a system of abstract elements, constructions, and rules.' Himmelmann (2006) presents it as the subfield of linguistics that is 'concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties' (Himmelmann 2006:v). Language documentation is by its nature multi-disciplinary, and as Woodbury (2011) notes, it is not restricted to theory and methods from linguistics but draws on 'concepts and techniques from linguistics, ethnography, psychology, computer science, recording arts, and more' (see Harrison 2005; Coelho 2005; Eisenbeiss 2005 for arguments).

Documentary linguistics has developed over the past 20 years as a response to the growing realisation among linguists, dating from the late 1980s,² that a majority of the world's 7,000 languages are endangered, in the sense that they are being spoken by decreasing and aging populations in reducing numbers of domains and are not being passed on to the next generation of speakers (Robins & Uhlenbeck 1991; Hale et al. 1992; Crystal 2000; Austin 2007; Whalen 2004; Grenoble 2011). A desire among some researchers to create a lasting, and potentially unrepeatably, record of language use in its social and cultural context was one of the driving forces behind the interest in this new approach. This involved a renewed attention to context, influenced by the ethnography of communication (pioneered by Hymes 1964), and the discourse-based approach of Sherzer (1987).

There was also a concern from the beginning of language documentation for supporting speakers and communities who wish to retrieve, revitalise or maintain their languages by providing documentation corpora that could be connected to revitalization work (but see Section 5 below). Also playing a role were advances in information, media, communication and archiving technologies (see Nathan 2010a, 2010b and Section 4) which made possible the collection, analysis, preservation and dissemination of documentary corpora in ways which were not feasible previously. Language documentation also paid attention to the rights and needs of language speakers and community members, and encouraged collaborative

Reuse, Julia Sallabank, Norval Smith, Mauro Tosco, Anthony Woodbury, Joshua Wilbur and an anonymous reviewer; I alone am responsible for any errors.

2. Himmelmann (2008:339) argues that the 'trigger ... was a short presentation by Johannes Bechert ... at the fourteenth International Congress of Linguists in East Berlin in 1987. ... [and] a motion drafted by Christian Lehmann, which was presented to the business meeting of the Comité International Permanent des Linguistes (CIPL) ... [urging] the committee to take action with the goal of bringing the issue of language endangerment to the attention of professional linguists and the general public'. Also important, especially in North America, was Hale et al. (1992).

approaches that would include their direct involvement in the documentation and support of their own languages (see Grinevald 2003; Austin 2010; Yamada 2007).

A concurrent and supporting development was the availability of extensive new funding resources for research on endangered languages from several sources, and the requirements of these funders to adopt a documentary perspective and to archive the recorded data and analyses. The new funders included the Endangered Languages Documentation Programme (ELDP)³ at SOAS (established in 2002 by Arcadia Fund, it has now provided around 350 documentation grants), the Volkswagen Foundation DoBeS⁴ project (which ran from 2001 to 2014 and funded 80 projects), and the Documenting Endangered Languages (DEL)⁵ inter-agency programme of the National Science Foundation and the National Endowment for the Humanities (established 2005, it has funded 320 projects to date). Other smaller sources also emerged (the Endangered Language Fund (ELF),⁶ Foundation for Endangered Languages (FEL),⁷ Gesellschaft für bedrohte Sprachen (GBS)⁸ and Unesco⁹) and have made more modest grants supporting scores of projects, many of which are community-based. This new funding influenced the topics that linguists (and others) chose to research, and the research methods they employed (see Sections 2 and 4 below).

The broader impact on the field of linguistics can be seen in the development of:

- academic journals specialising in language documentation topics (*Language Documentation and Conservation*,¹⁰ *Language Documentation and Description*¹¹), and special issues of other linguistics journals dedicated to documentation and revitalisation (e.g. Volume 34/4 (2013) of the *Journal of Multilingual and Multicultural Development*¹²);

3. <<http://www.hrelp.org/grants>> (11 March 2015).

4. <<http://dobes.mpi.nl/dobesprogramme>> (12 March 2015).

5. <http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12816> (27 March 2015).

6. <<http://www.endangeredlanguagefund.org/>> (27 March 2015).

7. <<http://www.ogmios.org/index.php>> (27 March 2015).

8. <<http://www.uni-koeln.de/gbs/>> (27 March 2015).

9. <<http://www.unesco.org/new/en/culture/themes/endangered-languages>> (27 March 2015).

10. <<http://nflrc.hawaii.edu/ldc/>> (14 March 2015).

11. <<http://www.elpublishing.org>> (10 March 2015).

12. <<http://www.tandfonline.com/toc/rmmm20/34/4#.VRTkmvmUeSo>> (27 March 2015).

- specialist conferences, such as the International Conference on Language Documentation and Conservation held biennially in Hawaii¹³ and the Language Documentation and Linguistic Theory (LDLT)¹⁴ conference held biennially since 2007 at SOAS;
- workshops and training courses, including the Summer Institutes of CoLang/InField¹⁵ run biennially in the United States since 2008, summer schools of the 3L consortium (Leiden-London-Lyon) that also commenced in 2008, and the DocLing¹⁶ training course held annually at Tokyo University of Foreign Studies since 2008;
- specialist MA and PhD programmes at SOAS¹⁷ (Austin 2008), University of Hawaii,¹⁸ and the increasing introduction of documentation topics in undergraduate and postgraduate Linguistics programmes elsewhere;
- a growing number of book publications on topics related to language documentation (for an annotated bibliography see Austin 2013);
- increased attention among linguists with a range of interests, objectives and theoretical persuasions to issues of data quality, portability, data citation, glossing standardization, and data sources (including elicitation, translation, story boarding, naturalistic observations, and experimentation).

Himmelman (2006:15) identified five major characteristics of language documentation that he proposed would distinguish it from other approaches to the study of human languages:

- *focus on primary data* – language documentation concerns the collection and analysis of an array of primary language data to be made available for a wide range of users (further elaborated in Himmelman 2012);
- *explicit concern for accountability* – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected;
- *concern for long-term storage and preservation of primary data* – language documentation includes a focus on archiving in order to ensure that documentary materials are made available to potential users now and into the distant future;

13. <<http://icldc-hawaii.org/>> (27 March 2015).

14. <<http://www.hrelp.org/events/>> (27 March 2015).

15. <<http://www.alaska.edu/colang2016/charter/>> (27 March 2015).

16. <<http://www.aa.tufs.ac.jp/en/training/fieldling-ws/docling>> (27 March 2015).

17. <<https://www.soas.ac.uk/linguistics/programmes/malangdocdesc/>> (27 March 2015).

18. <<http://ling.hawaii.edu/>> (27 March 2015).

- *work in interdisciplinary teams* – documentation requires input and expertise from a range of disciplines and is not restricted to linguistics alone;
- *close cooperation with and direct involvement of the speech community* – language documentation requires active and collaborative work with community members both as producers of language materials and as co-researchers.¹⁹

The application of these principles results, according to Himmelmann (1998, 2002, 2006), in the creation of a record of the linguistic practices and traditions of a speech community together with information about speakers' metalinguistic knowledge of those practices and traditions. This is achieved by systematic recording, transcription, translation and analysis of a variety of spoken (and written) language samples collected within their appropriate social and cultural context. Analysis within language documentation under this view is aimed at making the records accessible to a broad range of potential users which includes not only linguists but also researchers in other disciplines, community members and others, who may not have first-hand knowledge of the documented language. The record is also intended for posterity (and hence should be preservable and portable, in the sense of Bird & Simons 2003), and so some level of processing is required. There is a need for systematic recording of metadata (data about the data) to make the archived materials understandable, findable, preservable and usable.

The core of a language documentation defined in this way was generally understood to be a corpus of audio and/or video materials with time-aligned transcription, annotation, and translation into a language of wider communication (Schulze-Berndt 2006), and relevant metadata on context and use of the materials. Woodbury (2003) argued that the corpus will ideally cover a diverse range of genres and contexts, and be large, expandable, opportunistic, portable, transparent, ethical and preservable. Austin (2006a, 2008, 2010) proposes that there are five activities (not necessarily sequential) which are identifiable in this documentation approach and which contribute to corpus creation, analysis, preservation and dissemination:

- *recording* – of media and text (including metadata) in context;
- *transfer* – to a data management environment;
- *adding value* – the transcription, translation, annotation and notation and linking of metadata to the recordings;

19. Issues concerning communities, collaboration and ethics of research have been an ongoing thread in papers published in the journal *Language Documentation and Conservation* over a number of years.

- *archiving* – creating archival objects and assigning them access and usage rights;
- *mobilisation* – creation, publication and distribution of outputs, in a range of formats for a range of different users and uses.

2. Best practices, tools and models

The establishment of the DoBeS project in 2001 saw the emergence of a unified ‘DoBeS model’ for language documentation that the funded projects were expected to adopt.²⁰ This included specifications for archival storage, recommendations about recording and analysis formats, and the development of new software tools to assist with audio and video annotation (such as ELAN²¹), and the creation and management of metadata (various IMDI tools²²). Researchers affiliated with DoBeS also proposed general principles (or ‘best practice’) for language documentation, such as sampling (to meet Himmelmann’s desideratum that the documentary record should be ‘representative’, see Seifart 2008), data collection methods (Lüpke 2009) and a typology of data types (Himmelmann 2012).

Definition of best practice, standards, tools and models was also a central goal of the E-MELD project²³ funded by the National Science Foundation which ran from 2001 to 2006 aiming to develop recommendations for metadata, annotation markup, language identification and linguistic ontology (essentially the sets of labels employed in interlinear glossing). This resulted in a series of papers²⁴ defining formats for lexical entries (Bell & Bird 2000), interlinear text (Bird & Liberman 2001; Bowe et al. 2003), paradigms (Penton et al. 2004) and a generalised ontology for glossing (Farrar et al. 2002; Farrar & Langendoen 2003a, b). E-MELD set up a ‘School of Best Practices’ (Aristar 2003; Aristar-Dry 2004)²⁵ with case studies, a reference list of readings and tools, and a classroom ‘designed to offer “lessons” and tutorials which explain the recommendations of best practices’.

20. <<http://dobes.mpi.nl/dobesprogramme> and <http://www.mpi.nl/corpus/a4guides/a4-guide-dobes-format-encoding.pdf>> (10 March 2015).

21. <<https://tla.mpi.nl/tools/tla-tools/elan/>> (10 March 2015).

22. <<http://www.mpi.nl/IMDI/>> (10 March 2015).

23. <<http://emeld.org/>> (10 March 2015).

24. <<http://emeld.org/documents/index.cfm#loc-papers>> (10 March 2015).

25. <<http://emeld.org/school/index.html>> (10 March 2015).

Probably the most ambitious attempt to define best practice and what would constitute a complete documentation of a language is to be found in CELP 2007, which attempted to define everything that an adequate documentation should cover: all the basic phonology, morphology, syntactic constructions (in context), and provide a lexicon covering all the basic vocabulary and important areas of special expertise in the culture, with at least glosses for all words/morphemes in the corpus, plus a full range of textual genres and registers. It offered a set of 'accounting standards' to determine adequacy, including quantitative measures such as a figure of 10,000 items for a lexicon, and a text corpus of one million words (around 1200 hours of recorded speech). Other qualitative measures were suggested such as the notion that research on an endangered language is completed 'when nothing new is coming up in non-elicited material and when any apparent lacunae in the phonological system can be shown to be real and not an accident of data collection.'

It is doubtful if linguists would ever suggest it is possible to qualitatively and quantitatively determine when a research project is 'complete' for non-endangered languages, yet this is precisely what was suggested for language documentation.

Both DoBeS and E-MELD were influential in encouraging linguists to begin to pay attention to data types, data structures, analytical processes and workflows, together with preservability and transparency, however the notion that there was a 'documentation model' or a 'best practice' (or a small number of 'best practices') was questioned by some researchers, beginning around 2004.

3. Critical responses

The role of archives in defining the goals and values of language documentation was challenged by Nathan (2004) who introduced the term 'archivism' to describe the idea that quantifiable properties such as recording hours, data volume, file parameters, and technical desiderata like 'archival quality' and 'portability' could be reference points in assessing the aims and outcomes of language documentation. He argued that these should not be measures of quality of a documentation project, and that there had been a lack of discussion of research methodology among language documenters, including about what such quality measures might be.

Nathan and Austin (2004) addressed the issue of metadata and argued that all value-adding that researchers provide for the audio or video records they make should be understood as metadata, and that it should be as rich as possible and designed for the documentation purpose at hand. This means that metadata should not be constrained by specifications in the form of an 'ontology' or standard

minimal set (such as that proposed by OLAC²⁶). The need for richer metadata and meta-documentation (documentation of the language documentation) was further elaborated on by Austin (2009, 2013) – see also Gawne et al. (2015).

Two important issues for the definition of language documentation were raised in 2006, namely the difference between documentation and description, which was considered fundamental in Himmelmann's seminal paper (see quotation in 1 above), and the approach to audio recording within documentation. Austin (2006b), revised and published as Austin and Grenoble (2007), noted that, as Himmelmann (1998) made clear, language documentation and description differ in terms of their goals, areas of interest, research methods, workflows, and outcomes. Language description focusses on languages as sets of structures and systems, and typically aims to produce grammars, dictionaries, and collections of texts, the intended audience of which is usually linguistics specialists. By contrast, documentation is discourse-centered: its primary goal is the representation of a range of instances and types of language use in their social and cultural context. Although description may draw on a corpus, it involves analysis of a different order, oriented to providing an understanding of language at a more abstract level, as a system of elements, rules, and constructions. Austin and Grenoble (2007: 22) challenged this sharp separation of description and documentation and argued that:

[d]ocumentation projects must rely on the application of theoretical and descriptive linguistic techniques in order to ensure that they are usable (i.e. have accessible entry points via transcription, translation and annotation), as well as to ensure that they are comprehensive. It is only through linguistic analysis that we can discover that some crucial speech genre, lexical form, grammatical paradigm or sentence construction is missing or under-represented in the documentary record. Without good analysis, recorded audio and video materials do not serve as data for any community of potential users.

In terms of workflow, they also differ. For description, linguistic knowledge and decision-making is applied to some event in the real world to make an inscription (e.g. an audio recording) that is not itself of interest but serves as a source which can then be selected, analysed and systematised in order to create analytical representations, typically in the form of lists, summaries and analyses (e.g. statements about phonology, morphology or syntax). It is these representations which are the main focus of interest and which are then presented and distributed to users, typically other linguists. For documentation, linguistic and cultural knowledge and documentary techniques are applied to some event in the real world to make

26. <<http://www.language-archives.org/OLAC/olacms.html>> (10 March 2015).

an inscription (audio or video recording) that recapitulates aspects of the original event (such as social or spatial relationships – see Nathan 2010a) and is itself a focus of interest (e.g. for archiving, preservation and distribution). The documentary researcher adds value to the inscription by making decisions and applying linguistic and other knowledge to create representations, typically in the form of transcriptions, translations and annotations. These representations are the second major focus of interest and will be archived and/or mobilized and distributed. The same representations could, of course, also be the input to the selection and analytical procedures of description, thereby linking the descriptive outcomes to the documentary corpus. From this viewpoint, documentation and description are complementary activities with complementary goals, methods and outcomes.

Nathan (2006) argued that despite the expressed concern by language documenters for recording language in its social and cultural context, many researchers took an unscientific approach to audio recording in particular, ignoring issues such as spatiality and microphone selection and placement. He extended this critique in Nathan (2009, 2010a) and argued for the need to establish an epistemology for audio recording within language documentation.

A broader critique of documentation and contemporary endangered languages research can be found in Dobrin et al. (2007) who identify and highlight tendencies towards objectification of languages, and reliance on familiar qualitative metrics to measure quality, progress and value. More specifically, they argue that ‘subtle and pervasive kinds of commoditisation (reduction of languages to common exchange values) abound, particularly in competitive and programmatic contexts such as grant-seeking and standard-setting where languages are necessarily compared and ranked’. Bovern (2011:468) also points to commoditisation and suggests that ‘community members report sometimes feeling that the linguist comes in, reifies the language, turns it into a commodity, and then takes it away.’

Dobrin et al. (2007) echo Nathan (2004) in pointing to archivism as problematic, and join Nathan (2006) in arguing that documentary linguists show little or no knowledge about recording arts, including microphone types, properties and placement, even though microphone choice and handling is the single greatest determiner of audio recording quality. They also note that evidence from archival deposits shows that video tends to be poorly used by documentary linguists, with video recordings being made without reference to articulated hypotheses, goals, or methodology, simply because the technology is available, portable and relatively inexpensive. Finally, in contrast to earlier approaches, they point to diversity as an important aspect of language documentation. As researchers respond to the unique and particular social, cultural and linguistic contexts within which the languages they are studying are spoken or signed, actual documentation projects, as evidenced by grant project proposals and materials deposited in archives,

show a diversity of approaches, techniques, methodologies, skills and responses. In the last 10 years we also find an increasing diversity of materials that can be included in corpora, so that alongside the traditional field interviews, observations, experiments and narrative collections that have been the bread and butter of documentation and description, we also find materials, much of them created by native speakers, from YouTube uploads, Twitter feeds, Facebook posts, blogs, email, chat, Skype calls, and local pedagogy developed for revitalization. Similarly, the outcomes of documentation are increasingly diverse so that alongside books, papers and archive deposits, today research projects are also generating YouTube uploads,²⁷ Twitter and Facebook posts, blogs,²⁸ multimedia (such as *Gayarragi Winangali*,²⁹ and mobile apps (such as *Ma! Iwaidja*)³⁰). Rather than aiming for comprehensiveness or representativeness, research funded recently by ELDP for example, rather shows specificity, focussing on topics such as traditional song in its diaspora context, language use by blacksmiths, bark cloth making, libation rituals, fishing practices, child language, interactive speech, and ethnobotany (projects funded in 2012 and 2014).³¹

In a recent handbook, Woodbury (2011: 159) presents a definition of language documentation which reflects this shift away from representative samples towards more specific goals as ‘the creation, annotation, preservation and dissemination of transparent records of a language’. He also identifies some gaps in the earlier conceptions of documentation, especially because ‘language encompasses conscious and unconscious knowledge, ideation and cognitive ability, as well as overt social behaviour’ (ibid.). The role of ideologies of language structure and use, attitudes of speakers to their and others’ speech, and the relationships of beliefs and attitudes to actual performance in the world are only beginning to be addressed by documentary linguists (see Austin & Sallabank 2014). As Woodbury (2011: 160) notes, ‘humans experience their own and other people’s languages

27. For example, Anthony Jukes’ subtitled video on Minahasan food and cooking methods at <<https://www.youtube.com/watch?v=wVy2QsFqdYI>> (9 June 2015); see also <<https://www.youtube.com/watch?v=hqNQ-z9sIBw>> for further details.

28. For example, Austin’s Dieri blog at <<http://www.dieriyawarra.wordpress.com>> (15 March 2015).

29. <<http://www.dnathan.com/projects/gw/>> (27 March 2015).

30. <<http://www.iwaidja.org>> (15 March 2015).

31. <<http://www.hrelp.org/grants/projects/index.php?year=2012>> <<http://www.hrelp.org/grants/projects/index.php?year=2014>> (14 March 2015). Note that there has not been a complete shift away from the ‘whole language documentation’ approach with quite a number of funded projects still taking such an approach.

viscerally and have differing stakes, purposes, goals and aspirations for language records and language documentation.

Woodbury (2011) has also highlighted a need to develop a theory of documentary corpora (covering the principles by which a particular corpus ‘hangs together’), as well as a need for accounts of individual documentation project designs. Austin (2013) extends this to a general call for reflexive meta-documentation of their work by researchers concerning their documentary models, processes and practices. This would include: the identity of stakeholders and their roles; the attitudes and ideologies of language consultants and the communities within which they are located (towards their languages as well as the documenter and documentation project³²); the relationships between researchers, research project participants and the wider community; the goals and methodology adopted by the project, including research methods and tools (see Lüpke 2009); corpus theorization (Woodbury 2011); theoretical assumptions embedded in annotation and translation (e.g. in abbreviations, glosses); and considerations of the potential for a project to contribute to revitalization. In addition, it is important to know the biography of the project, including background knowledge and experience of the researcher and main consultants (e.g. how much fieldwork the researcher had done at the beginning of the project and under what conditions, what training the researcher and consultants had received). Austin (2013) suggests that such meta-documentation can draw upon knowledge and practices in other disciplines (such as social and cultural anthropology, archaeology, archiving and museum studies), and from considerations that surface in the interpretation of past documentations (of legacy materials). The many parallels between language documentation and ethnomusicology in terms of these and other topics are explored in detail by Grant (2014).

Austin and Sallabank (2015) point out that the early emphasis on ‘compiling a representative and lasting multipurpose record of a language’ has led documenters to focus on defining and describing individual languages in isolation with a narrow attention to what Woodbury (2011: 177) calls ‘the ancestral code’, rather than documenting dynamic language practices and real-life interactions in their sociolinguistic context (see also Sugita 2007; Amery 2009; Childs et al. 2014). By definition, endangered languages do not exist in isolation but are always spoken in relationships with other languages, varieties, codes, styles, registers, etc., in a complex linguistic ecology (Haugen 1972; Mühlhäusler 1992, 2000; Calvet 2006). Grenoble (2011) has argued that linguists should aim to document language ecologies, not just what they define as individual languages or varieties (the ancestral

32. See Kroskrity (2015) for an example relating to a dictionary project.

code approach). At the very least they should pay attention to multilingual repertoires, mixed codes, the sociolinguistic and structural effects of contact, and language variation and change (Lüpke & Storch 2013). Gullberg (2012) has explored the interplay between multilingualism and multimodality, arguing that ‘language documentation data has the potential to inform theoretical and empirical studies of linguistics, bilingualism and multimodality in entirely new ways, and, conversely, that documentation work would benefit from taking the bilingual and multimodal nature of its data into account’ (Gullberg 2012: 46).

It is also important to consider extra-linguistic factors such as language attitudes and ideologies (Sallabank 2013; Austin & Sallabank 2014). The dominant model of language documentation from 1995 to 2010 could be described as ‘saving the morphemes two-by-two’ in a ‘Noah’s arc(hive)’, salvage-linguistics approach which reflects a purist notion of single languages in isolation. From 2010, for at least some language documenters, the approach has become more particular, dynamic, pluralistic and socially-engaged.

4. Developments in archiving

The rise of language documentation has also seen the development of a number of internet-accessible digital archives focusing in particular on the preservation of materials on endangered languages. These include DoBes in the Netherlands,³³ Paradisec in Australia,³⁴ Pangloss in Paris,³⁵ the California Archive in Berkeley,³⁶ AILLA in Texas,³⁷ and ANLA in Alaska.³⁸

One of the most dramatic developments of the 21st century has been the rise of social network models on the internet (so-called Web 2.0) that aim to link people rather than documents, with a focus on interaction and collaboration instead of passive downloading and viewing of content. These new models have been taken up in the last 10 years by some language documentation archives (such as ELAR³⁹ at SOAS) leading to what Nathan (2010b) calls ‘Archives 2.0’.

33. <<http://dobes.mpi.nl/>> (9 June 2015).

34. <<http://paradisec.org.au/>> (9 June 2015).

35. <http://lacito.vjf.cnrs.fr/pangloss/presentation_en.htm> (9 June 2015).

36. <<http://cla.berkeley.edu/>> (9 June 2015).

37. <<http://www.ailla.utexas.org/site/welcome.html>> (9 June 2015).

38. <<https://www.uaf.edu/anla/>> (9 June 2015).

39. <<http://www.elar-archive.org>> (10 March 2015).

Traditionally, archiving has focused heavily on preservation (and on cataloguing and standards – see Section 3 above), however language documentation raises a number of new methodological challenges, especially in relation to endangered languages where speakers ‘tend to use their language more and more to speak of private, local, sensitive and secret matters. So the primary data of documentary linguistics maximises the likelihood of including content that can cause discomfort or harm to the recorded speakers’ (Nathan 2014: 191) or their families and descendants. Thus documentation corpora often contain ritual or sacred material that may be restricted in terms of who can be exposed to them, as well as gossip which may contain references to private knowledge or events. As a result, language documentation archives need powerful but flexible access management that is transparent, easy to understand, and able to be changed as circumstances develop. The basis for access will be via relationships between the providers of the materials (archive depositors and the stakeholders they work with) and those who wish to use them. Beginning in 2005, the ELAR archive at SOAS developed a richly articulated system of ‘access protocols’ designed to formulate and implement speakers’ rights and sensitivities, together with rigorous methods and processes for controlled access to the archival materials. Each resource is assigned one of five levels of access: U (open to all registered users), R (for registered researchers only), C (for community members only) and S (for subscribers who negotiate access with the depositor), X (closed to all but the depositor). Registered users are then categorized by archive staff and their access to particular materials depends on their status (e.g. they are R by virtue of being associated with an academic programme, and/or C because they explain that they have links to a particular community⁴⁰) and the access type of the materials they wish to use. A similar access protocol system is in use by TLA, *The Language Archive*, at the Max Planck Institute for Psycholinguistics (which includes the DoBeS endangered languages archive).⁴¹ Endangered language archiving thus requires a special response to the well-publicised movement for complete open access that is current in much other academic research and publication.

In this view, such an archive can also be seen as a place for establishing and transacting relationships and sharing, and Web 2.0 models provide a technology for instantiating this. The general model of the ELAR archive is presented by Nathan (2010b) as in Figure 1.

There are several other archiving developments that have been pioneered by ELAR in the last 10 years. The first, called *progressive archiving*, sees archiving

40. This can be one of the most difficult and complex statuses for an archive to determine.

41. <http://dobes.mpi.nl/access_registration/> (10 March 2015).

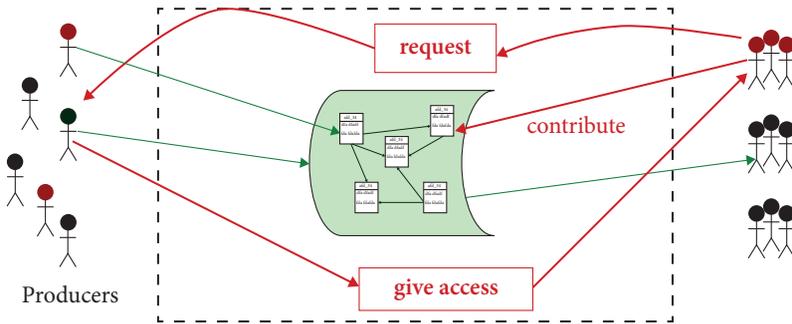


Figure 1. ELAR Archive 2.0 model

as a whole-of-project relationship: depositor accounts are established at the beginning of a research project, and researchers add and manage or update their materials over time, as well as managing and engaging in interactions with the curators and users. Secondly, ELAR have developed a web accessible archive interface that has been designed to provide contextualization, different degrees of presentation for different projects, and ease of navigation for users. The interface directly reflects the interests and needs of the materials providers and the users, rather than being, for example, a unified tree structure across the whole collection, as other archives such as DoBeS and AILLA. Thirdly, ELAR has promoted increased participation so that users can negotiate access to particular materials and bookmark their favourites, while depositors can negotiate access requests and monitor usage. A communication channel has also been established in order for both groups to exchange and share information. Nathan (2014) gives examples of these exchanges and how they can lead to creative outcomes and collaborations between researchers and members of the community of users.

Possible future developments in endangered languages archiving may include community curation of archived materials (Linn 2014), participant identification and expression of rights (Garrett 2014), and the creation of new kinds of outputs that draw upon a range of materials drawn from several collections within the archive (just as museums and galleries choose, select and exhibit their resources for educational or other purposes – see Holton 2014). The overall flavour of archiving in the last five years has changed from finality and completeness to being open and evolutionary. These developments also raise questions for archives about what a ‘deposit’ or ‘depositor’ really is, and recast archives as providers of services within a revised, holistic concept of language documentation.

5. Language documentation and revitalization

The term ‘language revitalization’ is used to describe principles and activities aimed at increasing the number of users of a language, and/or the range of domains within which it is used (Fishman 1991, 2001; Hinton & Hale 2001; Hornberger 2010; Hinton 2011; Romaine 2007; Grenoble & Whaley 2006). It has been in operation for more than 20 years longer than language documentation as its origins go back to community-based activities by Māori in New Zealand in the 1970s (Spolsky 1989, 2003; Bentahila & Davies 1993) and by other groups such as North American indigenous people (Niedzielski 1992; Kaponi 1995; Hinton 1993, 2002, 2013), and European minorities such as the Catalan, Welsh and Basque.

The relationship between language documentation and language revitalization is a rather complex one, and is explored in some detail in Austin and Sallabank (2015). For many language documenters revitalization has been seen as a waste of resources, a viewpoint connected to the ‘language-as-system’ ideology that sees linguistic data as the only thing worth collecting and preserving, in contrast to ‘linguistic social work’ (Newman 2003:6; see also Dimmendaal 2004:84 and Blench 2008:153).⁴²

Although documentation defined itself from the beginning as a field that set out to create a multipurpose record for a wide range of users, including community members, language revitalization has been treated as a simple ‘technical add-on’ that involves creation of orthographies, dictionaries, sub-titled videos, and primers and multimedia, including websites, rather than as a field of research or activity that requires theoretical and applied knowledge. This view was also strongly supported by the funding agencies (including ELDP, Volkswagen Stiftung, NSF-NEH), who excluded revitalization-oriented projects from grants and severely limited the amount of money that could be included for revitalisation materials creation or ‘community publication’ of research results.

Much of the material that has ended up in language documentation archives is unsuitable for revitalization for a variety of reasons, including inappropriate genres or topics, recordings and analyses in difficult to access archival formats that require specialised software (such as ELAN or FLE_x), or glossed and translated into languages such as English that have little or no place in the local linguistic ecology. Documentation is also heavily biased towards the performances of older fluent speakers, resulting in language that may be too fast, heavily context dependent and include slurring or elisions, or even be affected by physiological factors (not least of which may be lack of teeth). Few, if any, documentary

42. Newman’s views were repeated and further elaborated in Newman (2013); see also the response by Whalen (2013).

corpora include samples of children's ordinary language use⁴³ or learner-directed speech; in addition, as noted by Cope (2014), documentary linguists are not trained in pedagogical materials design, and applied linguists are rarely included in language documentation teams. The relationship between language documentation and revitalization has thus varied from avoidance or subordination to, at best, only an indirect connection (Sallabank 2012). There is a need for much more exploration and development of this area in the future (see also Austin & Sallabank 2015 for further discussion).

6. Documentation and academia

The development of language documentation as a field with its own principles and practices appeared to many researchers in its foundation period at the end of the 20th century to offer an opportunity to change the socio-political academic balance between fieldworkers and so-called 'armchair linguists' (typologists, theoreticians) (Fillmore 1992; Aikhenvald 2007:4; Crowley 2007:11–13) by providing a foundation (theory, best practices) for corpus creation, data collection and analysis. Many perceived that fieldwork and language description were in a subordinate sociological position. Newman (2009:124)⁴⁴ states explicitly that 'theoreticians belittle descriptivists as linguistically second-class citizens'⁴⁵, and hoped that language documentation and the work of corpus creation and associated activities would raise their status in academic linguistics. Indeed, lobbying by documenters and others led in 2010 to the Linguistic Society of America 'Resolution Recognizing the Scholarly Merit of Language Documentation' which states that:

[a] shift in practice has broadened the range of scholarly work to include not only grammars, dictionaries, and text collections, but also archives of primary data, electronic databases, corpora, critical editions of legacy materials, pedagogical works designed for the use of speech communities, software, websites, or other

43. An exception is the DoBeS Chintang/Puma project – see <http://dobes.mpi.nl/projects/chintang/>, accessed 9 June 2015. There is incidental children's language material in the ELAR archive, such as children's retellings of the Frog Story book (Mayer 1969), however this material has not been systematically collected.

44. Originally published in 1992.

45. Newman (2009:124) considers this to be an 'unintended consequence of Chomsky's (1964) hierarchy of levels of adequacy in grammar, namely, from the bottom up, observational adequacy – "A grammar that aims for observational adequacy is concerned merely to give an account of the primary data" (p. 63, italics mine) –, descriptive adequacy, and explanatory adequacy.'

digital media. The products of language documentation and work supporting linguistic vitality are of significant importance to the preservation of linguistic diversity, are fundamental and permanent contributions to the foundation of linguistics, and are intellectual achievements which require sophisticated analytical skills, deep theoretical knowledge, and broad linguistic expertise.

The resolution ‘support[ed] the recognition of these materials as scholarly contributions to be given weight in the awarding of advanced degrees and in decisions on hiring, tenure, and promotion of faculty’. In addition, the resolution encouraged ‘the development of appropriate means of review of such works so that their functionality, import, and scope can be assessed relative to other language resources and to more traditional publications’.

To date, criteria for this kind of review of documentary corpora, or examples of such reviews (parallel, say, to book reviews), have not appeared. In the five years since this resolution was passed there still remains what we can call an ‘output gap’: traditional products of language description and typological and theoretical research (grammars, book chapters, journal articles) are understood and accorded value in determining promotion, award of tenure and in decision making about new job appointments, but the newer outputs in the form of digital archival deposits, multimedia products, and pedagogical materials for revitalization are either not valued or discounted.

According to Thieberger (2012) similar discussions have taken place in Australia beginning in 2011 between the Australian Linguistic Society (ALS) and the Australian Research Council (ARC), and ‘although the ARC accepted that curated corpora could legitimately be seen as research output, it would be the responsibility of the ALS (or the scholarly community more generally) to establish conventions to accord scholarly credibility to such products’. He reports on proposals for a possible review procedure but recognizes that ‘the question of what criteria to use in evaluating a corpus is more problematic’. For some suggestions for criteria see Thieberger (2012) and Thieberger et al. (2012); again no action appears to have been taken to date to actually implement these Australian proposals.

In my view, to address this output gap, there is a need for experimentation and the development of new genres, so far unfamiliar to linguists, that link and contextualise analytical outputs and the archival corpus. These could include ethnographies of documentation project designs, accounts of data collection (cf. the genre of research publications in archaeology called ‘field reports’), finding-aids to corpus collections, or ‘exhibitions’ or ‘guided tours’ of archival deposits (along the lines of exhibitions and associated products regularly mounted by museums to display parts of their collections, see also Woodbury 2014). Similarly, reviews of corpora or these new kinds of writing could also be attempted.

There has been a very recent development in Linguistics of free online open access publication platforms (e.g. *Language Science Press*, established in April 2013,⁴⁶ and *EL Publishing*, launched in July 2014),⁴⁷ with all the usual academic requirements such as double-blind reviewing and professional editing, design and layout. While *Language Science Press* publishes digital versions of traditional books, *EL Publishing* has set out to provide and encourage new opportunities for language documenters to publish multimedia and the other innovative types of output mentioned above. It remains to be seen whether these opportunities will be taken up by practitioners, and whether they will go some way to addressing the output gap in the future.

7. Conclusions

The past 20 years has seen the emergence and gradual development of a new field of research called ‘documentary linguistics’ or ‘language documentation’ which has concentrated on recording, analysing, preserving and disseminating records of languages in use in ways that can serve a wide range of constituencies, particularly the language communities themselves. In the early period of its development there was a concentration on defining a model for language documentation and specifying best practices, tools and analytical categories, however the past 10 years have seen a shift in perspective responding to criticism of these early concerns. Today, there is more recognition of diversity of contexts, goals, methods and outcomes of language documentation, and indications of the introduction of social models of research, especially in the area of archiving. Much work remains to be done however, to engage better with language revitalization and to establish reliable and replicable measures for evaluating the quality, significance and value of language documentation research so that its position alongside such sub-fields as descriptive linguistics and theoretical linguistics can be assured and enhanced.

References

- Aikhenvald, A.Y. 2007. Linguistic fieldwork: Setting the scene. *Sprachtypologie und Universalienforschung* 60(1): 3–11.
- Amery, R. 2009. Phoenix or relic? Documentation of languages with revitalization in mind. *Language Documentation and Conservation* 3(2): 138–148.

46. <<http://langsci-press.org/index>> (9 June 2015).

47. <<http://www.elpublishing.org>> (1 March 2015).

- Aristar, A.R. 2003. *The school of best practice*. London: SOAS Workshop on Archives for Endangered Languages, November 21–22.
- Aristar-Dry, H. 2004. E-MELD School of best practices in digital language documentation. Detroit: Presentation at *E-MELD Conference 2004: Workshop on linguistic databases and best practice*, July 15–18.
- Austin, P.K. 2006a. Data and language documentation. In *Essentials of Language Documentation* [Trends in Linguistics. Studies and Monographs 178], J. Gippert, N. Himmelmann & U. Mosel (eds), 87–112. Berlin: Mouton de Gruyter.
- Austin, P.K. 2006b. *Defining language documentation*. Washington DC: Paper presented at the Georgetown University Roundtable on Linguistics, March.
- Austin, P.K. 2007. Survival of languages. In *Survival: Darwin College Lectures*, E.F. Shuckburgh (ed.), 80–98. Cambridge: CUP.
- Austin, P.K. 2008. Training for language documentation: Experiences at the School of Oriental and African Studies. In *Documenting and Revitalising Austronesian Languages* [Language Documentation and Conservation Special Publication 1], M. Florey & V. Rau (eds), 25–41. Hawaii HI: University of Hawaii Press.
- Austin, P.K. 2009. Meta-documentary linguistics. Kioloa: Paper given at the *Aboriginal languages workshop*, March 2009.
- Austin, P.K. 2010. Current issues in language documentation. *Language Documentation and Description* 7, 12–33. London: SOAS.
- Austin, P.K. 2013. Language documentation and meta-documentation. In *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*, S. Ogilvie & M. Jones (eds), 3–15. Cambridge: CUP. doi:10.1017/CBO9781139245890.003
- Austin, P.K. 2014. Language documentation in the 21st century. *JournaLIPP* 3: 57–71.
- Austin, P.K. & Grenoble, L. 2007. Current trends in language documentation. *Language Documentation and Description* 4, 12–25. London: SOAS.
- Austin, P.K. & Sallabank, J. 2014. *Endangered Languages: Ideologies and Beliefs in Language Documentation and Revitalization*. London: British Academy. doi:10.5871/bacad/9780197265765.001.0001
- Austin, P.K. & Sallabank, J. 2015. Language documentation and language revitalization: Partners or just good friends? Ms, SOAS.
- Bell, J. & Bird, S. 2000. Preliminary study of the structure of lexicon entries. Philadelphia: Proceedings of the *Workshop on Web-Based Language Documentation and Description*.
- Bentahila, A. & Davies, E.E. 1993. Language revival: Restoration or transformation? *Journal of Multilingual and Multicultural Development* 14(5): 355–374. doi:10.1080/01434632.1993.9994542
- Bird, S. & Liberman, M. 2001. A formal framework for linguistic annotation. *Speech Communication* 33(1–2): 23–60. doi:10.1016/S0167-6393(00)00068-6
- Bird, S. & Simons, G. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3): 557–582. doi:10.1353/lan.2003.0149
- Blench, R. 2008. Endangered languages in West Africa. In *Language Diversity Endangered*, M. Brenzinger (ed.), 140–162. Berlin: Mouton de Gruyter.
- Bowe, C., Hughes, B. & Bird, S. 2003. Towards a general model for interlinear text. Proceedings of *EMELD-03*, 12 August, 2013. <<http://emeld.org/workshop/2003/bowbadenbird-paper.html>>

- Bowern, C. 2011. Planning a language documentation project. In *The Cambridge Handbook of Endangered Languages*, P.K. Austin & J. Sallabank (eds), 459–482. Cambridge: CUP. doi:10.1017/CBO9780511975981.023
- Calvet, J.-L. 2006. *Towards an Ecology of World Languages*. Cambridge: Polity.
- CELP. 2007. Adequacy of documentation. Document circulated at the *January 2007 meeting of the Linguistic Society of America Committee on Endangered Languages and Their Preservation*.
- Childs, T., Good, J. & Mitchell, A. 2014. Beyond the ancestral code: Towards a model for sociolinguistic language documentation. *Language Documentation and Conservation* 8: 168–191.
- Chomsky, N. 1964. Current issues in linguistic theory. In *The Structure of Language*, J.A. Fodor & J.J. Katz (eds), 50–118. Englewood Cliffs NJ: Prentice Hall.
- Coelho, G. 2005. Language documentation and ecology: Areas of interaction. In *Language Documentation and Description* 3, P.K. Austin (ed.), 63–74. London: SOAS.
- Cope, L. (ed.). 2014. *Applied Linguists Needed: Cross-disciplinary Networking in Endangered Language Contexts*. Abingdon: Routledge.
- Crowley, T. 2007. *Field Linguistics: A Beginners Guide*. Oxford: OUP.
- Crystal, D. 2000. *Language Death*. Cambridge: CUP. doi:10.1017/CBO9781139106856
- Dimmendaal, G. 2004. Capacity building in an African context. In *Language Documentation and Description* 2, P.K. Austin (ed.), 71–89. London: SOAS.
- Dobrin, L., Austin, P.K. & Nathan, D. 2007. Dying to be counted: Commodification of endangered languages in documentary linguistics. *Language Documentation and Description* 6, 37–52. London: SOAS.
- Eisenbeiss, S. 2005. Psycholinguistic contributions to language documentation. *Language Documentation and Description* 3, 106–140. London: SOAS.
- Farrar, S. & Langendoen, D.T. 2003a. A linguistic ontology for the semantic web. *GLOT International* 7(3): 97–100.
- Farrar, S. & Langendoen, D.T. 2003b. Markup and the GOLD ontology. *Proceedings of EMELD-03*, 12 August 2013. <<http://staff.washington.edu/farrar/documents/inproceedings/FarLang03a.pdf>>
- Farrar, S., Lewis, W.D. & Langendoen, D.T. 2002. A common ontology for linguistic concepts. Seattle WA: Proceedings of the *Knowledge Technologies Conference*, 12 August 2013. <<http://staff.washington.edu/farrar/documents/inproceedings/FarLewLang02a.pdf>>
- Fillmore, C. 1992. 'Corpus linguistics' or 'computer-aided armchair linguistics'. In *Directions in Corpus Linguistics*, J. Svartik (ed.), 35–45. Berlin: Mouton de Gruyter.
- Fishman, J.A. 1991. *Reversing Language Shift: Theory and Practice of Assistance to Threatened Languages*. Bristol: Multilingual Matters.
- Fishman, J.A. (ed.). 2001. *Can Threatened Languages Be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective*. Bristol: Multilingual Matters.
- Garrett, E. 2014. Participant driven language archiving. In *Language Documentation and Description* 12, D. Nathan & P.K. Austin (eds), 68–84. London: SOAS.
- Gawne, L., Kelly, B.F., Berez, A. & Heston, T. 2015. Putting practice into words: Fieldwork methodology in grammatical descriptions. Hawaii: Paper presented at *ICLDC 4 Conference*.
- Grant, C. 2014. *Music Endangerment: How Language Maintenance Can Help*. Oxford: OUP. doi:10.1093/acprof:oso/9780199352173.001.0001

- Grenoble, L.A. 2010. Language documentation and field linguistics: The state of the field. In *Language Documentation: Practice and Values*, L.A. Grenoble & N. L. Furbee (eds), 289–309. Amsterdam: John Benjamins. doi:10.1075/z.158.28gre
- Grenoble, L.A. 2011. Language ecology and endangerment. In *The Cambridge Handbook of Endangered Languages*, P.K. Austin & J. Sallabank (eds), 27–45. Cambridge: CUP. doi:10.1017/CBO9780511975981.002
- Grenoble, L.A. & Whaley, L.J. 2006. *Saving Languages: An Introduction to Language Revitalization*. Cambridge: CUP.
- Grinevald, C. 2003. Speakers and documentation of endangered languages. *Language Documentation and Description* 1: 52–72.
- Gullberg, M. 2012. Bilingual multimodality and language documentation. In *Potentials of Language Documentation: Methods, Analyses, and Utilization* [Language Documentation and Conservation Special Publication 3], F. Seifart, G. Haig, N. P. Himmelmann, D. Jung, A. Margetts & P. Trilsbeek (eds), 46–53. Honolulu HI: University of Hawai'i Press.
- Hale, K., Krauss, M., Watahomigie, L.J., Yamamoto, A.Y., Craig, C., Jeanne, L.M. & England, N.C. 1992. Endangered languages. *Language* 68(1): 1–42.
- Harrison, K.D. 2005. Ethnographically informed language documentation. *Documentation and Description* 3, 22–41. London: SOAS.
- Haugen, E. 1972. *The Ecology of Language*. Stanford CA: Stanford University Press.
- Himmelmann, N.P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–195. doi:10.1515/ling.1998.36.1.161
- Himmelmann, N.P. 2002. Documentary and descriptive linguistics. In *Lectures on Endangered Languages*, Vol. V, O. Sakiyama & F. Endo (eds), 37–83. Kyoto: Endangered Languages of the Pacific Rim.
- Himmelmann, N.P. 2006. Language documentation: What is it and what is it good for? In *Essentials of Language Documentation* [Trends in Linguistics. Studies and Monographs 178], J. Gippert, N. P. Himmelmann & U. Mosel (eds), 1–30. Berlin: Mouton de Gruyter. doi:10.1515/9783110197730
- Himmelmann, N.P. 2008. Reproduction and preservation of linguistic knowledge: Linguistics' response to language endangerment. *Annual Review of Anthropology* 37: 337–350. doi:10.1146/annurev.anthro.37.081407.085226
- Himmelmann, N.P. 2012. Linguistic data types and the interface between language documentation and description. *Language Documentation and Conservation* 6: 187–207.
- Hinton, L. 1993. *Flutes of Fire*. Berkeley CA: Heyday Books.
- Hinton, L. 2002. *How to Keep your Language Alive*. Berkeley CA: Heyday Books
- Hinton, L. 2011. Revitalization of endangered languages. In *The Cambridge Handbook of Endangered Languages*, P.K. Austin & J. Sallabank (eds), 291–311. Cambridge: CUP. doi:10.1017/CBO9780511975981.015
- Hinton, L. 2013. *Bringing Our Languages Home: Language Revitalization for Families*. Berkeley CA: Heyday Books
- Hinton, L. & Hale, K. (eds). 2001. *The Green Book of Language Revitalization in Practice*. San Diego CA: Academic Press. doi:10.1163/9789004261723
- Holton, G. 2014. Mediating language documentation. In *Language Documentation and Description* 12, B. Nathan & P.K. Austin (eds), 37–52. London: SOAS.
- Hornberger, N.H. 2010. Language shift and language revitalization. In *The Oxford Handbook of Applied Linguistics*, R.B. Kaplan (ed.), 365–373. Oxford: OUP.

- Hymes, D. 1964. Introduction: Toward ethnographies of communication. *American Anthropologist* 66(6): 1–34. doi:10.1525/aa.1964.66.suppl_3.02a00010
- Kapono, E. 1995. Hawaiian language revitalization and immersion education. *International Journal of the Sociology of Language* 112: 121. doi:10.1515/ijsl.1995.112.121
- Kroskrity, P.V. 2015. Designing a dictionary for an endangered language community: Lexicographical deliberations, language ideological clarifications. *Language Documentation and Conservation* 9: 140–157.
- Lehmann, C. 2001. Language documentation: A program. In *Aspects of Typology and Universals*, W. Bisang (ed.), 83–97. Berlin: Akademie Verlag.
- Linn, M.-A. 2014. Living archives: A community-based language archive model. In *Language Documentation and Description* 12, D. Nathan & P.K. Austin (eds), 53–67. London: SOAS.
- Lüpke, F. 2009. Data collection methods for field-based language documentation. In *Language Documentation and Description* 6, P.K. Austin (ed.), 53–100. London: SOAS.
- Lüpke, F. & Storch, A. 2013. *Repertoires and Choices in African Languages*. Berlin: Mouton de Gruyter. doi:10.1515/9781614511946
- Mayer, M. 1969. *Frog Where Are You?* New York NY: Dial Books for Young Readers.
- Mühlhäusler, P. 1992. Preserving languages or language ecologies? A top-down approach to language survival. *Oceanic Linguistics* 31(2): 163–180. doi:10.2307/3623012
- Mühlhäusler, P. 2000. Language planning and language ecology. *Current Issues in Language Planning* 1(3): 306–367. doi:10.1080/14664200008668011
- Nathan, D. 2004. Documentary linguistics: Alarm bells and whistles? Seminar presentation, SOAS. 23 November 2004.
- Nathan, D. 2006. *Sound and unsound documentation: Questions about the roles of audio in language documentation*. Washington DC: Paper presented at the Georgetown University Roundtable on Linguistics.
- Nathan, D. 2009. The soundness of documentation: Towards an epistemology for audio in documentary linguistics. *Journal of the International Association of Sound Archives* 33. <www.iasa-web.org/book/iasa-journal-no-33-june-2009> (22 March 2012).
- Nathan, D. 2010a. Sound and unsound practices in documentary linguistics: Towards an epistemology for audio. *Language Documentation and Description* 7, 1–17. London: SOAS.
- Nathan, D. 2010b. Archives 2.0 for endangered languages: From disk space to MySpace. *International Journal of Humanities and Arts Computing* 4(1–2): 111–124. doi:10.3366/ijhac.2011.0011
- Nathan, D. 2014. Access and accessibility at ELAR, an archive for endangered languages documentation. In *Language Documentation and Description* 12, D. Nathan & P.K. Austin (eds), 187–208. London: SOAS.
- Nathan, D. & Austin, P.K. 2004. Reconceiving metadata: Language documentation through thick and thin. In *Language Documentation and Description* 2, P.K. Austin (ed.), 179–187. London: SOAS.
- Newman, P. 2003. The endangered languages issue as a hopeless cause. In *Language Death and Language Maintenance* [Current Issues in Linguistic Theory 240], M. Janse & S. Tol (eds), 1–13. Amsterdam: John Benjamins. doi:10.1075/cilt.240.03new
- Newman, P. 2009. Fieldwork and fieldmethods in linguistics. *Language Documentation and Conservation* 3(1): 113–125.
- Newman, P. 2013. The law of unintended consequences: How the endangered languages movement undermines field linguistics as a scientific enterprise. Seminar presented at SOAS, 15th October.

- Niedzielski, H.Z. 1992. The Hawaiian model for the revitalization of native minority languages and cultures. In *Maintenance and Loss of Minority Languages* [Studies in Bilingualism 1], W. Fase, K. Jaspaert & S. Kroon (eds), 369–384. Amsterdam: John Benjamins.
doi:10.1075/sibil.1.24nie
- Penton, D., Bow, C., Bird, S. & Hughes, B. 2004. Towards a general model for linguistic paradigms. In *Proceedings of EMELD-04*, 12 August 2013. <<http://emeld.org/workshop/2004/bird-paper.html>>
- Robins, R.H. & Uhlenbeck, E.M. 1991. *Endangered Languages*. New York NY: Berg.
- Romaine, S. 2007. Preserving endangered languages. *Language and Linguistics Compass* 1(1–2): 115–132. doi:10.1111/j.1749-818X.2007.00004.x
- Sallabank, J. 2012. From language documentation to language planning: Not necessarily a direct route. In *Potentials of Language Documentation: Methods, Analyses, and Utilization* [Language Documentation and Conservation Special Publication 3], F. Seifart, G. Haig, N.P. Himmelmann, D. Jung, A. Margetts & P. Trilsbeek (eds), 118–125. Honolulu HI: University of Hawai'i Press.
- Sallabank, J. 2013. *Endangered Languages: Attitudes, Identities and Policies*. Cambridge: CUP.
doi:10.1017/CBO9781139344166
- Schultze-Berndt, E. 2006. Linguistic annotation. In *Essentials of Language Documentation*, J. Gippert, N.P. Himmelmann & U. Mosel (eds), 213–251. Berlin: Mouton de Gruyter.
- Seifart, F. 2008. The representativeness of language documentations. *Language Documentation and Description* 5, 60–76. London: SOAS,
- Sherzer, J. 1987. A discourse-centered approach to language and culture. *American Anthropologist* 89(2): 295–309. doi:10.1525/aa.1987.89.2.02a00010
- Spolsky, B. 1989. Maori bilingual education and language revitalization. *Journal of Multilingual and Multicultural Development* 10(2): 89–106. doi:10.1080/01434632.1989.9994366
- Spolsky, B. 2003. Reassessing Maori regeneration. *Language in Society* 32: 553–578.
doi:10.1017/S0047404503324042
- Sugita, Y. 2007. Language revitalization or language fossilization? Some suggestions for language documentation from the viewpoint of interactional linguistics. In *Proceedings of First Conference on Language Documentation and Linguistic Theory*, P.K. Austin, O. Bond & D. Nathan (eds), 243–250. London: SOAS.
- Thieberger, N. 2012. Counting collections. *Paradisec Blog*, 29 November 2012. <<http://www.paradisec.org.au/blog/2012/11/counting-collections/>> (6 June 2015).
- Thieberger, N., Margetts, A., Morey, S., Musgrave, S. & Schembri, A. 2012. Assessing curated corpora as research output. Paper presented to the *Annual Conference of the Australian Linguistic Society*, University of Western Australia.
- Whalen, D. 2004. How the study of endangered languages will revolutionize linguistics. In *Linguistics Today – Facing a Greater Challenge*, P. van Sterkenburg (ed.), 321–342. Amsterdam: John Benjamins. doi:10.1075/z.126.16wha
- Whalen, D. 2013. Response to Paul Newman. Blog post dated 3 December 2013. <<http://elar-archive.org/blog/response-to-paul-newman-by-doug-whalen/>> (27 March 2015).
- Woodbury, T. 2003. Defining documentary linguistics. *Language Documentation and Description* 1, 35–51. London: SOAS.
- Woodbury, T. 2011. Language documentation. In *The Cambridge Handbook of Endangered Languages*, P.K. Austin & J. Sallabank (eds), 159–186. Cambridge: CUP.
doi:10.1017/CBO9780511975981.009

- Woodbury, A.C. 2014. Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In *Language Documentation and Description, 12: Special Issue on Language Documentation and Archiving*, D. Nathan & P.K. Austin (eds), 19–36. London: SOAS.
- Yamada, R.-M. 2007. Collaborative linguistic fieldwork: Practical application of the empowerment model. *Language Documentation and Conservation* 1(2): 257–282.