

LINGUISTIC RESEARCH: METHODS AND MOTIVES

James T. Collins
Universiti Kebangsaan Malaysia
jtcukm@yahoo.com

Abstract

Before the late nineteenth century, linguistic research relied heavily on dictionaries, but by the end of the century two European scholars, G. Wenker and J. Gillieron, revolutionized linguistic methodology with two very different approaches to data collection in Europe. One hundred years later, just a few years before the founding of *Masyarakat Linguistik Indonesia*, two other European linguists, G. Diffloth and B. Nothofer, began their fieldwork projects in Island Southeast Asia. Their methodologies—connected yet differing in some details—have been explained in print and online sources. This essay focuses on linguistic fieldwork and research methodology based on empirical and inductive principles by reviewing the methods of Diffloth and Nothofer and comparing their approaches to those procedures used in a few recent linguistic studies of local languages in Malaysia and Indonesia. The underlying importance of studying the social history and the geography of the language communities involved in research projects as well as the need for cooperative interaction between researchers and language informants are the only basis for credible linguistic analysis. As hundreds of regional languages in Southeast Asia are threatened with obsolescence, linguists must not only document these languages but explore modes of cooperation with communities seeking to maintain their heritage languages from generation to generation. Computer programs such as ELAN can be used not only to document endangered languages but to revitalize them as well. As linguists, we need to rethink our procedures and enhance our interactions with the language communities of these islands.

Keywords: linguistic methodology, motives, fieldwork approaches

Abstrak

Sebelum akhir abad ke-19, penelitian linguistik sangat tergantung pada kamus. Tetapi pada akhir abad ke-19, dua linguis Eropa, G. Wenker dan J. Gillieron, merombak metodologi linguistik dengan dua pendekatan pengambilan data yang sangat berbeda di Eropa. Seratus tahun kemudian, beberapa tahun sebelum Masyarakat Linguistik Indonesia dibentuk, dua orang linguis Eropa lainnya, G. Diffloth dan B. Nothofer, mengawali proyek penelitian mereka di Kepulauan Asia Tenggara. Metodologi mereka – saling terkait meskipun berbeda dalam beberapa hal – telah dijelaskan baik secara tertulis maupun lisan secara daring. Tulisan ini berfokus pada penelitian linguistik lapangan dan metodologi penelitian yang berbasis pada prinsip-prinsip empiris dan induktif dengan menelaah metode Diffloth dan Nothofer, dan membandingkan pendekatan keduanya dengan prosedur yang digunakan dalam beberapa penelitian linguistik mutakhir di Indonesia dan Malaysia. Satu-satunya basis untuk analisis linguistik yang kredibel adalah pentingnya mempelajari sejarah sosial dan geografi masyarakat penutur bahasa yang terlibat dalam proyek penelitian dan juga perlunya kerja sama yang interaktif antara peneliti dan informan bahasa. Ratusan bahasa daerah di Asia Tenggara terancam punah, oleh karena itu para linguis tidak hanya harus mendokumentasikan bahasa-bahasa tersebut tetapi juga mengeksplorasi berbagai moda kerja sama dengan komunitas bahasa yang berupaya mempertahankan bahasa warisan mereka dari generasi ke generasi. Program komputer seperti ELAN dapat digunakan tidak hanya untuk mendokumentasikan bahasa-bahasa yang terancam punah tetapi juga

merevitalisasi bahasa-bahasa tersebut. Sebagai linguis, kita perlu meninjau ulang prosedur penelitian kita dan juga memperbaiki interaksi kita dengan masyarakat penutur bahasa di pulau-pulau tersebut.

Kata kunci: metodologi linguisitik, motif, pendekatan penelitian lapangan

INTRODUCTION

In the nineteenth century, when the field of modern linguistics had only begun to develop, many European scholars realized that they needed linguistic data that differed from the materials found in dictionary entries. The documentation, analysis and mapping of regional dialects and languages required new approaches and credible methodologies. The names of two scholars who addressed this problem more than 125 years ago have long been prominent in textbooks and university lectures: G. Wenker and J. Gillieron. Wenker (1876-1887) developed a standard list of 40 sentences mailed to school teachers in 40,000 German-speaking regions of Europe. Gillieron (1897-1901), on the other hand, trained E. Edmont, to collect 1900 specific words in 639 French-speaking rural areas.ⁱ These methodologies have shifted, merged, improved and those diverse versions have expanded beyond dialectology to include fieldwork in historical linguistics, phonology, morphology, syntax and sociolinguistics.

All of these contemporary methods of linguistic research are based on two research principles. Research must be both empirical and inductive. Empirical research draws on experience, experimentation and observation, that is evidence gathered from the real world. These real data allow for triangulation to ensure the validity of the assembled data. Having collected the data, the inductive procedure can be followed. Based on the empirical data, we can propose a generalization. A generalization must be tested to lead to a hypothesis. Testing includes reviewing the data by those involved in collecting it but also presenting a generalization to others by discussing results with colleagues, presenting papers at conferences and publishing articles in print or online. Once a hypothesis is reached it also must be tested by presenting it in person, in print and online in order to ensure rigorous evaluation by others so that a theory can emerge. The process of presentation and evaluation must be repeated in order to revise and enhance the hypothesis.

This brief paper has two related components tied to fieldwork and methodology. First, we will look at the research methods of B. Nothofer and G. Diffloth, two linguists who worked extensively in Southeast Asia during the late twentieth century and the first decades of the twenty-first century. Both scholars have written or spoken in detail about their fieldwork methods. Second, based on published and spoken reports, we will look at some of the methods or techniques used in a few recent linguistic fieldwork projects about local languages (*bahasa daerah*) in Indonesia and Malaysia.

In the conclusion, two shifts in 21st century fieldwork approaches are touched upon as well as the rise of language documentation. There is a brief discussion of the need for sustainable initiatives to revitalize Indonesia's heritage languages using modern technologies as well as relying on solid fieldwork that connects with heritage language communities.

DATA AND ANALYSIS: TWO SCHOLARS OF SOUTHEAST ASIAN REGIONAL LANGUAGES

In the late 20th century two linguists, one German like Wenker of the 19th century and the other French like Gillieron of the same century, committed their lives to collecting language data, analyzing those data and advancing linguistics. Bernd Nothofer and Gérard Diffloth, both specializing in historical linguistics, studied local languages in Southeast Asia and both have written and spoken about their research methodologies. As noted above, the nineteenth century methods of Wenker and Gillieron have diverged, merged and emerged as complex and diverse procedures that may appear simple and straightforward to observers. Almost sixty years ago, the book, *Discovering American Dialects* (Shuy, 1967), delineated procedures to **prepare** for fieldwork. These included non-linguistic tasks, for example the selection of the fieldwork area itself, followed by background research about the selected area, including learning about the patterns of settlement history and population shift as well as the physical geography and social structure.

Nothofer's (1980) book, *Dialektgeographische Untersuchungen in West-Java und im westlichen Zentral-Java*, based on his fieldwork in 1971-1972 (Nothofer, 1981: 1),ⁱⁱ focuses on the dialect geography (mapping) of lexical items in the Sundanese language area in West Java, western areas of Central Java and Jakarta areas that border on the Sundanese region. It begins with a summary of earlier research on the topic and then proceeds with a historical overview of the region which includes patterns of population shift and contemporary social structure. The 162 villages and towns selected for data collection are specified. The choice of informants in each fieldwork site was based on specified characteristics, such as age, gender, education and place of birth. The basis of the data collection was 327-word list elicited in Indonesian with clarification, when needed, in Sundanese or Javanese. When there was confusion, sometimes a short discussion about the purpose of the project would take place. A more detailed presentation about data collection can be found in the introductory materials of *Dialektatlas von Zentral-Java*. (Nothofer, 1981: 1-30), a collection of 337 dialect maps of 34 Javanese-speaking villages to the west of Yogyakarta. For example, Nothofer explained the criteria for choosing a village in the project, namely a location distant from a large city or main road, a population of no more than 5000 and a low rate of population shift. The standard for selecting language informants is detailed: a male aged 35-45, whose parents, as well as he, were born and raised in the village, with an education no higher than primary school and whose work was as a farmer or a laborer. Moreover, Nothofer discussed four possible methods of eliciting vocabulary and grammatical forms:

- Ask in the standard dialect of the informant's mother tongue;
- Ask in Indonesian or the regional language of the area;
- Present a paraphrase, for example to elicit the word for **egg**, ask 'What does a chicken lay?', using the informant's mother tongue, Indonesian or the regional language;
- Best of all, when possible, present a picture of the term that is sought.

Of these four methods he considered the first two methods problematic because the term used in the question might act as a stimulus and simply be repeated by the informant because the word belongs to his passive vocabulary or is regarded as the prestige form. Nothofer also suggested strategies for eliciting grammatical morphemes.

The brief descriptions of only two of his books (Nothofer, 1980, 1981) presented here indicates that his research was based in part on the Swadesh 200-word questionnaire and that he

was fully cognizant of the direction of contemporary dialect geography issues, for example the four methods of eliciting responses.ⁱⁱⁱ Moreover, he profited from Teeuw's (1951, 1958) study of Sasak dialects, "the basis for all modern dialect geographic studies" in Indonesia (Nothofer, 1981: 32) and Uhlenbeck's (1964: 64) critical bibliography of earlier dialect studies in Java which were not influenced by the advances in dialect study since Gillieron (Nothofer, 1980: 2). We must remember that in the 1970s Nothofer studied two of Indonesia's regional languages, probably the two languages with largest number of speakers. He estimated that at the time of his research there were 50 million speakers of Javanese on the island of Java alone and 20 million speakers of Sundanese (Nothofer, 1980: 1).^{iv} Because of the geographic distribution of these two 'local' languages (*bahasa daerah*) and the depth of his findings that yielded cartographic depictions of linguistic elements of the Javanese and Sundanese dialects west of Yogyakarta, his analyses of the comparative linguistic implications for the history of these languages is unmatched.

In 1965, Diffloth completed his doctoral research about Irula, a Dravidic language spoken in the Nilgiri mountains southwest of Chennai, India. Then, in 1966 he arrived in Malaysia to present a paper at the First International Tamil Conference in Kuala Lumpur (Diffloth, 1968: viii). Apparently, it was in 1966 that he began his research on the Austroasiatic languages spoken in west Malaysia, such as Semai, Temiar, Cheq Wong, Jah-Hut and Semelai (Diffloth, 1976: 73).^v Eventually, his commitment to the study of the Austroasiatic languages expanded to include almost all the Austroasiatic languages spoken in Southeast Asia. While Nothofer's principal work focused on languages with tens of millions of speakers, most of the Austroasiatic languages that Diffloth studied were spoken by small communities, for example Semai, the largest Austroasiatic group in Malaysia with only 16,000 speakers (Diffloth, 1977), Jah-Hut another Austroasiatic language of Malaysia with perhaps 1300-1600 speakers (Diffloth 1976) and Nyah Kur, an Austroasiatic language spoken in central Thailand by 1600 people (Joshua Project 2025). Nonetheless, like Nothofer, Diffloth was committed to the study and mapping of the dialects of each of these languages and the classification of these languages based on the historical comparative method. For example, in a nine-year period, he had collected data in 117 Semai settlements in Malaysia and was able to identify and map 42 dialect areas to gain "a picture of the linguistic system, called Proto-Semai..." (Diffloth, 1977: 264). In his detailed study of Nyah Kur, he presented lexical data from eleven villages which indicated that the variants spoken in those villages belonged to three different dialect areas, all closely related to the Mon language (Diffloth, 1984: 33-34).^{vi}

Most of these Nyah Kur communities of Thailand had only been recently resettled from other areas; usually these new villages included Thai-language primary schools and Thai-speaking villagers. He observed that in some villages the children of Nyah Kur families spoke Thai when they played with each other. In other villages grandparents spoke Nyah Kur with people their age but only spoke Thai to their children and grandchildren. As noted above, typically in language or dialect surveys, persons born and raised in the target village are selected as informants; moreover, Nothofer's criteria specified that informants should be persons born in the village whose parents were also born in the village. But because many of the Nyah Kur villages were new settlements of perhaps only 25 years, no such requirement could be followed. Moreover, in contrast to the informants (men aged 35-45 years) that Nothofer sought, most of the informants in Diffloth's Nyah Kur project were in their 70s or even older, although in one village he did work with a husband and wife in their 50s. In a large village with several hundred Nyah Kur speakers two "older women" were chosen for the survey. In other villages, the informants were "a former

headman (79)”, “a man in his mid-seventies”, “an old woman 89 years old”, “a man, seventy-five”, “four older women and a few men” (Diffloth 1984: 34-40). Clearly, strategies for collecting data in villages where the language is on the verge of extinction differ from strategies for collecting data in villages connected to fifty million other speakers.

In addition to differences in assembling informants for the corpus of a “language before it disappears forever” (Diffloth 1984: 40), apparently there were other differences between Nothofer’s and Diffloth’s methods. First, Nothofer’s and Diffloth’s data both included more than Swadesh’s 200 words. For example, Nothofer’s (1981: 16) wordlist included at least 50 animal names. Although there is no explicit list of the items in Diffloth’s questionnaire, in his book about the Nyah Kur and Old Mon languages, he proposed a comparative lexicon—a lexicon that included nineteen names of mammals, ranging from tree shrews and bats to muntjacs and macaques (Diffloth 1984:61-68); in total there were 52 animal names. But, in addition to a shared interest in animal names, we know that Diffloth was willing to veer from a specified wordlist. One of his colleagues, N. Badenoch, witnessed a session with the “last speaker” of Kuay, an Austroasiatic language spoken in northwestern Cambodia. This elderly woman “was not responding to basic questions like ‘how do you say eat rice?’ But when Gérard asked her about catching elephants, she sat up and spoke her Kuay variety clearly and lucidly for him for an hour and half” (Badenoch, 2024: 4).^{vii}

Indeed, as a linguist, Diffloth placed the highest value on the **interaction** between the linguist and the informant, as he made explicit in an interview with Badenoch in 2014 about his perspective of fieldwork methods, for example tape recording oral data.

“Actually, the term ‘recording’ is ambiguous. In its superficial sense, it implies that we take along some high-tech equipment, make hundreds of hours of excellent recordings and then go home to study them, or store them for the future. But unless you have been constantly interacting, on a word-by-word basis, with the very person who spoke, who told the stories, who sang the songs, such recordings are practically worthless. Recordings, in the deeper sense of the term, are witnesses of detailed and elaborate agreements” (Diffloth, 2014).

The critical importance of seeking and maintaining rapport through “agreements” with one’s language informants is the basis for meaningful, in-depth language research.

This feature of fieldwork methods is implied as well by Nothofer’s description of data collection sessions in central Java.

Die Anzahl der Gewährsleute betrug für ein Interview drei bis vier Personen. Einer der Informanten wurde als Hauptrespondent ausgewählt (Nothofer, 1981: 6).^{viii}

The focus of Nothofer’s data collection even in the study of languages spoken by millions of persons remained the response of the chief informant. There are perhaps differences between the methods practised by these two European scholars^{ix} but the steady interaction of the fieldworker and his informant is at the core of excellence in their research methodology.

ENDANGERED LANGUAGES AND FIELDWORK IN THE ARCHIPELAGO: 21st CENTURY

A few years ago, an article was published with the title “*Bahasa Melayu Negeri Sembilan dan Isu Imperialisme Linguistik*” (Idris Aman et al., 2018). I was interested because in 1968 the first

Malay village I visited was in Negeri Sembilan and because I assumed that the article was about American imperialism, as usual and all the more so now. But I was surprised to read that the authors' intention was not actually about American imperialism but about refuting the idea that Negeri Sembilan Malay is Minangkabau, as Winstedt and modern authors (Malaysian and Indonesian) have stated, an idea perhaps caused by Minangkabau domination (imperialism?), that is "*akibat pengaruh budaya Minangkabau yang dominan*". The authors of this article stated that they had collected modified 100-word Swadesh wordlists only(!) in Malaysia from 15 informants who spoke Negeri Sembilan Malay and from seven Malaysians who spoke Minangkabau as well as from eight Indonesians who came from Padang, Bukit Tinggi and Payakumbuh and resided in Malaysia.

It is not clear how the authors selected their informants or how they elicited the lexical items (see Nothofer's four methods) but they wrote that each session lasted only 15-20 minutes! How did they ask 100 items in just 15 minutes? Did they simply hand out a print copy of the 100 words in Malay and wait for a response? Furthermore, the article does not explain how they prepared for their project. Did they conduct background research? For example, the region around Payakumbuh in Sumatra Barat is well known for its unique dialectal features, but there is no hint of that dialect complexity in their essay, as if the vast and mountainous province of Sumatra Barat speaks a uniform Minangkabau language with no dialectal variation. Moreover, in Malaysia between Alor Gajah in the state of Melaka and the border between Melaka and the state of Negeri Sembilan, there are numerous villages where Negeri Sembilan Malay is spoken as the first language. This region was only incorporated into Melaka after a long, armed struggle with the British in the early 19th century. Perhaps, the authors did not know about the dialect boundaries, only the political one; so, they limited their data collection to today's Negeri Sembilan with borders determined long ago by the imperial British! Both the questionable method of data collection in a mere 15 minutes and the failure to study the social and sociolinguistic background of the research areas do not lend any credence to the specious conclusions in Idris Aman et al. (2018).

Although Idris et al. (2018) cite Chambers and Trudgill's (1980) book, they did not mention the issue of heteronomy and autonomy, an early topic in the book (Chambers and Trudgill 1980:9). For example, the dialects of eastern Netherlands share the same basic features as the dialects spoken in the bordering regions of Germany, but the local dialects in one country are heteronymous to standard Dutch and in the other country they are heteronymous to standard German—to use Chambers and Trudgill's terminology.^x This is the result of international political boundaries. This notion of heteronymy seems to have played a role in insisting that the Malay dialect of Negeri Sembilan, Malaysia, cannot be related to Minangkabau in Indonesia. Similarly, the study and classification of local languages in Indonesia sometimes reflects this notion of heteronymy. However, the Indonesian examples for discussion below involve provincial borders, not international ones.

At the southeastern tip of the province of Sulawesi Tenggara, the Tukang Besi islands mark the entrance to the Flores Sea to the south. See Map 1.



Map 1. The Tukang Besi Islands of Sulawesi Tenggara, bordering the Flores Sea

On three of these islands, the Binongku language is the indigenous language spoken by the local population (Collins 2020). Three hundred kilometers southwest of the Tukang Besi islands, in the province of Sulawesi Selatan, lies Bonerate, one of the southeastern islands of that province, where the local language is known as “*bahasa Bonerate*” (Bakkers 1862). However, more than one hundred years ago, Elbert (1911-1912) identified the Bonerate language as a language of the Tukang Besi islands.^{xi} Moreover, to the southeast of Bonerate, lies the island of Pamana Besar off the north coast of Flores island in the province of Nusa Tenggara Timur. A large community of Binongku speakers live on that island.^{xii} In the opposite direction, 500 kilometers to the northeast of the Tukang Besi islands, for at least one hundred years, the Binongko language has been the home language of the inhabitants of two small islands off the village of Asilulu on the westernmost cape of Ambon island in the province of Maluku (Collins 2020).^{xiii}

Nonetheless, based on the material in Mahsun et al. (2008: 75, 89), the language data collected by language offices (*Balai Bahasa*) in at least two provinces of Sulawesi were not compared with each other (Collins, 2020). So, Bonerate is a language reported by the Sulawesi Selatan office and Tukang Besi (Binongko) is the language noted by the Sulawesi Tenggara office. The Binongko language communities of Nusa Tenggara Timur and Maluku were not even noted in the reports of their respective provincial language offices. This apparent lack of communication perhaps reflects a sort of heteronymy based on provincial borders! That the language of Bonerate is the same as the of Binongko language was overlooked,^{xiv} an approach apparently based on political boundaries **not** the geographic reality of the Flores Sea, has obscured the distribution of the Binongko language communities in four or five Indonesian provinces.

Even when language data is compared, errors may occur. In the same Pusat Bahasa book (Mahsun et al., 2008), we find a curious paragraph (Figure 1).

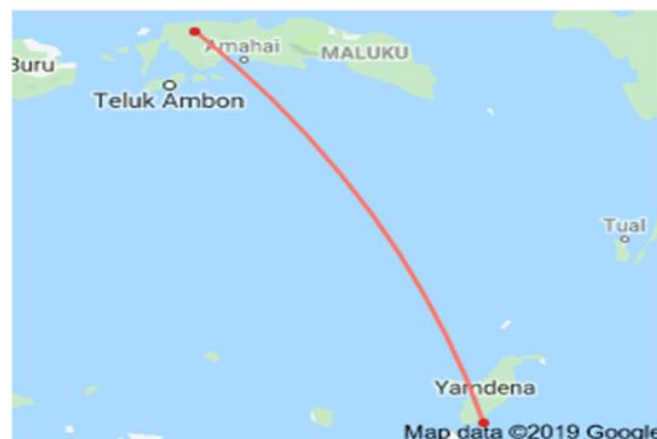
2.7.1.49 Bahasa Yamdena

Bahasa Yamdena dituturkan oleh masyarakat di (1) Desa Woroloin, Taniwel, Hulung, Uweh, dan Banggoi, Kabupaten Seram Bagian Barat dan (2) Desa Olilit Raya, Kabupaten Maluku Tenggara Barat, Provinsi Maluku. Bahasa ini terdiri atas enam dialek, yaitu (1) dialek Wemale yang dituturkan di Desa Woroloin, Kecamatan Taniwel Timur, Kabupaten Seram Bagian Barat; (2) dialek Nakaela dituturkan di Desa Taniwel, Kecamatan Taniwel, Kabupaten Seram Bagian Barat; (3) dialek Hulung dituturkan di Desa Hulung, Kecamatan Taniwel, Kabupaten Seram Bagian Barat; (4) dialek **Alune** dituturkan di Desa Uweh, Kecamatan Taniwel, Kabupaten Seram Bagian Barat; (5) dialek Banggoi dituturkan di Desa Banggoi, Kecamatan Bula Barat, Kabupaten Seram Bagian Timur; (6) dialek Yamdena Olilit Lama yang dituturkan di Desa Olilit Raya, Kecamatan Tanimbar Selatan, Kabupaten Maluku Tenggara Barat. Persentase antardialek tersebut berkisar 60–80%. Berdasarkan hasil penghitungan dialektometri, isolek Yamdena merupakan sebuah bahasa karena persentase perbedaannya dengan bahasa lain di Maluku berkisar 81%–100%, misalnya dengan bahasa Kei, Fordata (Iyaru), dan Banda.

Figure 1. Six “dialects” of the Yamdena language (Mahsun et al. 2008: 131-132)

Why are five languages spoken by former mountain dwellers (not seafarers!) in a few villages in northern Seram considered “dialects” of a language spoken by 25,000 villagers on the east coast of Tanimbar island? The point-to-point distance between Taniwel (Seram) where four of the “dialects” are spoken and Saumlaki, the capital of the Tanimbar regency is 643 kilometers; refer to Map 2.

643 km Taniwel-Saumlaki



Map 2. The distance between Taniwel (Seram) and Saumlaki (Tanimbar)

The *fatwa*-like book, *Bahasa dan peta bahasa di Indonesia*, published by Pusat Bahasa (Mahsun et al. 2008), provides no data. However, an article published by one of the Pusat Bahasa staff (Mukhamdanah, 2015) who apparently collected the data for Figure 1 provides information about the analysis of the data^{xv} and more importantly the intentional denial of **autonomy** to the speakers who provided the data. As noted in Collins (2019):

“The speakers of the languages in the Taniwel kecamatan, i.e. Alune, Wemale, Hulung and Naka’ela, told the Pusat Bahasa staff that they spoke different languages. ... But indigenous knowledge (*kebijakan lokal!*), which coincides with the findings of historical linguistics (Collins 1983) as well as lexicostatistics (Ethnologue 2014), meant nothing to the project managers. The calculations and simplifications in Pusat Bahasa’s Jakarta office superseded everything else.”

By ignoring the informants' insistence that their languages were not the same as well as the results of earlier research (summarized in Collins (2018: 73-79)), thereby denying their **autonomy**, the editors produced a book of dubious value.^{xvi}

CONCLUSION

The 21st century has yielded many innovations in data collection. Do the methods of collecting and storing data now used in insular Southeast Asia reflect these innovations and improvements? Two or three decades ago, to collect audial data, a few cassette recorders and a supply of cassette tapes and batteries might be all the equipment needed. Today a budget in a typical research proposal to a foreign funding agency supporting anthropological or linguistic research for a short-term project in Indonesia would include not only travel costs and allowances for the research team but also the cost of equipment for audio-video recordings as well as payments to informants. In some cases, these reflect a changing focus in collection and purpose. A budget of US \$10,000 for audio-visual equipment and a mere US\$500 distributed among dozens of informants would not be unusual.

In view of the enhanced role that audial and video recording now plays in fieldwork methodology and the subsequent use of the materials for analysis, it is not surprising that at the very beginning of this century, the Language Archive released ELAN,^{xvii} a computer system designed for transcribing and annotating audial video recordings. There is a range of modes of annotation, such as layering multiple languages.^{xviii} In addition to its usefulness in storing empirical data, by analyzing that stored, annotated data, a user can propose a generalization that can be published or presented to test its credibility. The advance of language description and linguistics should be the purpose and motive of language data collection. Unfortunately, as Diffloth (2014) pointed out a decade ago, many of these audial and video recordings are stored and face an uncertain future. Indeed, after publishing a brief article in a pay-if-you-can journal, linguists often move on to a new research project, wherever the funding leads them. The demands of an academic career now have superseded the scholarly task of proposing generalizations and hypotheses that can be tested by others.

Nonetheless, although many linguists see ELAN only as a site to store materials for the sake of language documentation, this computer system can be organized as a language revitalization resource. Language revitalization is often mentioned by linguists but usually suggested by them as a task for local language communities. However, linguists can encourage local language communities to launch language revitalization programs by using ELAN, for example, to organize and format video clips of the communities' local languages in everyday use (building houses, planting cassava and so forth), based on data collected through intergenerational interaction (Collins 2021b). Linguists can assist local language communities to produce "products" in their languages aimed at specific consumers. For example, they can make video games, animations or live action films for primary school students or set up blogging communities for adults so members can post and comment on blogs, and join discussion forums^{xix}

Some writers estimate that there are 706 heritage ("local") languages spoken in Indonesia. Many of these languages are endangered languages spoken only by the oldest members of small communities, especially in eastern Indonesia. In the provinces of the Maluku and Papua regions alone, 384 that is 54% of Indonesia's 706 heritage languages are spoken by a diminishing number of elderly villagers. Languages we may have studied by collecting brief wordlists for a limited

project may now be close to extinction (Collins 2021a). We need to act now--not just to document but also to join in the revitalization of Indonesia's heritage languages.

NOTE

This paper was presented at the *Kongres Internasional Masyarakat Linguistik Indonesia* (KIMLI), held at the University of Indonesia on July 16–18, 2025.

REFERENCES

- Aman, I.; Awal, N. M.; & Jaafar, M. F. (2018). Bahasa Melayu Negeri Sembilan dan isu imperialisme linguistik. *International Journal of the Malay World and Civilisation* 6(1), 11 – 26.
- Anceaux, J. C. (1958). Languages of the Bomberai peninsula. *Nieuw Guinea Studiën* 2:109-121.
- Badenoch, N. (2024). In memoriam: Gérard Diffloth (1939-2023) *Austroasiatic linguistics in honour of Gérard Diffloth (1939-2023)*, pp. 1-5, P. Sidwell (ed.). Chiang Mai University.
- Bakkers, J. A. (1862). De eilanden Bonerate en Kalao. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 11, 215-264.
- Bloomfield, L. (1935). *Language*. Allen & Unwin.
- Chambers, J. K. and P. Trudgill. (1984). *Dialectology*. Cambridge: Cambridge University Press.
- Collins, J. T. (1983). *The historical relationships of the languages of Central Maluku, Indonesia*. Pacific Linguistics D47. The Australian National University.
- Collins, J. T. (2018). *Penelitian bahasa di Maluku*. Ambon: Kantor Bahasa. Kementerian Pendidikan dan Kebudayaan.
- Collins, J. T. (2019). The Naka'ela language: Studying and appreciating the story of an extinct language. Paper presented at Universitas Haluoleo, Kendari, 15 November 2019.
- Collins, J. T. (2020). Krisis di Indonesia Timur: Fokus revitalisasi bahasa daerah. In *Prosiding Kongres Internasional III Bahasa-Bahasa Daerah di Sulawesi Tenggara: Peran bahasa dan sastra daerah sebagai negosiasi budaya dalam menciptakan perdamaian bangsa*, Sandra Safitri Hanan (Ed.), pp. 18-31. Kendari: Kantor Bahasa Sulawesi Tenggara.
- Collins, J. T. (2021a). "Culture is a shadow", language as a shade: Fragments of a dead language, Naka'ela. *Wacana. Jurnal Ilmu Pengetahuan Budaya* 22, 152-181.
- Collins, J. T. (2021b). Digitalisasi dan revitalisasi bahasa daerah: Pengalaman dua proyek di Indonesia. Invited presentation at Revitalisasi Bahasa Daerah di Era Digital, Program Studi Magister Linguistik, Universitas Sam Ratulangi, 29 Oktober 2021.
- Collins, J. T. (2024). The role of education in developing students' ethnic character. Invited presentation at International Seminar, Transforming Education for Building Students' Character Quality in the Era of Society 5.0, Universitas Muhamadiyah Gorontalo, 24 October 2023.
- Diffloth, G. (1968). The Irula language, a close relative of Tamil. Dissertation, University of California at Los Angeles. <https://linguistics.ucla.edu/general/dissertations/Diffloth.1968.pdf>
- Diffloth, G. (1976). Jah-Hut, an Austroasiatic language of Malaysia. In *South-East Asian Linguistic Studies, Pacific Linguistics* C, 42, 73-118. Australian National University.

- Diffloth, G. (1977). Towards a history of Mon-Khmer: Proto-Semai vowels. *Tônan Ajia Kenkyû (Southeast Asian Studies)* 14(4), 463-95.
- Diffloth, G. (1984). *The Dvaravati Old Mon language and Nyah Kur*. Chulalongkorn University Printing Press.
- Diffloth, G. (2014). Language and history in Southeast Asia. With N. Badenoch. *CSEAS Newsletter* No. 69. Kyoto: Center for Southeast Asian Studies. <https://www-archive.cseas.kyoto-u.ac.jp/www/2016/2014/05/language-and-history-in-southeast-asia-an-interview-with-gerard-diffloth/>
- Donohue, M. (1999). *A Grammar of Tukang Besi*. Mouton de Gruyter.
- Elbert, J. (1911-1912). *Die Sunda-Expedition des Veriens für Geographie und Statistik zu Frankfurt am Main*. Minjon.
- Ethnologue. Languages of the world*. (2014). <http://www.ethnologue.com/>.
- Lewis, M. Paul (2009). *Ethnologue: Languages of the World*. Sixteenth edition. SIL International.
- Francis, W.N. (1983). *Dialectology: An Introduction*. London: Longman.
- Gilliéron, J. and E. Edmont. (1902-1910). *Atlas linguistique de la France*. Vol. 1-13. Paris: H. Champion.
- Joshua Project. 2(025). Nyahkur in Thailand. https://joshuaproject.net/people_groups/14077/TH.
- Mahsun et al. (Ed.) (2008). *Bahasa dan Peta Bahasa di Indonesia*. Pusat Bahasa. Departemen Pendidikan Nasional.
- Mukhamdanah. (2015). Bahasa Yamdena di Maluku: Korespondensi, variasi dialektal dan sebaran wilayah pakainya. *Widyaparwa* 43(1), 39-54.
- Nothofer, B. (1972a). Parean jeung Lelea. Jumplukan Sunda di wilayah Jawa. *Mangle* 341.13 and 31.
- Nothofer, B. (1972b). Dialek bahasa Sunda dan nasibnya. *Pikiran Rakyat* 31/7/1972.
- Nothofer, B. (1980). *Dialektgeographische Untersuchungen in West-Java und im westlichen Zentral-Java*. 2 Bände. Otto Harrassowitz.
- Nothofer, B. (1981). *Dialektatlas von Zentral-Java*. Otto Harrassowitz.
- Petyt, K. M. (1980). *The Study of Dialect: An Introduction to Dialectology*. André Deutsch.
- Premisrirat, S. (2002). The future of Nyah Kur. *Collected papers on Southeast Asian and Pacific languages*. Edited by R.S. Bauer. *Pacific Linguistics* 530:3-12:155-165. Australian National University, Research School of Pacific and Asian Studies.
- Shuy, R. W. (1967). *Discovering American Dialects*. National Council of Teachers of English.
- Teeuw, A. (1951). *Dialect-atlas van/of Lombok (Indonesia)*. Biro Reproduksi Djawatan Topografi.
- Teeuw, A. (1958). Lombok. Een dialect-geographische studie. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*, XXV). 's-Gravenhage: Martinus Nijhoff.
- Uhlenbeck, E.M. (1964). *A Critical Survey of Studies on the Languages of Java and Madura*. 's-Gravenhage: Martinus Nijhoff.
- Warnk, H. (2006a). Badminton in Brunei and beyond: Some personal notes on Bernd Nothofer. In *Insular Southeast Asia. Linguistic and Cultural Studies in Honour of Bernd Nothofer*, edited by F. Schulze and H. Warnk, pp. xi-xvi. Otto Harrassowitz Verlag.
- Warnk, H. (2006b). Publications of Bernd Nothofer in *Insular Southeast Asia. Linguistic and Cultural Studies in Honour of Bernd Nothofer*, edited by F. Schulze and H. Warnk, pp. xvii-xx. Otto Harrassowitz Verlag.

Wenker, G. (1878). *Sprach-Atlas der Rhein-provinz nordlich der Mosel sowie des Kreises Siegen*. Marburg: Universitätsbibliothek.

ⁱ Chambers and Trudgill (1984:15-21) provide a brief overview of these two pioneering projects. See also Petyt (1980) and Francis (1983).

ⁱⁱ Nothofer (1972a, 1972b) published his earliest (?) articles about Sundanese dialects.

ⁱⁱⁱ See also his citations (Nothofer 1980:261-262) of Meillet (1921) and Bloomfield (1933) each of whom wrote extensively about the methods of dialect geography.

^{iv} Of course, Nothofer subsequently conducted research on many other languages of insular Southeast Asia, including the Enggano language, the Malay dialects of Bangka island, all the indigenous languages of Brunei and some of the local languages of Kalimantan Barat. See Warnk (2006b).

^v As he said in his interview: "... from Tamil to the mountains of Malaysia" (Diffloth 2014).

^{vi} By contrast, the Jah Hut language with a similar number of speakers displayed no significant dialect differences, except in one village. (Diffloth 1976:82-83).

^{vii} Was this shift to elephants caused by Diffloth trying to elicit the word for 'eat'? Note that "The Kuy [Kuay] are known as skilled mahouts, or elephant trainers, and many Kuy villages are employed in finding, taming, and selling elephants" (Elephant database <https://www.elephant.se/index.php?id=199>).

^{viii} Translation: "The number of informants for an interview was three to four persons. One of the informants was chosen as the chief respondent."

^{ix} Both of these leaders in the field studied for the PhD's at well-known American universities: Nothofer at Yale University and Diffloth at the University of California at Los Angeles. Perhaps this "standard" training and course work in the U.S. may have contributed to the similarities in methods despite the differences in the number of speakers of the local languages—a theoretical consonance reflected in similar approaches and priorities?

^x Bloomfield (1933:344) demonstrates the borders of Low and High German but includes the complexities of the Rhenish Fan.

^{xi} More recently see Donohue (1999:8) identifying the Bonerate language.

^{xii} I am grateful to Ibu Veronika Genua at Universitas Flores (Ende) who confirmed reports that I had received about the Binongku community on "Pamana".

^{xiii} Indeed, Anceaux (1958:112) wrote of a large community of Binongku ("Buton") speakers in the town of Fakfak now in the Province of Papua Barat. See Donohue (1999) as well.

^{xiv} Probably when the staff were preparing their research on local languages, they did not refer to earlier studies, including the patterns of settlement history, as Shuy (1967) recommended.

^{xv} "*Keenam dialek ini, menurut pengakuan penutur, merupakan bahasa yang berbeda. Masyarakat atau penutur dialek-dialek tersebut menyebut isolek yang mereka pakai sebagai bahasa Waymale Utara, bahasa Nakaela, bahasa Hulung, bahasa Alune, bahasa Bonggoi, dan bahasa Olilit Lama. Dalam buku Bahasa dan Peta Bahasa di Indonesia (2008:31-32), keenam isolek tersebut bukan merupakan bahasa yang berbeda, tetapi hanya merupakan dialek-dialek yang berbeda atau merupakan variasi dialektal dalam satu bahasa yang sama, yaitu bahasa Yamdena*"

^{xvi} Apparently in Mukhamdanah. (2015), if words looked "the same" they were considered words of the same language. Should we apply this "method" to German and English basic vocabulary for body parts? These two distinct languages are not mutually intelligible but the words are similar. Note, a few examples:

German	English
Auge	eye
Ohr	ear
Nase	nose
Lippe	lip
Nacken	neck
Schulter	shoulder
Fuss	foot

Is English a dialect of German?

^{xvii} Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands (<https://archive.mpi.nl/tla/elan>).

^{xviii} Note that in December 2024 it was announced that “Indonesian has been added as a new language for the user interface” (<https://archive.mpi.nl/tla/elan/release-notes>).

^{xix} See, for example, Victorian Government (Australia) “Teaching materials and methods” (<https://www.education.vic.gov.au/school/teachers/teachingresources/digital/Pages/activities.aspx>).