



The Extinction of Minor Languages: The role of technology in the protection and preservation of endangered languages as linguistic heritage in Indonesia

Fanny Henry Tondo¹, Maya Pinkan Warouw², Meilany M. P. Talumewo³, Handrini Ardiyanti⁴, Fachri Helmanto⁵

¹Research Center for Society and Culture - BRIN, ²Sam Ratulangi University, ³Sekolah Tinggi Teologi IKAT, ⁴Research Center for Society and Culture – BRIN, ⁵Universitas Djuanda

Email: ¹fann001@brin.go.id, ²maya_warouw@unsrat.ac.id, ³meilanytalumewo@gmail.com, ⁴hand006@brin.go.id, ⁵fachri.helmanto@unida.ac.id

ABSTRACT

Indonesia's extensive linguistic diversity represents not only a system of communication but also a repository of cultural identity, historical memory, and local knowledge. However, this diversity is increasingly threatened by language shift and extinction. Of the 719 languages identified in Indonesia, several have become extinct, while many others are categorized as endangered or dying, indicating an urgent need for more effective preservation strategies. This paper examines the role of modern technology in documenting, protecting, and preserving endangered languages in Indonesia. Using a descriptive literature-based approach, the discussion compares traditional linguistic documentation—commonly limited to printed grammars, dictionaries, written texts, and physical archives—with documentary linguistics supported by digital technology. The analysis shows that technology transforms language documentation from a relatively restricted research product into a lasting, accessible, searchable, and multipurpose record consisting of audio-video data, metadata, transcription, translation, linguistic annotation, and ethnographic commentary. Digital archives, multimedia resources, electronic dictionaries, computer-assisted language learning, web portals, and online communities expand both the accessibility of linguistic data and the participation of speaker communities. The paper concludes that the preservation of endangered languages requires a shift from traditional documentation toward technology-based documentary linguistics, supported by collaboration among researchers, government institutions, universities, and language communities.

Keywords: endangered languages, language documentation, documentary linguistics, digital technology, linguistic heritage, Indonesia.

INTRODUCTION

Language as one of the cultural heritages plays an important role in human life. It functions as a tool of communication so that a person can convey his or her ideas or messages to other people. Besides that, it can play a function of symbolizing cultural reality and identity. In some cases, a certain language can also emerge as a unification language like Indonesian language which is actually derived from Malay language as one of the local

languages in western Indonesia. Meanwhile, in some regions, a language can also function as an interaction language among diverse ethnics such as local Malays which still exist in present Indonesia. Some of the interethnic languages are Jakarta Malay, Jambi Malay, Banjar Malay, Ambon Malay, Manado Malay, Papua Malay, and Kupang Malay.

The diverse condition of Indonesian languages which can also be divided into two main language families, namely Austronesian and Papuan languages, has made many linguists from all over the world coming and conducting various researches in this archipelagic country with various purposes in relation to thesis, dissertation, and other academic intentions. Some researchers have been working with many Austronesian languages whereas the rest works with Papuan languages. However, there are some linguists that have been conducting researches on the contacts between the two language families as well. The researches of languages here, of course, will be very interesting not only on contacts among those languages but also regarding the history of the ethnics which can also be known by analysis the linguistic data. The history could be on migration and other aspects of the ethnics.

In relation to the numbers of languages, Indonesia takes the second rank in the world after Papua New Guinea. The later has presently about 852 languages whereas Indonesia comprises 719 languages (Lewis *et al.*, 2016). It means that Indonesia constitutes a country which does not only have a high biodiversity that is already well-known all over the world but also have a high linguistic diversity. What can be done with the mentioned condition? By having the high linguistic diversity it implicitly means that this country keeps abundant cultural knowledge including local knowledge in many parts the country because language stores not only the history of the people using it but also their cultural values.

However, the high diversity of languages in Indonesia nowadays is undergoing a threat of extinction or in the process of extinction. Many factors can cause the condition occurs, including the late responses to overcome the problem. I argue that this happen because of the activities in related to linguistic researches especially language documentation are still conducted with traditional sight which is not parallel with the rapid extinction process of the languages in the country. This paper aims to explain the role of the application of modern technology in language documentation to protect and preserve the endangered languages.

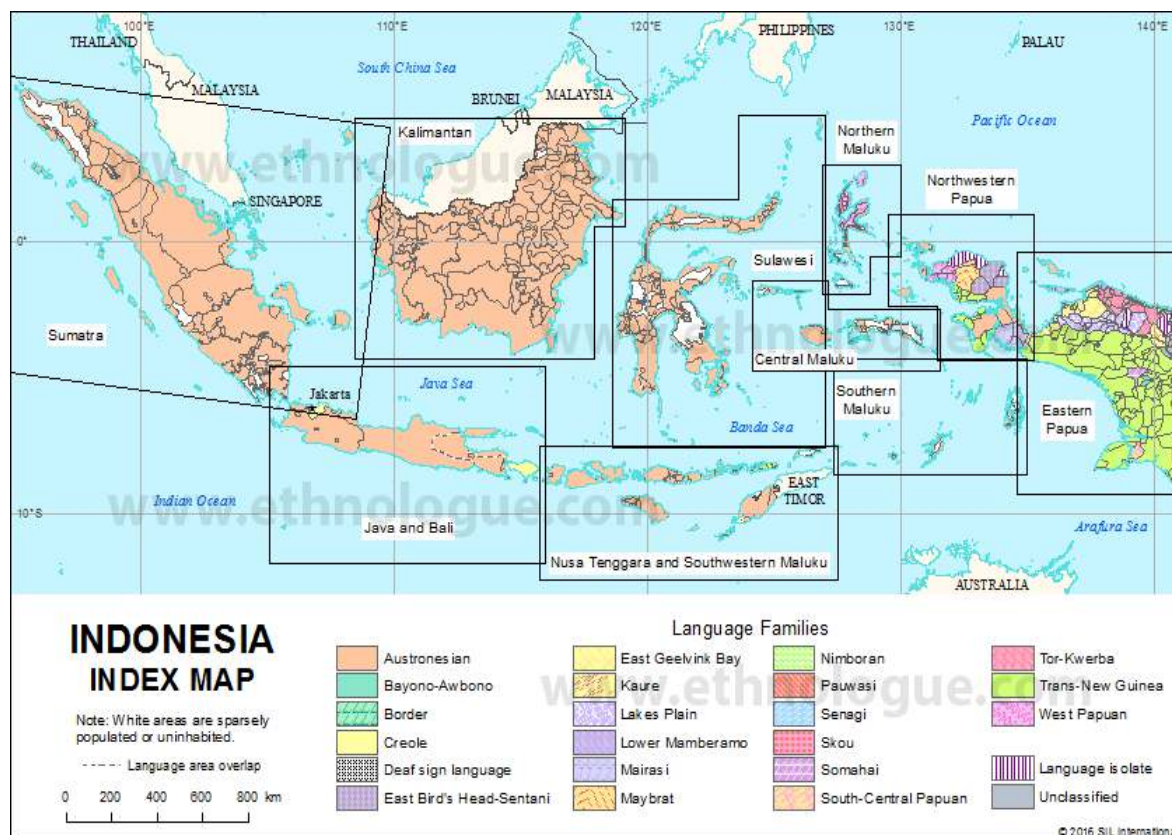
Linguistic Diversity and the Threat of Extinction

It has been mentioned before that there are two primary language families in Indonesia. They are Austronesian and Papuan language families. The first one spreads all over Indonesia particularly in western parts whereas the second one can be found mainly in eastern parts of the country. If we compare the numbers of languages of the world and Indonesia, it is obvious that the country constitutes the abundant source of knowledge especially in relation to linguistic studies with the widely linguistic laboratory. From 7.097 living languages in the world, around 700 languages could be found in Indonesia (Lewis *et al.* 2016).

According to Lewis *et al.* (2016) Indonesia has 719 languages. Of these, 12 are extinct and 76 are dying. Based on that condition, all stakeholders including government, linguists, and language communities should start to think on quick and proper efforts to overcome the problem so that languages could be prevented from extinction or death. Even though a language will come to the extinct situation, at least it has been documented so that it could be

learned by new generations or even just to be studied for academic purposes. The following map shows us the spread of the languages existing in Indonesia.

Map 1: Languages of Indonesia



Source: Lewis *et al.* (2016)

The above map shows the spreading of the languages in the archipelagic country of Indonesia. What can be explained from the map is that Austronesian languages spread almost on all regions of Indonesia whereas the non-Austronesian languages can specifically find in eastern parts of Indonesia only namely in the two Papuan provinces (Papua and West Papua), North Moluccas Province, and East Nusa Tenggara Province. The Austronesian languages are indicated in brown color whereas the non-Austronesian languages are indicated in various colors except brown. Contacts among the languages could certainly bring some impacts to them. Many linguistic phenomena could happen in the diverse condition such as code switching, language shifting, and even language extinction.

A phenomenon that should be given much attention is certainly the extinction. According to the SIL (2001) data as quoted from Lauder (2006: 368-369), there are about 100 languages have 250 or less speakers which could be categorized as endangered, but probably are seriously endangered or even moribund. Wurm 1998, as quoted in Crystal (2000: 21) classify five-level on endangered languages: (1) *Potentially endangered languages*: Languages coming under significant pressure from the majority language and whose speakers form a social and economic minority; the younger generation of speakers has begun to switch to the majority language; (2) *Endangered languages*: Languages with no younger generation mother-tongue speakers; (3) *Seriously endangered languages*: Languages with speakers above the age of 50; (4) *Moribund languages*: Languages whose only speakers are above the age of 70; and (5) *Extinct languages*: Languages with no living speakers or languages with only one living speaker.

The classification of endangered languages is made to make easier to measure the endangerment of a certain language so that we can know the way or steps to overcome the condition. To care a certain endangered language has certainly some reasons. Crystal (2000: 32) try to explain the reasons why we should care if a language dies. He gave five reasons. *Firstly*, we need diversity. The arguments that pertain to biological diversity also pertain to linguistic diversity. Monocultures are not sustainable and this applies in dealing with problems such as development, democracy, and the reduction of poverty. *Secondly*, languages express identity. Thinking people care about who they are, about what makes them the way they are. A language is a mark of community's existence and uniqueness. *Thirdly*, languages are repositories of history. The written records of a language allow us to trace its origins, or ancestry. Written language is an archive of a community's past. *Fourthly*, languages contribute to the sum of human knowledge. Languages uniquely encapsulate and interpret human existence. Each language is a unique mosaic of the ways that its speakers see the world and live in it. *Fifthly*, languages are interesting in themselves. The study of language is the subject matter of linguistics, a branch of human knowledge in its own right. Each language encapsulates a fresh combination of sounds, grammar and vocabulary to form a communication system.

Based on the explanation on the significances of caring a language, we, therefore, should plan to document the languages as soon as possible with the right way. It means we should leave the traditional way of documenting but start to use the modern one.

Traditional Method in Linguistic Research

Previous efforts to study a language usually use traditional method. Regarding archives, for instance, they are traditionally focused on storing physical objects – be it sculptures, artifacts used in daily life, or information engraved in clay tablets or on “old” paper. In other words, not many people are allowed to have access to these physical objects, and creating copies – if it all possible – is an expensive and time-consuming process. It also need special environment in order to meet preservation goals (Trilsbeek and Wittenburg 2006: 312).

In modern digital archives, it is the information on the carrier that we have to preserve independently, whether it represents texts, sounds, videos, three-dimensional representations of artifacts, etc. However, the digital archives have to meet some requirements (Trilsbeek and Wittenburg 2006: 312-313). Those are a) the objects stored in the archive can be subject to change, b) users of the archive may want to add information about resources based on their specific expertise, and c) the objects in the archive need to be accessible and searchable for different purposes.

Many forms of data are not digital, such as an audio-recording on a cassette tape. Traditional written and printed content on paper can thus be regarded as in a sense ‘digital’ (Nathan 2011: 256). What is happening in Indonesia nowadays is that the usage of traditional tools is still an ongoing phenomenon. Many linguists are still using traditional tool of archiving language data and is not parallel with the rapid process of language extinction especially in eastern Indonesia.

Meanwhile, in traditional paradigm, linguists have equated documentation with traditional products of linguistic description, namely a grammar, a dictionary, and a set of texts (Woodbury 2003: 35). Besides that, most of the audiences are the linguists or language researchers.

Documentary Linguistics and the Use of Technology for Language Documentation



What is a language documentation? It is a lasting, multipurpose record of a language. It can also be defined as a field of linguistic inquiry and practice in its own right which is primarily concerned with the compilation and preservation of linguistic primary data and interfaces between primary data and various types of analysis based on these data (Himmelman 2006: 1). The primary data as the core of a language documentation comprises of audio or video recordings. Woodbury (2011: 159) explains that language documentation is the creation, annotation, preservation, and dissemination of transparent records of a language.

Meanwhile, Himmelman (2006: 5) states that there are essentially three reasons for engaging in language documentation. Firstly, the fact that a substantial number of the languages still spoken today are threatened by extinction. In this sense, language documentations are not only seen as data repositories for scientific inquires, but also as important resources for supporting language maintenance. Secondly, to establish open archives for primary data. In language documentation for making analyses accountable.

The basic format for language documentation consists of primary data and apparatus. (Himmelman 2006: 7-14) which can be summarized in the following table.

Table 1. Basic format of a language documentation

<i>Primary data</i>	<i>Apparatus</i>	
Recordings/records of observable linguistic behavior and metalinguistic knowledge (possible basic formats: session and lexical database)	Per session	For documentation as a whole
	Metadata - time and location of recording - participants - recording team - recording equipment - content descriptors Annotations - transcription - translation - further linguistic and ethnographic glossing and commentary	Metadata - location of documented community - project team(s) contributing in documentation - participants in documentation - acknowledgements General access resources - introduction - orthographical conventions - ethnographic sketch - sketch grammar - glossing conventions - indices - links to other resources

Source: Himmelman (2006: 14)

Peter K. Austin and Lenore A. Grenoble (2007: 12) stated that documentary linguistics is a newly emerging field 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” (Gippert, Himmelman and Mosel 2006: v). Documentary linguistics has developed over the last decade in large part in response to urgent need to make an enduring record of the world’s many endangered languages and to support speakers of these languages in their desire to maintain them (Whalen 2003, Austin 2007). It is

fueled by development in information, communication, and media technologies which make documentation and the preservation and dissemination of language data possible in ways which could not previously be envisioned. In addition it essentially also concerns itself with the roles of language speakers in documentary projects and their rights and needs in ways not previously considered within linguistics (see Thieberger and Musgrave, this volume).

Himmelman (2006: 15) identifies the following as important new features of documentary linguistics: (a) *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; (b) *Explicit concern for accountability* – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected; (c) *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 into the distant future; (d) *Work in interdisciplinary teams* – documentation requires input and expertise from a range of disciplines and is not restricted to linguistics alone; (e) *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.

Language documentation refers to the activities carried out by researchers and communities engaged in work that adopt a documentary linguistic approach. The historical genesis of the field of documentary linguistics has meant that the term ‘language documentation’ is sometimes used loosely to refer to any kind of language record, but documentary linguistics uses it in a more specific way, to refer to an activity with much larger and more specific goals. Holton (2011: 375-389) mentioned some types of products which may support endangered languages, namely multimedia, computer-assisted language learning, electronic dictionaries, web portals, and making use of commercial off-the-shelf tools. Besides that, online technologies such as email listservs could also foster communities of language learners and language users.

Today, there are many researchers in Indonesia starts to use new paradigm by applying the documentary linguistics in their language documentation. One of foreign researchers which conduct research by applying technology is Brandon Yoder (2012). In his research in Enggano island, Bengkulu, Sumatera, he introduced two sets of audio recordings of language data in Enggano, a language spoken off the southwestern coast of Sumatera in Indonesia. The words recorded in his research are in sentence frames for acoustic study of the vowels, consonants and vocoid sequences in Enggano. LIPI’ language team also is in the process of application of the paradigm. It is started from the researches on minor language and ethnics in eastern parts of Indonesia. Meanwhile, some universities have also been conducting such a research paradigm. For instance, Atmajaya Catholic Univesity in collaboration with Max Planck Institute.

CLOSING REMARKS

The reality of Indonesian languages which are in the process of extinction should actually make various sectors or stakeholders aware and try to plan and conduct all endeavors to avoid the worst result with the dying languages. Government, academic sector, and language communities should work hand in hand to overcome the problem of language extinction. We have to be aware of the importance of saving a dying language.

In present Indonesia condition with high linguistic diversity, what can be done to protect and preserve from the endangerment conditions? The best solution for the situation is



certainly the application of modern technology to document the languages. All researchers from various institutions should start to shift from traditional method to the modern one. That is the application of documentary linguistics.

REFERENCES

- Austin, P. K. (2006). Data and language documentation. In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation* (pp. 87–112). Mouton de Gruyter.
- Austin, P. K., & Grenoble, L. A. (2007). Current trends in language documentation. *Language Documentation and Description*, 4, 12–25.
- Crystal, D. (2000). *Language death*. Cambridge University Press.
- Himmelmann, N. P. (2006). Language documentation: What is it and what is it good for? In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation*. Mouton de Gruyter.
- Himmelmann, N. P. (2008). Reproduction and preservation of linguistic knowledge: Linguistics' response to language endangerment. *Annual Review of Anthropology*, 37, 337–350.
- Holton, G. (2011). The role of information technology in supporting minority and endangered languages. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge handbook of endangered languages*. Cambridge University Press.
- Lauder, M. R. M. T., & Ayatrohaedi. (2006). The distribution of Austronesian and non-Austronesian languages in Indonesia: Evidence and issues. In T. Simanjuntak, I. H. E. Pojoh, & M. Hisyam (Eds.), *Austronesian diaspora and the ethnogeneses of people in the Indonesian Archipelago: Proceedings of the International Symposium*. LIPI Press.
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2016). *Ethnologue: Languages of the world* (19th ed.). SIL International. <http://www.ethnologue.com>
- Nathan, D. J. (2006). Thick interfaces: Mobilizing language documentation with multimedia. In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation*. Mouton de Gruyter.
- Nathan, D. (2011). Digital archiving. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge handbook of endangered languages*. Cambridge University Press.
- Trilsbeek, P., & Wittenburg, P. (2006). Archiving challenges. In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation*. Mouton de Gruyter.
- Woodbury, A. C. (2011). Language documentation. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge handbook of endangered languages*. Cambridge University Press.
- Yoder, B. E. (2012). *Introduction to the Enggano audio recordings*. Department of Linguistics, University of North Dakota.