On conceptualizing documentary linguistics as an independent field of

linguistic research: An effective approach to preserving endangered languages

Mark Nartey

Department of Language and Literature

Norwegian University of Science and Technology

Trondheim, Norway.

Email: narteynartey60@gmail.com

Telephone: (+47) 45785962

Biographical Information

A budding academic, Mark Nartey holds a First Class B.A (Hons) degree in English and Linguistics from the University of Cape Coast, Ghana, where he also graduated as the overall Best Graduating Arts Student as well as the Best Graduating Student of English. Currently, he is a Master of Philosophy (Linguistics) student in the Department of Language and Literature at the Norwegian University of Science and Technology, Trondheim, Norway. His research interest areas include Academic Communication, English for Specific Purposes (ESP), Discourse Studies and Applied Linguistic Theory. Email address: narteynartey60@gmail.com

English for Specific Purposes World, ISSN 1682-3257, http://www.esp-world.info, Issue 41, Vol. 14, 2013

Abstract

Much of the work that is labeled 'descriptive' within linguistics comprises two activities – the

collection of primary data and a (low level) analysis of these data. These are indeed two separate

activities as established by the fact that the methods employed in each activity differ

substantially. Till date, the field concerned with the first activity – called documentary linguistics

in this study – has received very little scholarly attention. In this paper, I argue in support of the

notion that documentary linguistics be conceived of as a fairly independent field of enquiry and

practice which is not necessarily (or no longer) linked exclusively to the descriptive framework. I

also establish a nexus between the re-conceptualization of documentary linguistics as an

independent framework and the preservation of endangered languages. The study holds

implications for language documentation, preservation and revitalization.

Key words: language documentation, language description, endangered language

1. Background

In the last two and half decades, we have seen the emergence of a branch of linguistics which

has come to be called 'documentary linguistics', owing to the quintessence of language

documentation. It is concerned with the making and keeping of records of the world's

languages and their patterns of use. In the words of Woodbury (2011), "Language

documentation aims to provide a comprehensive record of the linguistic practices

characteristic of a given speech community, and therefore seeks to create as thorough a

record as possible of the speech community for both posterity and language

revitalization" (p.15). Language documentation, it can thus be posited, provides a firmer

foundation for linguistic analysis in that it creates a citable set of materials in a language on which claims about the structure of the language can be based.

The emergence of documentary linguistics, it must be noted, has taken place alongside major changes in the technology of linguistic data representation and maintenance, new attention to linguistic diversity and an increasing focus on the threats to that diversity by the endangerment of languages and language practices around the world, especially in small indigenous communities. Perhaps, and more importantly, this emergence has taken place alongside the discipline's growing awareness that linguistic or language documentation has crucial stakeholders well beyond the academic community, in endangered language communities themselves, and even beyond these communities. Given this, interest in the state of the world's linguistic heritage has increased dramatically, such that there is now a considerably greater effort being put towards language development in situations where such development is a viable proposition. In cases where development or stabilization is no longer feasible, for example due to reduced numbers of speakers, documentation of such languages is still nonetheless desirable, as a means of preserving both the unique linguistic structures that might exist in these languages, as well as the cultural and other forms of knowledge embodied in them (Akinlabi & Connell, 2008; Evans, 2008).

Accordingly, research programs have been established by existing funding agencies such as the DOBES program of the Volkswagen Foundation (Germany) or the Documentation of Endangered Language program set up by the National Science Foundation in the US and the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) in Australia. Other organizations, such as the EMELD consortium, have also come into

existence with the aim of coordinating efforts towards universal standards. Not surprisingly, this interest in endangered languages has brought on a number of what Akinlabi and Connell (2008) call 'knock-on' effects: increased recognition and understanding of the relationship between language and culture, and language and knowledge. Additionally, there has been landmark development in methodological issues, procedures involved in fieldwork, the availability of increasingly sophisticated software tools and, in particular, documentation techniques.

Although it has been fairly suggested in the extant scholarship that language documentation and/or documentary linguistics has arisen as a new sub-discipline within linguistics, many scholars continue to use the terms 'language description' and 'language documentation' interchangeably, while other scholars simply consider language documentation and documentary linguistics merely as ancillary to language description and descriptive linguistics (Akinlabi & Connell, 2008; Gippert, 2006). Against this backdrop, the present paper argues in favor of the conceptual separation of language documentation from language description, and more crucially advances the notion that such a separation is an effective approach to preserving endangered languages. In the subsequent paragraphs, I present the rationale undergirding the study, after which I sketch a short description of documentary linguistics. Thereafter, I discuss the disciplinary distinction and conceptual separation of documentary linguistics from descriptive linguistics and the concomitant relationship with preserving endangered languages, followed by the conclusion.

2. Aim of the study

Previous definitions of documentary linguistics have focused on its inter – and trans – disciplinary nature, thereby giving rise to some misconceptions regarding its basis and role in linguistics (Himmelmann, 2012). One of such misconceptions, undeniably, is the often subsidiary role and appendage designation given to documentary linguistics viz-a-viz descriptive linguistics. Following from this, some previous studies, including Himmelmann (2004, 2012), Lehmann (2001) and Gippert (2006), have argued on the need for documentary linguistics to be kept as an independent field of scientific linguistic enquiry. Owing to these studies and similar ones, it can be conveniently opined that the distinction between documentary linguistics and descriptive linguistics has both tacitly and explicitly been suggested in the literature. However, there appears to be silence on the relationship between or the ramifications such a distinction has for the preservation of endangered languages.

In this regard, this study, first and foremost, is a modest contribution to the conceptual distinctiveness of documentary linguistics as a field of linguistic research, and second, sets out to fill the niche in the literature by establishing a correlation between this distinctiveness and the preservation of endangered languages.

3. A vignette of documentary linguistics

3.1 *Documentation is not a modern-day phenomenon*

There has long been concern for the perspicuous documentation and description of the world's languages. We see this in the now century-old tradition of monograph series and journals of record in which texts, dictionaries, grammars, vocabularies and other works have

been published. We can see too that such work has been foundational for the discipline's more theoretical endeavors since at least the time of Franz Boas (Woobury, 2011). Referred to as the Boasian trio (Thieberger, 2009), dictionaries, grammars and texts have informed historical linguistics and the re-construction of linguistic pre-history, of genetic language families, and of patterns of pre-historic linguistic contact. They have informed enquiry into the methods and tools for linguistic description and discovery, and they have informed the development and testing of linguistic typology and of universal grammar.

Both documentation and description have been foundational too in having kept linguists in the field, observing language in its social context, and through that it has led directly to work on the use and function of language in specific speech communities. Further, as Evans (2008) rightly notes, practitioners of documentary and descriptive linguistics, since ancient times, have always operated in an atmosphere of urgency and impending language loss, making lasting records and, in some cases, taking part in community efforts at language preservation, teaching, planning and revival.

3.2 *Data collection and language documentation in linguistics*

I subscribe to Lehmann's (2001:2) view that, as in any other empirical science, a first step in any linguistic work devoted to a language is data collection. And in linguistics, as in other empirical sciences, the status of the individual datum has been accorded a relatively low status, given that it appears the individual datum is not, generally, of interest in and of itself, but only a representative of a class of data that share particular characteristics. Data are considered or assumed to be easily reproducible, so data collection is accorded a correspondingly low status relative to the theoretical ends they serve.

Language, however, as the object of linguistic research, is unlike the objects of other scientific endeavors. This stems from the fact that language is rooted in a speech community, in its history and its culture; as a result, a language is a living object. In so far as this is true, each language is unlike any other language – linguistic data are not easily replicable and utterances, even of a living language, are not replaceable at will; better yet, data from an extinct language are irreplaceable. One implication of this is that linguistic data should be treated with a degree of care and dignity not typically given to data in linguistics, let alone other empirical sciences, where theory has been of the uttermost importance (Lehmann, 2004). This notwithstanding, it can be intimated that data collection is in and of itself not a sufficient goal, nor can such an activity be seen as equivalent to documentation. This is because it is entirely plausible, and perhaps and all too frequently the case, that collected data may be inaccessible, in that it is poorly organized, insufficiently annotated, or in any other ways difficult to use for every one save the person who collected it.

It stands to reason, therefore, that a documentation includes a considerably wider range of material than is, normally, included when one thinks of 'data collection' in linguistics, and it involves organizing and annotating data in such a way that it is accessible to others, as is extensively discussed by scholars such as Bird & Simons (2002), Himmelmann (1998, 2006) and Lehmann (2001, 2002). It goes without saying, then, that data should be shared, indeed made publicly accessible, rather than be permitted to languish – this, I dare say, is especially true in instances of endangered language data.

3.3 *On defining language documentation*

What then is 'documentation', and how does it differ from 'data collection' on one hand, and as suggested already, from 'description' on the other hand? As a starting point, Lehmann (2001:5) draws attention to the etymology of the word, where the Latin term *documentum* can be glossed as 'a thing for teaching', so that with respect to language, a documentation involves creating (and archiving) record of linguistic material that may serve to teach others about the language. The primary purpose of a documentation, then, is "to represent the language for those who do not have direct access to the language itself', independent of the motivation of the user (Lehmann, 2001:5). That is, documentation involves the archiving, corpus-building and (large) lexico-graphic projects which serve as the basis of orthography development, the production of a dictionary or a grammar, or teaching primers, as well as providing language materials for linguistic analysis at all levels.

The implication of the preceding paragraph, I argue, is that language documentation requires a deliberate and specific curation of the data collected, together with a high degree of representativeness of the structures of the language (at all linguistic levels), how it is used in as wide a range of contexts as possible, and the role(s) that it plays in its society. In the context of current work on endangered languages, it must be borne in mind that a documentation also needs to serve these purposes in the potential "permanent absence of native speakers" (Hyman, 2003). These characteristics make it clear that a documentation is quite different from collections of data, which are often focused on a particular topic and/or purpose, and constructed without thought to more general accessibility.

Now until recently, and still in much current usage, the terms 'language documentation' and 'descriptive work' have been used to describe activities that included both data collection and some degree of analysis and presentation of linguistic data. The term 'descriptive linguistics' has been used in opposition to 'formal linguistics', the latter referring to work that is in its orientation theoretical and formal, while the former refers to an informal presentation of the facts of a language often considered to be outside the confines of a particular theoretical framework. As is sometimes claimed, this is not to suggest that such work is 'atheoretical' or 'pretheoretical' (Hyman, 2001). As Himmelmann (1998: 165) points out, descriptive linguistics has also come to be associated with work on under — or undescribed languages, little known languages and by extension has also been taken to mean the 'documentation' of a language. Owing to this, the two terms, 'description' and 'documentation' have frequently been used interchangeably, and linguists doing descriptive work on a language have often assumed that they are 'documenting' the language.

While the above may be true in a certain sense, it is not documentation of the sort described above, nor is a documentation as described above a description of a language as one expects, for example, of a descriptive grammar or an exposé of the phonology of a language (Dryer, 2006). A description of a language is an account of the system underlying the data. Thus, the nature of a language description is determined, on one hand, by the nature of human language and its components and, on the other, by a system (or sub-system) of the individual language in question, as well as by the methodological perspectives taken on it by the researcher/analyst. A description is, therefore, at a different level, a meta-level, with respect to a documentation (Lehmann, 2001), and is from this perspective distinct from a documentation.

To summarize, thus far, I quote again from Lehmann (2001) that a documentation is an activity (or its result) that,

"gathers, processes and exhibits a sample of data of the language that is representative of its linguistic structure and gives a fair impression of how and for what purposes the language is used. Its purpose is to represent the language for those who do not have access to the language itself,

where as a description of a language is an activity (or its result) that,

"formulates, in the most generable way possible, the patterns underlying the linguistic data. Its purpose is to make the user of the description understand the way the language works". (Lehmann, 2001:7).

From the foregoing, it can be deduced simply that documentation is specific and concrete; conversely, description is general and abstract.

It should be clear from the above that language documentation is not simply or merely a collection of data, but that it pre-supposes a certain degree of analysis; call it a description. Nonetheless, such an analysis does not make language documentation equivalent to language description or as an addendum to it.

4. Re-conceptualizing documentary linguistics as an independent field of linguistic enquiry

As already mentioned, the task of recording a little-known language comprises two activities, the first being the collection, transcription and translation of primary data, and the second, a low-

level (i.e. descriptive) analysis of these data (Himmelmann, 1996). Given that the two activities are closely inter-related for various epistemological, methodological and practical reasons, there is a strong tendency to neglect the difference (s) between them and to consider them part of a single uniform project called 'describing language'. However, following from Himmelmann (1998) and Lehmann (2001, 2002), I argue against this widespread conception in favor of the view that documentary linguistics be conceived of as a fairly independent field of linguistic enquiry and practice which is no longer linked exclusively to the descriptive framework – such a position, I maintain, is an effective approach in the quest to preserve endangered languages given its rewarding potential. At the outset, I wish to state that this paper does not imply that a link between documentation and description does not exist. Rather, it refutes the conventional view that the documentary activity is ancillary to the descriptive activity (i.e. primary data are collected in order to make a descriptive statement of the language). And instead suggests that descriptive techniques are part of a broad set of techniques applied in compiling and presenting a useful and representative corpus of primary documents of the linguistic practices found in a given speech community. In the subsequent paragraphs, I shed light on the need for documentary linguistics to be conceived of as an independent field of linguistic enquiry and the ramification such a position has for the preservation of endangered languages.

Because of the association of descriptive linguistics with work on little-known languages, descriptive linguistics is also deemed to be competent for and, perhaps, in charge of the data collection and handling procedures necessary when working on little-known languages. A further extension of this view, as Himmelmann (1998) notes, then leads to the widespread belief that 'describing a language' is – in the case of previously unrecorded languages – more or less

coterminous with 'documenting a language'. This view is, however, problematic for the following reason: The two activities, i.e. the collection of primary data and descriptive analysis of these data, differ substantially with respect to the methods and procedures employed as well as to their immediate results.

For example, the collection of primary data may result in some 50 utterances, in many of which a segment lu occurs. The methods used in putting together this sample may include participant observation, various forms of elicitation or recording, transcribing and translating a text. Methodological issues arise with respect to the reliability, naturalness and representativeness of the data. The second activity (the analysis of these kind of data) leads to statements such as that in language X, an ergative case exists, formally expressed by a suffix lu, which is part of a case paradigm and has A, B and C further syntactic and semantic properties. The procedures used in arriving at such statements, *inter alia*, involve distributional tests like commutation and substitution, and the analysis of the semantic properties of the utterances containing lu. Methodological issues arise with respect to the definition of the notions 'ergative', 'suffix', etc. and the kind of evidence adduced for analyzing a certain segment as an ergative suffix.

Indeed, the dichotomy between documentation and description expounded above, for which it is fair for documentary linguistics to be conceptualized a distinct discipline from descriptive linguistics, despite the partial overlap that may exist between the two fields of study is reinforced by Himmelmann (1998) when he asserts that the fact that in current usage the term 'descriptive' also often comprises the activity of collecting the data prerequisite to a descriptive analysis is but a historical coincidence and should not be permitted to gloss over the fact that collecting activity and analytic activity differ substantially, both with respect to their immediate results as well as

the methodological issues involved. He goes on further to say that the primary goal of descriptive linguistics has been, and continues to be, describing the system of a language in synchronic and non prescriptive terms. Clearly, thus, the question or the issue of on which data a description be based has always been of secondary relevance to descriptive linguists. To the extent that the primary focus of descriptive linguistics is not 'the data' (here, an endangered language), I assert that descriptive linguistics, strictly speaking, is not the best approach in a bid to preserve an endangered language. By conceptualizing the documentation process as a discipline on its own, it makes it, pragmatically, feasible for some extensive work to be done on an endangered language since the language is now of primacy.

A second point on the need for documentary linguistics to be considered as an independent field of linguistic research stems from the fact that in recent decades, hardly any comprehensive collections of primary data have been produced. Largely, this is because the collection and presentation of primary data is often conceived of as an ancillary procedure within another research framework, and not as a goal in itself (Kenneth et al. 1992). However, a clear separation between documentation and description will ensure that the collection and presentation of primary data receive the theoretical and practical attention they deserve. Ultimately, such an approach will, in no doubt, prove useful for endangered languages. The concept of language documentation as a field of linguistic research and activity in its own right proceeds based on the assumption that the linguistic practices and traditions in a given speech community are worthy of documentation in the same way as material aspects of its culture (for example, arts and crafts) are, generally, deemed worthy of documentation. The aim of a language documentation, then, is to provide a comprehensive record of the linguistic practices characteristic of a given speech

community. Indeed, as Pawley (1993) intimates, linguistic practices and traditions manifest in two ways: 1) the observable linguistic behavior (which manifests in everyday interaction between and among members of the speech community), and 2) the native speaker's metalinguistic knowledge (which manifests in their ability to provide interpretations and for linguistic units and events).

I must stress at this point that the aim of a language documentation explicitly provided above differs, fundamentally, from the aim of language description in that a language description, as already suggested in this paper, aims at the record of a *language*, with language being understood as a system of abstract elements, constructions and rules which constitute the invariant underlying structure of the utterances observable in a speech community (Himmelmann, 1998). To this end, it is reasonably appropriate for me to postulate that reconceptualizing language documentation and indeed documentary linguistics as an enterprise in itself will be an effective and pragmatic approach worth adopting in the quest to preserve endangered languages.

Third, my thesis that language documentation and documentary linguistics, in general, be reconceptualized as an independent framework proceeds from the further notion (though related to the second) that a corpus of extensively annotated primary data documenting linguistic practices and traditions is of use for a variety of purposes. These include further analysis in the framework of a language-related discipline as well as projects concerning the cultivation and maintenance of its linguistic practices administered by the speech community. Conversely, no single one of the possible specific uses of a language documentation provides the major guidelines for data collection in this framework. Instead, as pointed out by Himmelmann (1998), Borsley (2005) and

Woodbury (2011), the make-up and contents of a language documentation are determined and influenced by a broad variety of language-related (sub) disciplines, including the following:

- sociological and anthropological approaches to language (variationist sociolinguistics, conversational analysis, linguistics and cognitive anthropology, language contact, etc)
- 'hardcore' linguistics (theoretical, comparative and descriptive)
- discourse analysis, spoken language research and rhetoric
- language acquisition
- phonetics and phonology
- ethics, language rights and language planning
- field methods
- oral literature and oral history
- philology and corpus linguistics
- educational linguistics

These approaches to language influence language documentation procedures, at least on two counts. First, they influence the collection process inasmuch as they contribute to the compilers' understanding of linguistic practices and traditions (and hence, influence the choice of data to be recorded). Second, they influence the recording and presentation of the data inasmuch as certain kinds of information are indispensable for a given analytical procedure (for instance, no phonetic analysis is possible without some high-quality sound recording; no analysis of gestures is possible without videotaping, etc.). From the foregoing, it is perceptible that text curation and documentary linguistics are much bigger enterprises to be considered merely complementary to something else, in this case descriptive linguistics.

Although various analytic frameworks with differing importance are applicable to language documentation, it must be mentioned that the important point to be noted here is that language documentation is not to be conceived of as an ancillary procedure to any of the research frameworks. On the contrary, I posit that the documentation itself is the central project, while the various analytic frameworks help to ensure the quality and usefulness of the documentation. Thus, within the documentary framework, aspects of various approaches to language (as mentioned above) – *including descriptive linguists* – are unified in the task of providing the linguistic practices and traditions of a speech community. Such a re-conceptualization, I believe, comes in handy and indeed is useful in a bid to document, and thereby preserve endangered languages.

A fourth point on the need for documentary linguistics to be conceived of as an independent framework emerges from the two immediate previous points discussed. This point concerns a very patent and salient distinction between language documentation and language description – the role of primary data within the two frameworks. Within the documentation framework, primary data are of central concern – the goal then is to present as many primary data with as much analytical information as possible (Himmelmann, 1998). Therefore, the analytic information is given in the form of a commentary (or apparatus) appended to the primary data. Within the descriptive framework, on the other hand, primary data are a means to an end, i.e. the analysis of the language system. Analytic statements are of central concern and primary data are, generally, not presented in full, but only as exemplifications of analytic statements. Given this distinction between documentation and description, it stands to reason that considering the documentation process as codicil or supplement to the descriptive framework will not augur well

for a comprehensive documentation of languages, in general, but of endangered languages, in particular. Obviously, the collection of primary data is critical and principal to the preservation of endangered languages, and against the backdrop that language documentation (and not language description) emphasizes the role and primal relevance of primary data, this paper advances the view that re-conceptualizing documentary linguistics as an independent framework is an effective approach to preserving endangered languages.

5. Conclusion

This paper argued in favor of the conceptual separation of language documentation and documentary linguistics from language description and descriptive linguistics. It underscored the need for a disciplinary distinction between the two partially overlapping frameworks and demonstrated, perceptibly, the concomitant relationship such a distinction has with the preservation of endangered languages. To this end, this paper bears implications for language documentation, preservation and revitalization.

Finally, it is hoped that this study would engender further research into language documentation, in general, and endangered languages, in particular, in order that the views expressed in this paper would be corroborated or repudiated. Indeed, one of the major theoretical challenges for documentary linguistics is the task of synthesizing a coherent framework for language documentation that takes into account all the various (sub) disciplines that inform a language documentation. Besides, documentary linguistics is not some kind of theory-free enterprise; instead, it is underpinned by a broad variety of theoretical frameworks and requires a theoretical discourse concerned with conceptual and procedural issues in language documentation.

Moreover, there is the need for a format of language documentation. Future studies may want to explore these issues.

References

- Akinlabi, A. & Connell, B. (2008). The interaction of linguistic theory, linguistic description and linguistic documentation. *Studies in the Linguistic Sciences*, 19 (2): 3-39.
- Bird, S. & Simons, G. (2002). Seven dimensions of portability for language documentation and description. Proceedings of the Workshop on Portability Issues in Human Language Technologies, Third Internal Conference on Language Resources and Evaluation. http://arxiv.org/abs/cs/0204020.
- Borsley, R.D. (ed.). (2005). Data in theoretical linguistics. Special issue of *Lingua*, 115: 1475-1665.
- Dryer, M.S. (2006). Descriptive theories and basic linguistic theory. In: F. Ameka, A. Dench, and N. Evans (eds.), *Catching Language: Issues in Grammar Writing*. Berlin, Mouton de Gruyter: 207-234.
- Evans, N. (2008). Review of Gippert et al' 2006. *Language Documentation and Conservation*, 2: 340-350.
- Gippert, J., Himmelmann, N.P.. Ulrike, M. (eds.). (2006). *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- Himmelmann, N.P. (1998). Documentary and descriptive linguistics. *Linguistica*, 36: 161-195.
- Himmelmann, N.P. (2004). Documentary and descriptive linguistics. In: O. Sakiyama and F. Endo (eds.), *Lectures on Endangered Languages 5*, Tokyo and Kyoto Conferences 2002, 37-83.
- Himmelmann, N.P. (2006). Linguistic types and the interface between language documentation and description. *Language Documentation and Conservation*, 6 (2012): 187-207.
- Hyman, L.M. (2001). Field work as a state of mind. In: P. Newman and M. Ratliff (eds.), *Linguistic Fieldwork*, 15-35. Cambridge: Cambridge University Press.

- Hyman, L.M. (2003). Why describe African languages? In: A. Akinlabi and O. Adesola (eds.), *Proceedings of 4th World Congress of African Linguistics*, New Brunswick, 2003. Rudiger Koppe Verlag.
- Kenneth, H., Craig, C., England, N., Jeanne, L., Krauss, M., Watahomigic, L. & Yamamoto, A. (1992). Endangered languages. *Language*, 68 (1): 1-42.
- Lehmann, C. (2001). Language documentation as a program. In: W. Bislang (ed.), *Aspects of Typology and Universals*. Berlin: Akademie Verlag.
- Lehmann, C. (2004). Data in linguistics. *The Linguistic Review*, 21: 175-210.
- Pawley, A. (1993). A language which defies description by ordinary means. In: Foley (ed.), *The Role of Theory in Language description*, 87-129. Berlin: Mouton de Gruyter.
- Thieberger, N. (2009). Steps towards a grammar embedded in data. In: P. Epps and A. Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 389-408. Berlin & New York: Mouton de Gruyter.
- Woodbury, A.C. (2011). Language documentation. In: P.K. Austin and J. Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*. 159-186. Cambridge: Cambridge University Press.