# Language documentation in the 21st century

**Peter K. Austin (SOAS, University of London)**

## Abstract

*Language documentation emerged as a new sub-field of linguistics in 1995 and has developed and expanded over the past 20 years. In this paper we outline the defining characteristics of language documentation as presented in the late 1990s and discuss some of the changes in the field that have occurred since. These include a move away from concern for best practices, standards and tools to a more critical and reflexive approach that highlights diversity and flexibility of individual documentation projects in their social, cultural and political contexts, as well as the need for greater attention to goals and outcomes. There have also been developments in archiving that build upon social networking models linking people to each other, rather than seeing documentation as being primarily about 'data' and 'resources'.*

## 1   Introduction[1]

In the last decade of the 20th century a new sub-field of linguistics emerged that has come to be known as 'language documentation' or 'documentary linguistics' (Himmelmann 1998, 2002, 2006, Lehmann 2001, Austin 2010a, Grenoble 2010, Woodbury 2003, 2011). In this paper we explore how it was defined in the seminal work of Himmelmann (1998) and others, including what were presented as significant characteristics that distinguished language documentation from language description. A focus on best practices, standards, tools and models for documentary corpora appeared in the following years, along with more critical discussions of the goals and methods of language documentation. The paper examines some current developments, including new approaches to archiving, and suggests that in the 21st century language documentation needs to adopt a more socially-engaged approach to linguistic research.

## 2   Defining language documentation

Language documentation (also known by the term 'documentary linguistics') is defined by Nikolaus Himmelmann 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 one response to the realisation among linguists, dating from around 1992, that a majority of the world's 7,000 languages are endangered, in that they are not being passed on to the next generation of speakers (Hale et al. 1992, Crystal 2000, Austin 2007, Whalen 2004). A desire among some researchers to create a lasting, and potentially unrepeatable, record of language use in its social and cultural context was one of the driving forces behind the interest in this new approach. There was also a concern for supporting speakers and communities who wished to maintain their languages by providing documentation that could feed into revitalisation efforts. Also playing a role were advances in information, media, communication and archiving technologies (see Nathan 2010a, 2010b) which made possible the collection, analysis, preservation and dissemination of documentary records 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 their direct involvement in the documentation and support of their own languages (see Grinevald 2003, Austin 2010).

A concurrent and supporting development that began around the year 2000 was the availability of extensive new funding resources for research from several sources, and the requirements of these funders to adopt a documentary perspective and to archive the recorded data and analyses. The new funding sources included:

- the Endangered Languages Documentation Programme (ELDP)[2] at SOAS which was established in 2002 by Arcadia Fund with a commitment of £15 million to sponsor documentation research across the world. ELDP has to date funded around 300 documentation projects and the Endangered Languages Archive (ELAR) at SOAS holds around 100 collections of material arising from the funded projects (see Figure 1);

- the Volkswagen Foundation DoBeS[3] project which was established in 2001 (after a pilot year in 2000) and funded 80 research projects to a value of over 60 million euros before its funding came to an end in 2013 (see Figure 2 for a map of DoBeS projects);

- the Documenting Endangered Languages (DEL) inter-agency programme of the National Science Foundation and the National Endowment for the Humanities which has funded 100 projects and awarded approximately $30 million in grants. The DEL programme is now a permanent component of the National Science Foundation budget;

- the European Science Foundation EuroBABEL initiative (Better Analyses Based on Endangered Languages)[4] which funded five projects between 2009 and 2012 with a budget of 8 million euros;

---

[2]  See http://www.hrelp.org/grants/ [accessed 2013-08-11]
[3]  See http://dobes.mpi.nl/dobesprogramme [accessed 2013-08-12]
[4]  See http://www.esf.org/?id=4632 [accessed 2013-08-12]

- smaller, more modest funders, such as the Endangered Language Fund (ELF), Foundation for Endangered Languages (FEL), Gesellschaft für bedrohte Sprachen (GfBS) and Unesco, which have provided hundreds of thousands of dollars in grants supporting scores of projects, especially ones that are community-based.

This level of funding had an influence on the topics that linguists (and others) chose to research, and the research methods they employed. The broader impact on the field of linguistics can be seen in the emergence of academic journals specialising in language documentation issues (Language Documentation and Conservation, Language Documentation and Description), specialist conferences, workshops and training courses (including InField and 3L summer schools, and the specialist MA and PhD programmes at SOAS (Austin 2008)), and a growing list of book publications on topics related to language documentation (for an annotated bibliography see Austin 2013).



Figure 1: Map of ELAR-deposits as at August 2013



Figure 2: Map of Volkswagen-funded DoBeS projects

Himmelmann (2006: 15) identified five main characteristics of language documentation that he proposed distinguish it from other approaches to human language:

- *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;

- *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;

- *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.

The application of these principles results, according to Himmelmann (1998, 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 and Simons 2003), and so some level of processing is required, and there is a need for systematic recording of metadata (data about the data) to make the archived documents understandable, findable 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, 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 (2006, 2008, 2010) proposes that there are five documentation activities which are identifiable in this 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;

- *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.

## 3   Best practices, tools and models

The establishment of the DoBeS project in 2001 (after a pilot year in 2000) gave a major boost to language documentation as an approach to linguistic data collection and analysis, and saw the emergence of a unified 'DoBeS model' for language documentation that the funded projects were expected to adopt[5]. This included specifications for archival formats, recommendations about recording and analysis formats, and the development of new software tools to assist with audio and video annotation (such as ELAN), and the creation and management of metadata (various IMDI tools). Researchers affiliated with DoBeS also wished to specify 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) and data collection methods (Lüpke 2009).

Definition of best practice, standards, tools and models was also a central goal of the E-MELD project[6] 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 papers[7] defining formats for lexical entries (Bell and Bird 2000), interlinear text (Bird and Liberman 2001, Bowe, Hughes, and Bird 2003), paradigms (Penton, Bowe, Bird and Hughes 2004) and a generalised ontology for glossing (Farrar, Lewis and Langandoen 2002, Farrar and Langendoen 2003a, b). E-MELD set up a 'School of Best Practices' (Aristar 2003, Aristar-Dry 2004)[8] 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'.

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 proposed that an adequate documentation should cover:

(i)   all the basic phonology, both low-level and morphophonemic

(ii)   all the basic morphology

(iii)  all the basic syntactic constructions (in context)

(iv)  a lexicon which (a) covers all the basic vocabulary and important areas of special expertise in the culture, and (b) provides at least glosses for all words/morphemes in the corpus

(v)   a full range of textual genres and registers

---

[5]   See http://dobes.mpi.nl/dobesprogramme and http://www.mpi.nl/corpus/a4guides/a4-guide-dobes-format-encoding.pdf [accessed 2013-08-12]

[6]   See http://emeld.org/ [accessed 2013-08-12]

[7]   See http://emeld.org/documents/index.cfm#loc-papers [accessed 2013-09-12]

[8]   See also http://emeld.org/school/index.html [accessed 2012-08-12]

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 1 million words (around 1200 hours of recorded speech). Other qualitative measures were suggested such as '[o]ne is done 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-minority languages, yet this is precisely what was suggested for language documentation, especially for projects involving endangered languages in particular.

Both DoBeS and E-MELD were influential in getting 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 challenged by some researchers, beginning around 2004.

## 4    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 notion that quantifiable properties such as recording hours, data volume, and file parameters, and technical desiderata like 'archival quality' and 'portability' had become 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 among language documenters about what such quality measures might consist in.

Nathan and Austin (2004) addressed the issue of metadata and argued that all value-adding that researchers do to the audio or video records they make should be understood as metadata, and that it should be as rich as possible and not constrained by specifications in the form of an 'ontology' or standard minimal set (such as the OLAC metadata set[9]). The need for richer metadata and meta-documentation (documentation of the language documentation) was further elaborated on by Austin (2009, 2013).

Two important issues for the definition of language documentation were raised at the Georgetown Round Table in Linguistics in 2006, namely the difference between documentation and description which was considered fundamental by Himmelmann (Austin 2006b), and the approach to audio recording within documentation (Nathan 2006). 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 linguistic structures and systems, and typically aims at the production of 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 types of language use. Although description may draw on a corpus, it involves analysis of a different order, aiming to provide an understanding of language at a more abstract level, as a system of elements, rules, and constructions.

---

[9]    See http://www.language-archives.org/OLAC/olacms.html [accessed 2013-08-12]

Austin and Grenoble (2007: 22) challenged the sharp separation of description and documentation advocated by Himmelmann 1998 and pointed out 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:

- in 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;

- in documentation, linguistic knowledge and documentary techniques are applied to some event in the real world to make a recording (audio or video) 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 and preservation). In relation to the recording, the researcher makes decisions and applies 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 may be archived or mobilised, or otherwise used to support language documentation and support goals. The 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 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 in their attempts to collect language data. 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 approaches to endangered languages research can be found in Dobrin, Austin and Nathan (2007) who argue against what they see as tendencies towards objectification, and reliance on familiar metrics to measure quality, progress and value in language documentation. More specifically, they claim 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'. They echo Nathan (2005) in pointing to archivism as problematic, and join Nathan (2006) in arguing that documentary linguists show little or no knowledge about recording arts and microphone types, properties and placement, even though microphone choice and handling is the single greatest determiner of recording quality. They also assert that evidence from archival deposits shows that video tends to be poorly used by documentary linguists, with video recordings being made without reference to hypotheses, goals, or methodology, simply because the technology is available, portable and relatively inexpensive. Finally, in contrast to earlier conceptions, 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, actual documentation projects, as evidenced by grant project proposals and materials deposited in archives, show a diversity of approaches, techniques, methodologies, skills and responses. Rather than aiming for comprehensiveness or representativeness, research funded recently by ELDP for example, shows rather specificity, focussing on topics such as bark cloth making, libation rituals, fishing practices, child language, interactive speech, and ethnobotany, to mention just some of the projects funded in 2012.[10]

Interestingly, in a recent handbook of language documentation Woodbury (2011) presents a definition which reflects this shift away from representative samples towards more specific goals: 'language documentation is 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'. 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 and Sallabank 2014). As Woodbury (2011) notes, 'humans experience their own and other people's languages 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:

- the identity of stakeholders involved and their roles in the project

- attitudes and ideologies of language consultants and the narrower and broader communities within which they are located, both towards their languages and towards the documenter and documentation project

- the relationships with researchers, research project participants and the wider community

- the goals and methodology adopted within the project, including research methods and tools (see Lüpke 2010), corpus theorisation (Woodbury 2011), theoretical assumptions embedded in annotation (abbreviations, glosses), and considerations of the potential for a project to contribute to revitalization

---

[10]   See http://www.hrelp.org/grants/projects/index.php?year=2012 [accessed 2013-08-12]

- 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 gained by neighbouring 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).

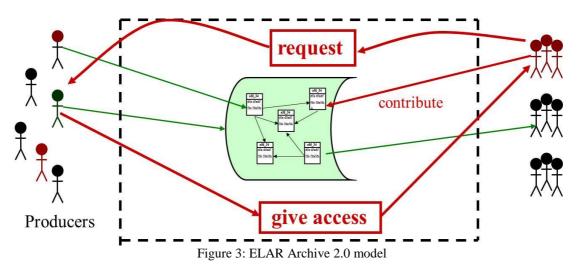## 5   Quality of documentation outcomes

- compliance with some widely agreed standards in data and metadata representation – currently Unicode for character encoding and XML for text encoding are widely recognised as de facto standards in language documentation (and elsewhere), however there seems to be little other agreement about any possible standards and compliance. Indeed, the community of documenters has been slow to adopt any of the 'best practices' proposed by E-MELD and other groups;

- architecture of the data and modelling of the knowledge domain so that representations comply with some expressed data model and show internal and rigorous consistency;

- range and comprehensiveness of the data and analysis, in terms of such things as the genres present in a speech community as determined by a well-grounded ethnography of speaking;

- uniqueness of the project in terms of the language(s) or ways of speaking documented, or the particular approach taken by the documentation team;

- the ethical context of the project, and the ways it responds to expressed needs of the participants and the community within which it is located.

In any given instance, these metrics may be in conflict and a delicate balance between them may have to be struck. Perhaps what are most needed for language documentation in the 21st century are examples of review assessments so that the field can establish accepted measures of 'functionality, import, and scope'.

## 6   Developments in archiving

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 by some language documentation archives (such as ELAR at SOAS) leading to what Nathan (2010b) calls 'Archives 2.0'.

Traditionally, archiving has focussed heavily on preservation, however language documentation often deals with highly sensitive topics (such as sacred stories that may be restricted in terms of who can be exposed to them, or gossip which may contain references to private knowledge or events). As a result, language archives need

powerful but flexible access management that is transparent in terms of being easy to understand and to change as circumstances develop. The basis for access needs to be through relationships established between the materials providers (archive depositors and the stakeholders they work with) and those who wish to use the materials. To achieve this the archive thus needs to be 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:



Figure 3: ELAR Archive 2.0 model

Further Archive 2.0 developments include:

- progressive archiving, where 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;

- reworking the archive interface to provide contextualization, different degrees of presentation, and ease of navigation. The interface directly reflects the interests and needs of the materials providers and the users;

- increasing participation so that users can negotiate access and bookmark their favourite materials, depositors can negotiate access requests and monitor usage, and both groups can exchange and share information

Possible future directions may include community curation of archived materials (Linn 2013), participant identification and expression of rights (Garrett 2013), 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 2013).

# 7 Conclusions

The past 20 years has seen the emergence and gradual development of a new sub-field of research called 'documentary linguistics' or 'language documentation' which has concentrated on recording, analysing, preserving and disseminating records of language 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 recognition of diversity of goals, methods and outcomes of language documentation, and the introduction of social models of research, especially in the area of archiving. Much work remains to be done however, 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.

# References

Aristar, Anthony Rodrigues. 2003. The school of best practice. SOAS Workshop on Archives for Endangered Languages, London, November 21-22.

Aristar-Dry, Helen. 2004. E-MELD School of best practices in digital language documentation. Presentation at E-MELD Conference 2004: Workshop on linguistic databases and best practice. Detroit, Michigan, July 15-18.

Austin, Peter K. 2006a. Data and language documentation. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation* (*Trends in Linguistics. Studies and Monographs*, 178), 87-112. Berlin: Mouton de Gruyter.

Austin, Peter K. 2006b. Defining language documentation. Paper presented at the Georgetown University Roundtable on Linguistics, Georgetown University, Washington, DC, March.

Austin, Peter K. 2007. Survival of Languages. In Emily F. Shuckburgh (ed.), *Survival: Darwin College Lectures*, 80-98. Cambridge: Cambridge University Press.

Austin, Peter K. 2008. Training for language documentation: Experiences at the School of Oriental and African Studies. In Margaret Florey & Victoria Rau (eds.), *Documenting and Revitalising Austronesian Languages*, 25-41. Language Documentation and Conservation Special Publication No. 1. Hawaii: University of Hawaii Press

Austin, Peter K. 2010. Current issues in language documentation. *Language Documentation and Description, Volume 7,* 12-33. London: SOAS.

Austin, Peter K. 2009. Meta-documentary linguistics. Paper given at the Aboriginal languages workshop, Kioloa, Australia, March 2009.
[Powerpoint slides available at http://www.slideshare.net/pkaustin/2010-march-kioloa]

Austin, Peter K. 2013. Language documentation and meta-documentation. In Sarah Ogilvie & Mari Jones (eds.), *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*, 3-15. Cambridge: Cambridge University Press.

Austin, Peter K. & Lenore Grenoble. 2007. Current trends in language documentation. *Language Documentation and Description, Volume 4*, 12-25. London: SOAS.

Austin, Peter K. & Stuart McGill. 2010. *Essential readings in endangered languages*. London: Routledge.

Austin, Peter K. & Julia Sallabank. 2011. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.

Austin, Peter K. & Julia Sallabank. 2014. *Endangered Languages: Ideologies and beliefs in language documentation and revitalization*. London: British Academy.

Bell, John & Steven Bird. 2000. Preliminary study of the structure of lexicon entries. Proceedings of the Workshop on Web-Based Language Documentation and Description, Philadelphia, December.

Bird, Steven & Mark Liberman. 2001. A formal framework for linguistic annotation. *Speech Communication*, 33 (1,2): 23-60.

Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3): 557-82.

Bow, Cathy, Baden Hughes & Steven Bird. 2003. Towards a general model for interlinear text. *Proceedings of EMELD-03*
http://emeld.org/workshop/2003/bowbadenbird-paper.html
[accessed 2013-08-12]

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.

Coelho, Gail. 2005. Language documentation and ecology: areas of interaction. *Language Documentation and Description, Volume 3*, 63-74**.** London: SOAS.

Crystal, David. 2000. *Language death*. Cambridge: Cambridge University Press.

Dobrin, Lise, Peter K. Austin & David Nathan. 2007. Dying to be counted: commodification of endangered languages in documentary linguistics. *Language Documentation and Description, Volume 6*, 37-52. London: SOAS.

Eisebeiss, Sonja. 2005. Psycholinguistic contributions to language documentation. *Language Documentation and Description, Volume 3*, 106-140. London: SOAS.

Farrar, Scott & D. Terence Langendoen. 2003a. A linguistic ontology for the semantic web. *GLOT International* 7 (3): 97-100.

Farrar, Scott & D. Terence Langendoen. 2003b. Markup and the GOLD ontology. *Proceedings of EMELD-03*.
http://staff.washington.edu/farrar/documents/inproceedings/FarLang03a.pdf
[accessed 2013-08-12]

Farrar, Scott, William D. Lewis & D. Terence Langendoen. 2002. A common ontology for linguistic concepts. Proceedings of the Knowledge Technologies Conference, Seattle, Washington
http://staff.washington.edu/farrar/documents/inproceedings/FarLewLang02a.pdf
[accessed 2013-08-12].

Garrett, Edward. 2013. Participant driven language archiving. *Language Documentation and Description*, *Volume 12*. London: SOAS.

Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.). 2006. *Essentials of language documentation* (*Trends in Linguistics. Studies and Monographs*, 178). Berlin: Mouton de Gruyter.

Grenoble, Lenore. 2010. Language documentation and field linguistics: The state of the field. In Lenore A. Grenoble & N. Louanna Furbee (eds.), *Language documentation: Practice and values*, 289-309. Amsterdam: John Benjamins.

Grinevald, Colette. 2003. Speakers and documentation of endangered languages. *Language Documentation and Description, Volume 1*, 52-72.

Hale, Kenneth, Michael Krauss, L. J. Watahomigie, Akira Y. Yamamoto, Colette Craig, Laverne M. Jeanne & Nora C. England. 1992. Endangered languages. *Language* 68 (1): 1-42.

Harrison, K. David. 2005. Ethnographically informed language documentation. *Documentation and Description, Volume 3*, 22-41. London: SOAS.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-95.

Himmelmann, Nikolaus P. 2002. Documentary and descriptive linguistics. In Osamu Sakiyama & Fubito Endo (eds.), *Lectures on endangered languages, Volume V*, 37-83. Kyoto: Endangered Languages of the Pacific Rim.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation* (*Trends in Linguistics. Studies and Monographs*, 178), 1-30. Berlin: Mouton de Gruyter.

Holton, Gary. 2013. Mediating language documentation. *Language Documentation and Description, Volume 12*. London: SOAS.

Lehmann, Christian. 2001. Language documentation: A program. In Walter Bisang (ed.), *Aspects of typology and universals*, 83-97. Berlin: Akademie Verlag.

Linn, Mary-Anne. 2013. Living archives: A community-based language archive model. *Language Documentation and Description, Volume 12*. London: SOAS.

Lüpke, Friederike. 2009. Data collection methods for field-based language documentation. *Language Documentation and Description, Volume 6*, 53-100. London: SOAS.

Nathan, David 2004. Documentary linguistics: alarm bells and whistles? Seminar presentation, SOAS. 23 November 2004.

Nathan, David. 2006. Sound and unsound documentation: Questions about the roles of audio in language documentation. Paper presented at the Georgetown University Roundtable on Linguistics, Georgetown University, Washington, DC, March.

Nathan, David. 2009. The soundness of documentation: Towards an epistemology for audio in documentary linguistics. *Journal of the International Association of Sound Archives* (IASA) 33.
Available online at www.iasa-web.org/book/iasa-journal-no-33-june-2009
[accessed 22 March 2012]

Nathan, David. 2010a. Sound and unsound practices in documentary linguistics: towards an epistemology for audio. *Language Documentation and Description, Volume 7*, 1-17. London: SOAS.

Nathan, David. 2010b. Archives 2.0 for endangered languages: From disk space to MySpace. *International Journal of Humanities and Arts Computing* 4 (1-2): 111-124.

Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: language documentation through thick and thin. In Peter K. Austin (ed.), *Language Documentation and Description, Volume 2*, 179-187. London: SOAS.
Penton, David, Cathy Bow, Steven Bird & Baden Hughes. 2004. Towards a general model for linguistic paradigms. *Proceedings of EMELD-04*.
http://emeld.org/workshop/2004/bird-paper.html
[accessed 2013-08-12]

Seifart, Frank. 2008. The representativeness of language documentations. *Language Documentation and Description, Volume 5*, 60-76. London: SOAS.

Whalen, Doug. 2004. How the study of endangered languages will revolutionize linguistics. In Piet van Sterkenburg (ed.), *Linguistics Today – Facing a greater challenge*, 321-342. Amsterdam: John Benjamins.

Woodbury, Tony. 2003. Defining documentary linguistics. *Language Documentation and Description*, *Volume 1,* 35-51. London: SOAS.

Woodbury, Tony. 2011. Language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 159-186. Cambridge: Cambridge University Press.