12-2022

Evolutionary cartographies of language diversification: Quantitative approaches to the geolinguistic mapping of the Kayanic languages (Central Borneo)

DeAndré A. Espree-Conaway

Evolutionary Cartographies of Language Diversification: Quantitative Approaches to the Geolinguistic Mapping of the Kayanic Languages (Central Borneo)

DeAndré A. Espree-Conaway
Bachelor of Arts, Sewanee: The University of the South, 2013

A Thesis Submitted to the Graduate Faculty of the University of North Dakota in partial fulfillment of the requirements for the degree of Master of Arts

Grand Forks, North Dakota
December 2022

This thesis, submitted on the 15th of December, 2022 by DeAndré A. Espree-Conaway in partial fulfillment of the requirements for the degree of Master of Arts from the University of North Dakota, has been read by the faculty advisory committee under whom the work has been done, and is hereby approved.

_____
Dr. Mark E. Karan

_____
Dr. John M. Clifton

_____
Dr. Karl R. Anderbeck

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota, and is hereby approved.

_____
Chris Nelson
Dean of the School of Graduate Studies

_____
Date

Permission

Title Evolutionary Cartographies of Language Diversification: Quantitative Approaches to the Geolinguistic Mapping of the Kayanic Languages (Central Borneo)

Department Linguistics

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this university should make it freely available for inspection. I further agree that permission for extensive cop ing for scholarly purposes may be granted by me, or in my absence, by the chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or any part thereof for financial gain will not be allowed without my written permission. It is also understood that due recognition should be given to me and to the University of North Dakota in any scholarly use which may be made of any material in this thesis.

DeAndré A. Espree-Conaway
15 December 2022

Table of Contents

List of Figures

List of Tables

Acknowledgements

Every thesis requires the support of a community to be produced. I would like to take this time to thank a number of key members of my community of friends, family, and colleagues for their support during this thesis writing process.

First, I would like to thank my wife, Sharazade Balouchi, for supporting me throughout the thesis process and my time at the University of North Dakota. The first summer after we were married (after just 2 months), I packed up and moved to Grand Forks for two months while she had to remain in Houston. The next summer, I moved her from Houston to Grand Forks to Indonesia and back again. I'm thankful for her being willing to go where my work takes me, for being a great travel companion (especially in Borneo), and for grounding me whenever things got too crazy.

A thank you goes to Brenda Boeger for help with research funding proposals and other questions in language documentation and linguistic fieldwork.

I'd like to thank RISTEK, Universitas Negeri Malang, and Universitas Borneo Tarakan for sponsoring my research visa along with Djaya Bakri and Eva Apriani there. A special thanks to Gatut Susanto for making it all possible and continued research collaborations.

I'd like to thank Sri Ningsih for her support in the growth of my Indonesian language proficiency and visa support as well.

I'd like to thank my advisor and committee chair Mark E. Karan for all the words of encouragement during hardtimes (especially healthwise in the summer of 2019) and for supporting my academic growth as a linguist. John M. Clifton, thank you for all of your support as well and for training me in language documentation methods–the focus of my future career. You two embody the kind of linguist that I hope to one day become. Thank you.

I would like to give a special thanks to Karl R. Anderbeck, one of my committee members, and the EPI/CTNA Asia Language Survey Research Fellowship program for supporting my fieldwork in Borneo. Thanks Karl for the training sessions and other support during this process. This was truly a dream come true that would have been impossible without your support. As part of this program, I'd like to thank Yayasan Suluh Insan Lestari, especially their employees and associates, Eva Wulandari and Daniella Situmorang, for support in the visa process and working with communities.

I must express my gratitude to Keith Slater and Albert Bickford along with the entire program at the University of North Dakota for having me in the program.

I would like to thank my in-laws Randall and Nina Deaton for their support during our stay with them during the height of COVID-19 along with my brother-in-law Ali Balouchi. I'd like to thank Mia Lopez for her support, encouragement, and the many co-writing sessions together.

A special thanks goes to Martinus Joko and the Wehea community for warmly welcoming me into their homes and sharing their culture with me. Thanks Martin for helping me make connections across Kalimantan and accompanying me to Samarinda, supporting my work. Thank you Pak Ledjie Be Leang Song as well for receiving me in your home.

I'd also like to thank Antonio Guerreiro, Lars Kaskija, Bernard Sellato, and Robert Blust for their many correspondences. A special thanks is also due to Antonia Soriente, Alexander Smith, and Jérôme Rousseau for taking the time to give me indepth introductions to the languages and cultures of Borneo.

I would like to thank the Kayanic-speaking communities for allowing me the privilege of working on their languages.

Abstract

Why are there so many languages in the world and how did this diversity come about? Every day, humans around the world speak over 7,000 languages. One of the most impactful theories describing the processes and mechanisms of language diversification is that of the Linguistic Niche Hypothesis (Lupyan & Dale 2010). The Linguistic Niche Hypothesis suggests that languages adapt to their environments (Lupyan & Dale 2010: 1). Lupyan and Dale describe it, stating that "just as biological organisms are shaped by ecological niches, languages structures...adapt to the environment (niche) in which they are being learned and used" (2010: 1). Taking the Kayanic languages of Borneo—a subgroup that shows variegated patterns of diversity (reportedly, homogeneity among some dialects while stark heterogeneity among others)—this study aims to test the claims of the Linguistic Niche Hypothesis. Within this framework, I have proposed an innovative taxonomy which identifies ten main language environments or ecological niches, adding to language ecology theory: (1) physical niche, (2) social niche (3) cultural niche, (4) symbolic niche, (5) cognitive niche, (6) linguistic niche, (7) technological niche, (8) developmental niche, (9) bio-corporeal niche, and (10) genetic niche. This study will focus on physical, social, cultural, and linguistic niches—niches which have emerge as most salient in the creation of linguistic diversity within the Kayanic subgroup.

CHAPTER 1

Evolutionary Cartographies of Language Diversification: An Introduction

> [A] remarkable fact about humans: we speak approximately 7,000 mutually unintelligible languages around the world. This means that a person plucked from one corner of the Earth is not able to communicate with another human in a different corner of the Earth, or often from next door. Apart from a few songbird species, and possibly some whales that learn their songs locally and show dialectical differences, this is unique among animals. … Large as the number of extant human languages is, it has probably reduced from a maximum of perhaps 12,000 to 20,000 different languages before the spread of agriculture, and it pales in comparison with the possibly hundreds of thousands of different languages humans have ever spoken.
>
> —Mark Pagel[1]

1.1 Cartographies of diversification: Mapping the mechanisms of diversity

Why are there so many languages in the world and how did this diversity come about? With people speaking over 7,000 languages today, the question of the production of language diversity rises to the forefront. Since that moment of the emergence of a postulated, perhaps singular, proto-human language, how have languages diversified to reach the thousands that exist today? One of the most impactful theories describing the processes and mechanisms of language diversification is that of the *Linguistic Niche Hypothesis* (Lupyan & Dale 2010). The Linguistic Niche Hypothesis suggests that languages adapt to their environments (Lupyan & Dale 2010: 1). Lupyan and Dale state that "just as biological organisms are shaped by ecological niches, language structures…adapt to the environment (or niche) in which they are being learned and used" (2010: 1). Taking the Kayanic languages of Borneo—a subgroup that shows variegated patterns of diversity (reportedly, homogeneity among some dialects while stark heterogeneity among others)—this study aims to test the claims of the Linguistic Niche Hypothesis.

[1] This quote is from Pagel (2009: 405).

Studying these languages on the island of Borneo, this thesis is an investigation into the complexities of a 'hotspot' of linguistic diversity (Adelaar 1995; Blust 2007; Smith 2017a). The differentiation activity that occurred on this island may shed light on our fundamental question. Borneo of insular Southeast Asia—a landmass the size of Texas—containing over 100 languages, serves as a great case study (Smith 2017a).

The world's third largest island, placed at the crossroads of some of the world's greatest civilizations (Indic and Sinitic) and human migrations (Austronesian and Non-Austronesian), Borneo provides a richness by which linguists might come to understand the dynamics of language change. Despite the abundance of languages, poor documentation of Borneo's linguistic diversity limits our understanding of those dynamics. In light of this, the present study aims to help fill this documentation gap, by enumerating the sufficiently differentiated language varieties and thereby making possible the revealing of those inherent processes and mechanisms of their formation. The standard of 'sufficient differentiation' is founded on a criterion of mutual intelligibility—the comprehensibility of language varieties among one another. After establishing 'where one language ends and another begins,' this case might reveal larger patterns about language diversification.

To avoid becoming lost in a sea of 100 languages, for feasability's sake, this thesis focuses on one corner of Bornean languages—the Kayanic language family. Kayanic is a subgroup of the Malayo-Polynesian family which itself is a subgroup of the massive Austronesian language family—the most linguistically dense and, until 1500 CE, the most geographically extensive language family in the world. This thesis is a language survey of the Kayanic subgroup. It aims to identify and map these language varieties through the collection of *calculated intelligibility* data.

1.2 Research question and argument

The guiding research question of this study is thus: Does the phonological structure and lexical composition of the Kayanic languages, and, consequently, the patterns of dialectal intelligibility derived from it, support the Linguistic Niche Hypothesis—showing adaptation to ecological factors in this language subgroup's environment? The idea is, here, that once unified language varieties experience a divergence or convergence in language structure, predicated on whether they are entering different or similar combinations of niche factors—that is, whether they are entering different or similar ecologies—to which they adapt. Adaptation to certain combinations of niche factors encourages differentiation while others encourage homogenization. I have proposed that there are ten such ecological niches, as we shall see in Chapter 2. These processes of adapting to the environment—the position of the Linguistic Niche Hypothesis—is what drives language diversification.

By spatially mapping the Kayanic language subgroup based on intelligibility, creating a kind of 'cartography of language diversification', this study demonstrates that these languages show signs of adaptation to ecological factors which drive languages either toward homogeneity (remaining a dialect of a single language) or heterogeneity (splitting into separate, "new" languages). I hypothesize that, indeed, the phonological structure and the lexical composition of the Kayanic languages and the patterns of intelligibility both show signs of ecological adaptation to physical factors (language contact through travel and geographic ease of travel) and social factors (language contact through trade/economic activity, warfare, population size, and intermarriage) and therefore supports the Linguistic Niche Hypothesis.

1.3 Claims of significance

This study, about how language diversification is driven by environmental factors,

captured through geolinguistic maps and intelligibility percentages, is of interest to dialectologists, Austronesianists, and scholars in the areas of language ecology, multilingualism, and evolutionary linguistics. By revealing the 'intelligibility zones' or 'languages,' the applied linguistic applications of this study might probably be of interest to the Indonesian Ministry of Education (Kementerian Pendidikan dan Kebudayaan) and the Indonesian Language Center (Pusat Bahasa), in particular, but also to developing countries, in general, aiming to create data-driven, cost effective, mother tongue-based multilingual education (MTB-MLE) programs (cf. Blair 1990; Ginsburgh & Weber 2011). By consolidating 'intelligibility zones,' countries with high linguistic diversity have fewer discrete groups to which they need to provide language-specific materials. This study also works to document endangered and minority languages previously understudied and provides a raised profile of Kayanic-speaking people groups aiming to protect the rights to their traditional forests.

1.4 Language maps

The Ethnologue (Eberhard et al. 2021) is often considered one of the most authoritative sources in diversity linguistics. It contains language inventories and maps based on the previous scholarly literature, historical sources, and, most importantly, original linguistic fieldwork. I am providing these maps in the introduction to give the 'lay of the land' in terms of the general location where Kayanic languages are spoken.

This map depicts the languages of Kalimantan (Indonesian Borneo). Numbers 36 (Mendalam Kayan), 43 (Mahakam Kayan), 58 (Modang), 59 (Bahau), 61 (Busang Kayan), 63 (Kayan River Kayan), 66 (Punan Merap), 83 (Segai), and 85 (Wahau Kayan) are all Kayanic languages shown on this map. Number 43 representing Mahakam Kayan appears in two areas

and number 58 representing Modang appears in four areas. Note that much diversity within the Kayanic subgroup is not indicated here.

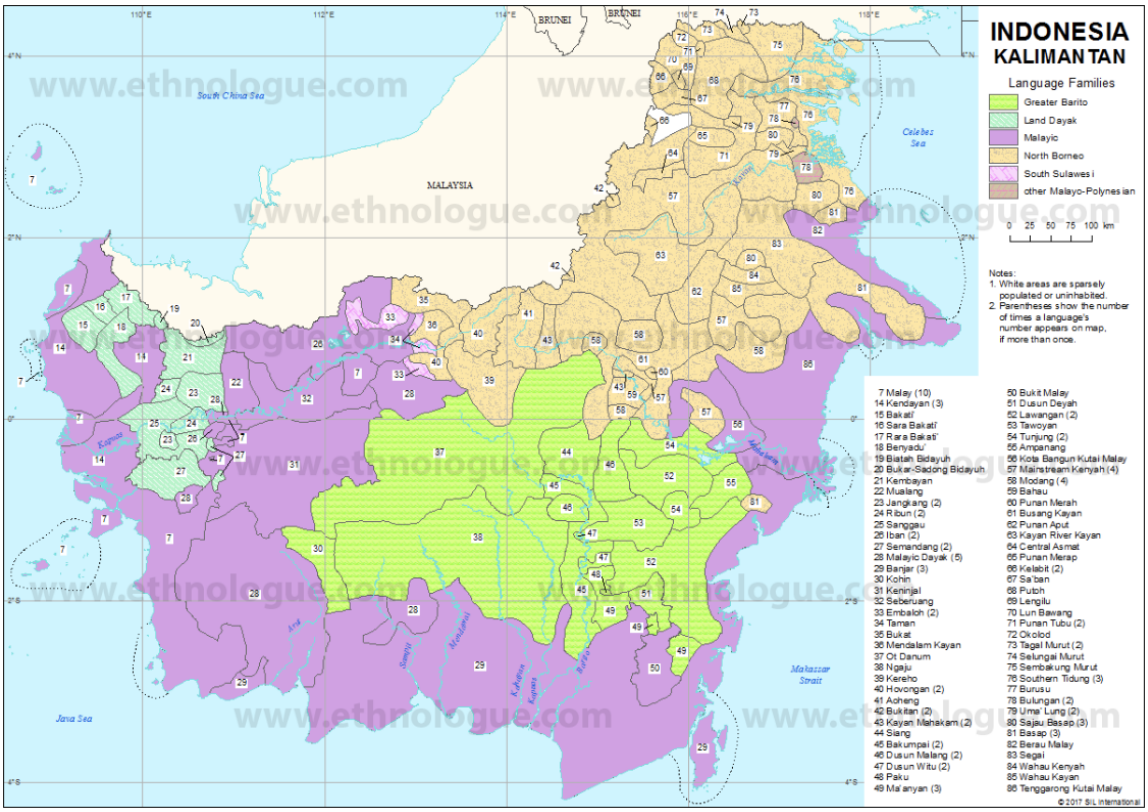


Figure 1. Ethnologue Map of the Languages of Kalimantan (Eberhard et al. 2021)

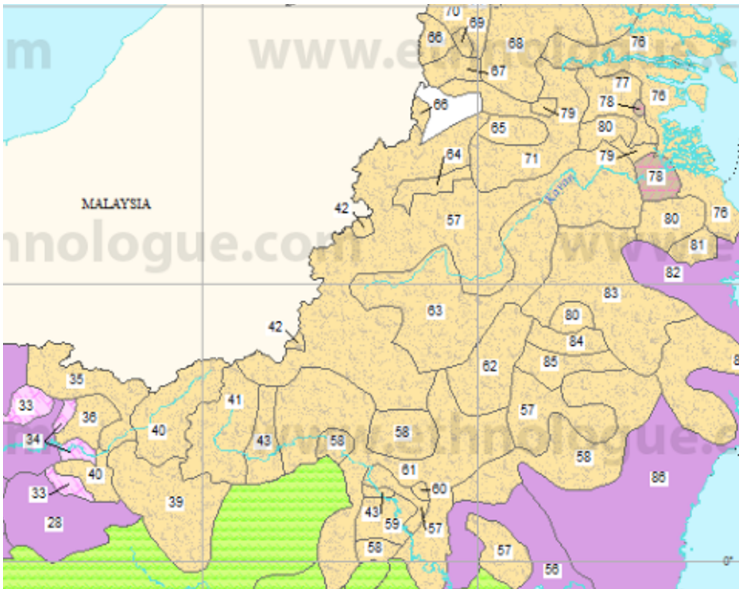


Figure 2. This map zooms in on the Kayanic area (Eberhard et al. 2021).

This map depicts the languages of Sarawak (Malaysian Borneo). Numbers 48 (Rejang Kayan), 58 (Murik), and 71 (Baram Kayan) are all Kayanic languages shown on this map. Number 48 representing Rejang Kayan appears in three areas and number 58 Murik appears in two. Note again that much diversity within the Kayanic subgroup is not indicated here.

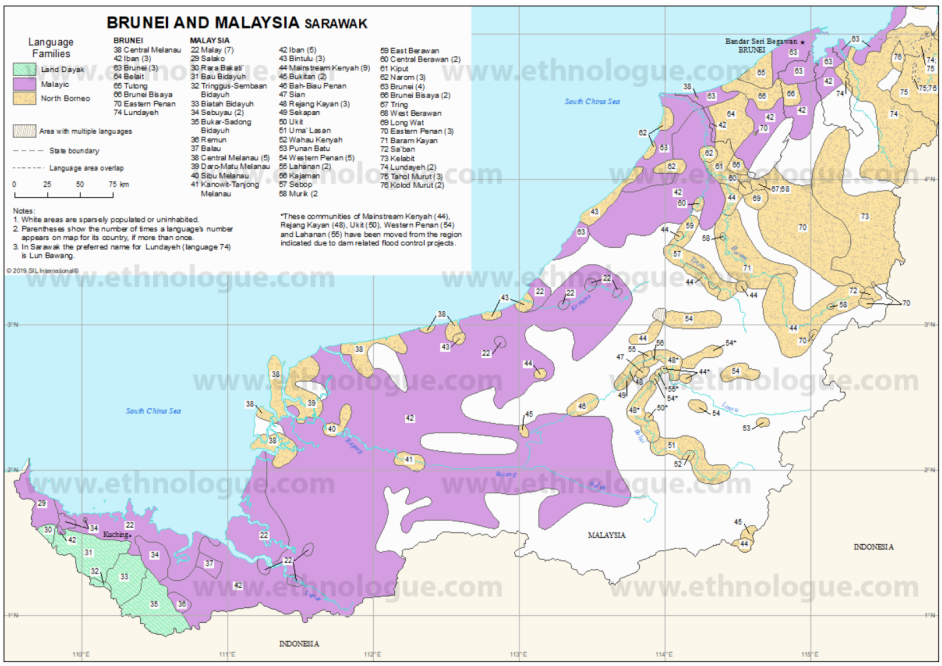


Figure 3. Ethnologue Map of the Languages of Sarawak (Eberhard et al. 2021)

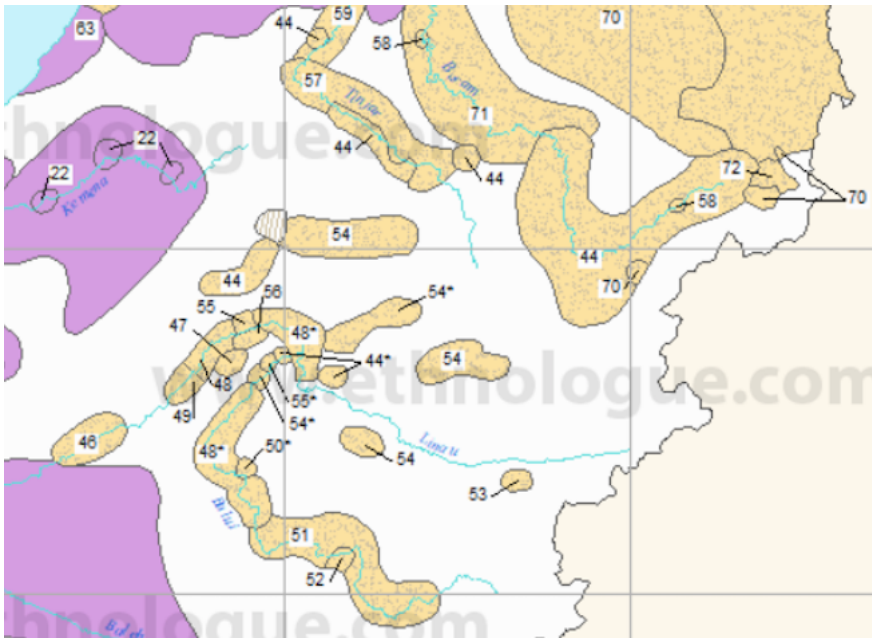


Figure 4. This map zooms in on the Kayanic area (Eberhard et al. 2021).

Finally, this map provides an impressionistic view of the location of the Kayanic language subgroup in Central Borneo. This gives us a ‘bird's eye view’ of the distribution of Kayanic languages across the entire island.

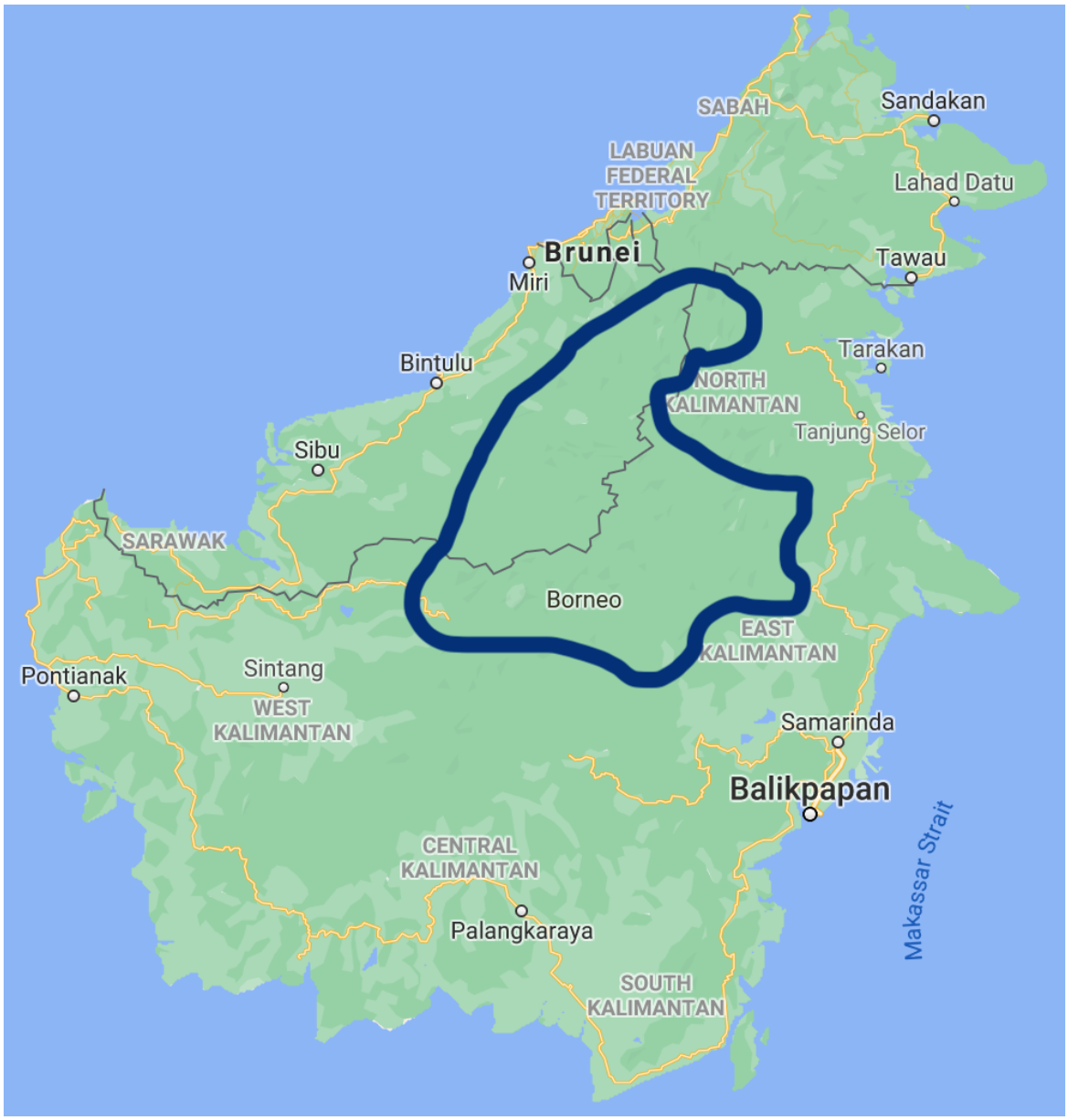


Figure 5. Impressionistic Distribution of Kayanic Languages (Borneo Map from Google Maps)

1.5 Problems and approaches to addressing them

Some of the main issues encountered when trying to understand language diversification are based in theory while others are based in the practical application of this knowledge. The

theoretical plausibility of studying the language diversification process from a synchronic perspective requires the inferential interpretation of present states. The assumptions that these inferences are built on must also be addressed in order to facilitate replicability and thus maintaining this approaches predictive validity within certain parameters. The practical use of these interpreted findings also naturally warrants discussion.

1.5.1 Can synchronic spatial dialectology alone account for language variation dynamics?

While Schmidt (2010: 205) suggests that the methodological issue of the incertitude about prior states of language (and thus the development of language change) limit the explanatory power of synchronic dialect mapping studies, unidimensional (non-real-time) 'cartographies of diversification' can be interpreted to illustrate the dynamics of language variation. In fact, they are the only kind of empirical studies possible on language dynamics in most cases.

Real-time studies that compare complete language maps or atlases drafted between two periods in time (for example, the Wenker survey comparing the German of 1880s to that of the 1960s (Schmidt 2010: 204)) for the same area of language varieties face the several methodological issues of their own, namely that of true comparability across lines of common sociolinguistic factors (age, socio-economic status, ethnic identity, gender, etc.). While Schmidt characterizes these dialect maps as "frozen random samples," this study escapes its seemingly static constraints by focusing on diversification as produced by *everyday practices in variation*, rather than seeking rationalist approaches to uncovering *change*. It must be emphasized that here I study language *variation*, and not language *change*. The evidence that this study provides may index change, but not demonstrate that change itself (a fact that, from a validity standpoint,

eludes even historical-comparative, real-time, and apparent time studies, because they also remain somewhat uncertain).

The "relics" or "vanguards of innovation" observed at the synchronic level can point to past ecological conditions and language changes without seeking to show the exact trajectory per se (Schmidt 2010). This study's reliance on lexical data can yield insight into possible past language contact situations. The implication is that current relations (and daily practices) that, if prolonged from the past to the present (or into the future for that matter), have or will have resulted in creating more difference or more similarity or stably maintaining a certain set of features among language varieties that can be measured through intelligibility. These measurements need to be coupled with an ethnographic or ethnohistorical account to provide the context of those current relations and daily practices. This ethnographic context creates a language use-based 'bridge' across the gap between interpreted language dynamics and their present geospatial distribution (cf. Haas 2010). By asking the following kinds of questions of the synchronic data and the ethnographic record, linguists might arrive at these language diversifying dynamics:

> What are the ecological motivations for divergence or convergence? Are there cases of dialect leveling? In what language varieties are people multilingual? Do they have complete proficiency in their bilingual ability or partial? Are there signs of the formation of contact languages (pidginization of local variants/languages?) Is there code-switching or mixing among dialects/languages? Are there lexical accommodations? Or phonological accommodations? Do both speakers accommodate? Does just one speech community accommodate? Perhaps across geographic (upriver-downriver; mountain-valley, etc.)? Do social class lines matter? Perhaps across social class lines within and between languages (e.g. upper class members only communicate with other upper class members in a given language or between languages or perhaps lower class or middle class members could serve as envoys among the upper classes between languages)? Is there egalitarian multilingualism in some areas? And hierarchical relations in other? [We know almost certainly the latter, since the ethnographic record shows that Kayanic cultures are very hierarchical].

In sum, the major question first and foremost is, who understands whom and under what conditions? This is the focus. This, what I call the *praxis of language variation*, provides insight into just that alone—language variation or diversification at the synchronic level. Language

change is thus rendered an epiphenomenal concern—a residual reflection to which the synchronic data points. This does not mean this method is not robust. Using the synchronic to reveal and predict the past (necessarily diachronic) language change, without actually touching diachrony, provides a plausible set of factors to later be judged against the diachronic, real-time, actual-time, and, I would add, more ethnohistorical evidence. This approach has the power of offering plausible explanatory evidence of the historical dynamics (further elaborated by true diachronic, real time, and/or apparent time studies) as well as a potential predictive power, predicated on current occurrences. A linguist might be able to say, "If the trajectory of X continues as it currently exists, then the language will actualize into a state of Y."

To index these implied trajectories based on the praxis of language variation requires a theory that asks these questions of both the intelligibility measurements (along with its geospatial visualization, i.e. language maps) and documented ethnographic accounts. The theory of 'language ecology' viewed through the lens of biological evolution fulfills this requirement.

In line with what Mufwene has demonstrated about the importance of the ethnography to understanding the evolutionary ecology within which languages diversify (2001), the ethnographic record affords us a "dynamic interpretation of static maps," (Haas 2010) providing the longitudinal sociolinguistic context whereby we might interpret the emergent variation, visually represented on the maps.

The ethnographic record, being relatively short (early 19th century to the end of the 20th century) allows us to control for the effects of the concerted evolution of neutral drift in language change (Hruschka et al. 2015)—the random, but regular sound changes, focused on in historical linguistics. Normal sound change, if it is thought to contribute to the discrete historical 'stages' of a given language (i.e. Old English to Middle English to Modern English, etc.), a perceptual

discreteness with a structural reality, occurs roughly every 385 years—a rate of 0.0026 individual changes per year (Hruschka et al. 2015: 5)[2]. By since our ethnographic record is securely less than this regular sound change window, we can be fairly certain that most of the observed differences, tabulated in the intelligibility percentages are due to forces of ecological adaptation (climate, landscape, multilingual contact, intermarriage migration, trade, etc). Another to put this is controlling for the accumulation of natural sound change, we can come to understand how the state of our 'cartography of diversification' is the result of evolutionary processes of adaptation to various niche factors in the language ecology, captured in the ethnography—a prediction of the linguistic niche hypothesis.

With our evolutionary ecology theory securely in place, we can now deal with the matter of how to actually interpret the cartography of diversification, i.e. the intelligibility measurements and its resulting language map. This calls for an exemplar.

Walter Haas's study of German dialects from the *Sprachatlas der deutschen Schweiz* (SDS) shows how we can go about performing a "dynamic interpretation of static maps" (Haas 2010). Taking an *areal diffusion* approach—'diffusion' describing the "dynamic process across time and space in which an increasing number of speakers add a linguistic feature to their repertoire" (Haas 2010: 649), Haas provides a framework for describing how geospatial distributions unfold. He writes (2010: 649):

> The distribution of a feature on a linguistic map is a frozen geographic reflex of a shift in linguistic behavior - a change in language. … Speakers with their attendant psychological, social and linguistic preconditions, linguistic features, time, and space must all be brought together into a systematic relationship capable of "explaining" the spatial patterns discovered, that is, able to suggest reasons for the patterns being as they are.

[2] This should not imply that languages always change at a constant rate. It is very well documented that various factors affect the rate of language change (e.g. language contact, cultural orientation, technology, language shift, among others). In sum, sound change is not the only factor driving language change.

Due to the psychological and social reality that idiolects of various individuals acquire features in a speaker-to-speaker manner and that this acquisition is linked through a gradual spatial trend, it can be said that dialects diffuse along this same pattern, from speaker to gradually more distant speaker, from dialect to gradually more distant dialect (see Mufwene 2001 for 'idiolect approach' and Haas 2010 for this diffusion principle). He calls this *extensive diffusion*. The process necessitates a certain amount of implied relative time to diffuse from group to group, even if we do not deal with the specifics of the time scale. These at-first individual features interact with the features already present in a given group's language variety at the phonological and morphological levels, gaining *intensive diffusion* throughout that variety as the new feature competes or merges with other features as it becomes fully integrated into the rules of that language variety's system. This is happening while the language variety is being passed down generationally, further structuring the integration of the new feature (Haas 2010: 654).

This intensive diffusion, internal to a given dialect, can create a chain of associated features, linked to the acquisition of the new feature. The configuration of this chain of accrued features is what constitutes the different dialects and their distribution of gradually smaller feature chains. Another way to put this is that from point X in space feature A will arise in dialect III. Dialect III will propagate feature A to neighboring dialect II while feature A offsets the rise of feature B within itself. Feature B offsets a feature C within dialect III as time progresses. With dialect II having acquired feature A it, in turn, propagates feature A to dialect I while adding a feature B to its chain (Haas 2010: 654). You ultimately end up with a distribution like this:

Table 1. The Preservation of "Historical Implicational Relations" (cf. Haas 2010: 656)

Geographical Reflection:	Zone III	Zone II	Zone I
Stages of Innovation:	ABC	AB	A

I call this distribution a preservation of historical implicational relations, because Haas notes that gradual distribution of features establishes a required trend which supports the integrity of this approach to interpretation. This *dialectal-geographical radiation*—a wave based theoretical concept, typical in dialectology and historical linguistics (Chambers & Trudgill 1998; Campbell 2013; Crowley & Bowern 2010)—provides *implicative coherence* which is a requirement of this interpretive approach. Simply put, dialectal-geographical radiation on a map demonstrates historical implicational relations, giving rise to an implicative coherence, emerging from that map.

In addition to implicative coherence, another requirement is that the distribution of language varieties maintains a *geographical coherence*. Establishing this trend ensures that the features are successively acquired from one language variety to another and did not happen to arise within each language variety independently of each other. The feature distribution must show a "geographic constellation" in "wave formation (*wellenbild*)" (Haas 2010: 651).

With the requirements of his model in place, we may now turn to interpreting his language maps. This first map is a case of short vowel lowering. In dialect III, ë > [æ] which also spread to dialect II and dialect I. Dialect III seems to have also acquired these two features: e > [ɛ]/ o > [ɔ] and i > [e]/ u > [o]. Only the e > [ɛ]/ o > [ɔ] feature was propagated to dialect II. The third feature was not. Also, note that only the first feature was propagated to dialect I. The maps show a left-to-right diffusion of these features. This represents a west-to-east diffusion in geographical space with the gradualness of the diffusion maintaining geographical coherence.

The fact that each propagated feature builds on a previous feature gives this map its implicative coherence.

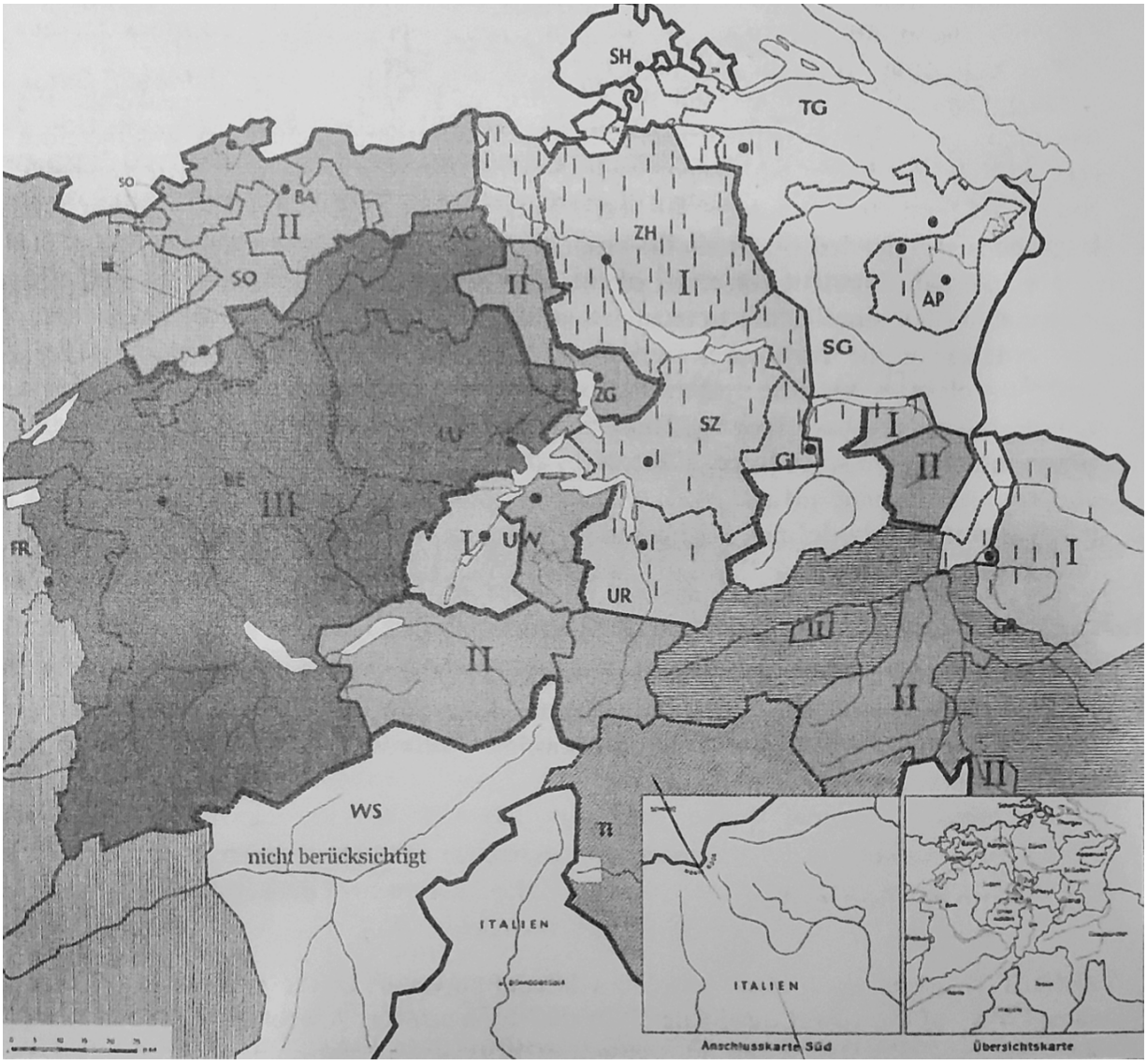


Figure 6. Map of Short Vowel Lowering from *Sprachatlas der deutschen Schweiz* (Haas 2010: 659)

This figure gives an idealized mapping of the geographical map above, demonstrating the model as it adheres to implicative and geographical coherence. It is based on a model figure in Haas (2010: 655-656).

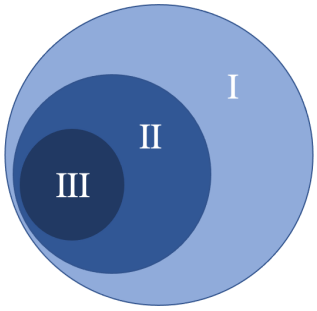


Figure 7. Idealized Map (Haas 2010: 655-656)

In this second case, we see long vowel raising.[3] Dialect III contains the feature ē > [eː] which also spread to dialect II and dialect I. This is the only feature propagated from dialect III to dialect I. Dialect III seems to have also acquired features ā > [ɔː] and æ > [ɛː]. The ā > [ɔː] feature was propagated to dialect II, but not the third. The map shows a right-to-left diffusion of these features. This represents an east-to-west diffusion in geographical space with the gradualness of the diffusion maintaining geographical coherence. The fact that each propagated feature builds on a previous feature gives this map its implicative coherence.

[3] Note that while Haas uses a circumflex to indicate a long vowel (e.g. 'ê'), I have decide to indicate vowel length with a macro which is perhaps a more modern notation convention (e.g. 'ē').

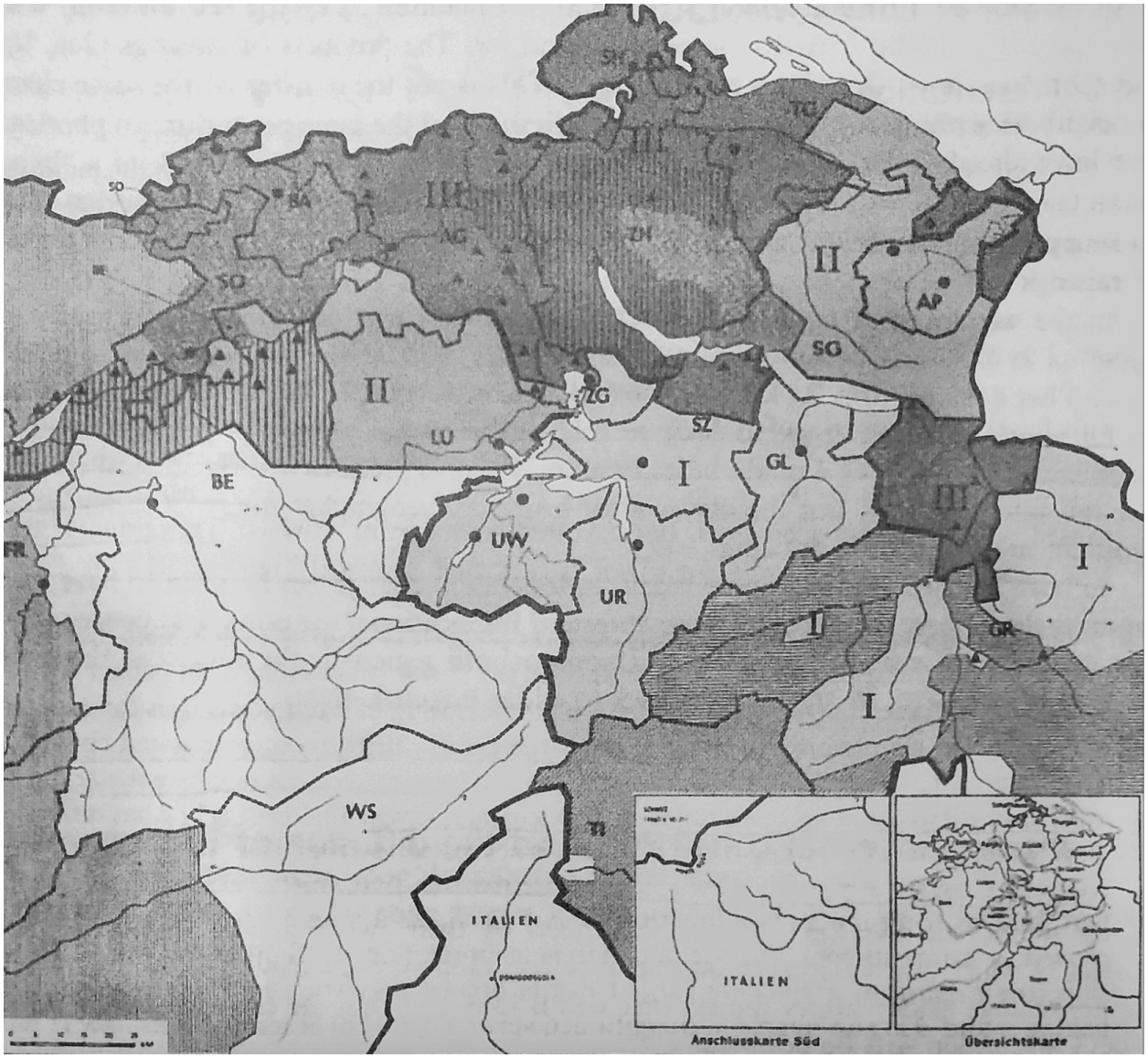


Figure 8. Map of Short Vowel Lowering from *Sprachatlas der deutschen Schweiz* (Haas 2010: 659)

It should be noted that there is something unusual about the ordering in this map. In certain areas, particularly in the southeast, the dialectal radiation appears to be disjointed, jumping from dialect III to dialect I. Any number of reasons could cause this (i.e. radiation from "parachute" dialect urban centers or perhaps the physical geography). This is not a grave matter since what is important for implicative and geographical coherence is the global trend. This allows our model to handle variability, inherent to any linguistic situation (cf. Haas 2010). The idealized map below depicts the global trend of east-to-west diffusion whereby the dialectal

features successively built upon each other, maintaining both implicative and geographical coherence.

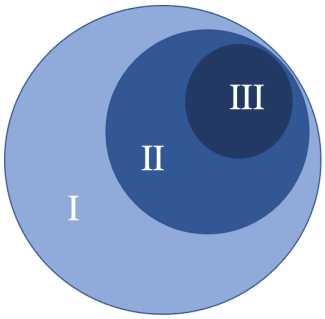


Figure 9. Idealized Map (Haas 2010: 655-656)

Haas's study (2010) provides a solid framework for interpreting language maps. This demonstrates that "dynamic interpretation of static maps" is indeed possible. What Haas's framework means for this study is that we can examine, not how one or a certain set of features propagate per se, but rather how sets of features that have solidified into various language varieties which can be globally measured through intelligibility have propagating creating a implicatively and geographically coherent distribution.

Table 2. The Preservation of "Historical Implicational Relations" Applied to Intelligibility

Geographical Reflection:	Zone III	Zone II	Zone I
Intelligibility Dynamics:	90%	80%	70%

The idealized expectation would be to have an area corresponding to Zone III where all language varieties are at perhaps 90% intelligibility (which is really a 90% benchmark past a limit that indicates that there is no structural barrier to communication, although social ones might still exist. More on this in the Methodology section). There might be a Zone II where the various languages are scored in the 80% range for intelligibility. Perhaps a Zone I will appear where the language varieties score in the 70s or even the 60s. These successive patterns of intelligibility, maintaining both implicative and geographical coherence could be depicted in an idealized map such as the one below. The black dots represent different language varieties in each zone.

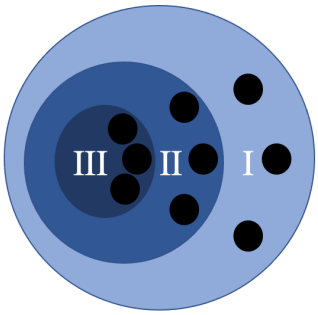


Figure 10. Idealized Model Map of Intelligibility Diffusion and Dialect Distribution

By showing that the trend of our model is represented among the Kayanic languages of Borneo while also describing seemingly disjointed areas due to specific niche factors in the linguistic environment, we can examine whether patterns of intelligibility captured in language maps support the linguistic niche hypothesis—i.e. the case that dialect intelligibility and its geospatial distribution adapt to the environment.

1.5.2 Assumptions of the study

In order to embark on understanding the evolutionary mechanisms of language diversification through ecological niche adaptation (the assertion of the linguistic niche hypothesis), we must take into account what assumptions our theory makes. By taking a certain theoretical stance, this study makes several assumptions a priori:

1) *Language ecology*: languages do not exist in a vacuum, discrete from the context within which they exist. Therefore, one of this study's assumptions is that linguists can observe and, thus, describe the dynamics of language variation, particularly at the synchronic level, in terms of speakers' daily "cognitive and interactions-cum-communicative" practices (Schmidt 2010: 202). These practices along with the physical, cultural, socio-economic, and political niches are all competing and complementary factors that compose an environment in which languages vary and diversify—the language ecology.

2) Ecocriticism and its methodology Ecocritical Discourse Analysis are a separate enterprise from the evolutionary theory of 'language ecology' which aims to describe language change (and evolution) through the apt metaphor with biological evolution.
3) While the diachronic approach to language evolution is a valid one, it is not the only approach to the study of language evolution. An overwhelming majority of studies in the evolution of language diversity takes a diachronic approach which is sensible since it most readily demonstrates that principle of evolution, 'descent with modification'—with 'modification' mapping directly onto the principles of descend established through 'shared innovations'—the most foundational of ideas in historical linguistics. However, a synchronic approach to language diversification is also possible (although understudied). This study takes that approach.
4) Only humans have language. Microevolutionary approaches to examining language diversification or speciation is completely conspecific. While other animals, particularly other primate might have complete systems of communication (cf. Seyfarth & Cheney 1980; Ouattara et al 2009), language is an exclusively human capacity.
5) The role of language contact as it acts upon the unit of the idiolect—one's individual language system—is crucial to understanding language adaptation. This study, in line with Mufwene (2001: 21) "cast[s] doubt on the position that the role of contact is negligible in "normal" language evolution." Language contact is 'front and center' in this study.
6) Language is a *complex adaptive system* (Benz 2018; Mufwene 2001; Beckner et al. 2009).

Knowing these assumptions allow us to move forward in applying our theory to the language data presented here.

1.5.3 Language development concerns

One of the practical applications of this knowledge (as mentioned above in the 'Claims of significance' section) is the use of the 'intelligibility zones' or 'languages,' represented on the language map as a guide to creating mother tongue-based multilingual education (MTB-MLE) programs. The findings of this study can help organizations locate the 'dialect of reference' for a given community. These language varieties can be used to create literacy materials and other scholastic tools that can be readily understood in various communities. This possibility of 'teaching in the mother tongue' can help countries build their human capital by helping its citizens learn to read better in general which has the effect of better national language and potentially foreign language skills. This increases the economic prospects for both the individual and the country as a whole. It is ethical that research in a developing part of the world has the capacity of giving some benefit back to the communities that participated in it.

1.6 Kayanic and the region of Central Borneo

Kayanic speakers consist of several Dayak people groups. Dayak is a term, referring to the non-Muslim, non-Malay, autochthonous groups of Borneo. This term has locally specific significance, covering over 400 ethnic subtribes. The current Dayak population for all of Borneo is estimated around 2.2 million. While most Kayanic people groups are classified as 'middle range' societies[4], some maintained hunter-gatherer lifestyles as well. Prehistorically, their homeland in the Apo Kayan region, Kayanic speakers were pushed out after extensive warfare

[4] Middle-range societies are "pre-state sedentary societies" (Rousseau 2001).

with the Kenyah, another major Dayak group of Borneo, settling the riverine areas that cover eastern part of Central Borneo.

1.6.1 Geography

It is important to remember that geography has played an important role in shaping Borneo history and the socio-cultural formation of its peoples. The abundance of rivers along with their, in some cases, impenetrable rapids have created natural barriers that often transform into social ones. Thick, equatorial rainforests and rough mountainous terrain are also dimensions of the physical ecology within which Bornean speech communities live and interact. Rainforest and seasonality differences also exist across the island probably due to elevation or forest coverage levels (Rousseau 1990: 10). Also, as of late, the political geography has played a role in forming part of the socio-cultural (or political) ecology of Kayanic speakers. The island's division between three countries (Indonesia, Malaysia, and Brunei, although only Indonesian and Malaysian Borneo have Kayanic people groups), with different approaches to language policy, environmental law, and indigenous people-government relations, has had a differential impact on the nature of language use within Kayanic communities.

1.6.2 History

All Kayanic-speaking people groups originate in the mountainous region of Central Borneo, known as the Apo Kayan (Apo or Apau Kayan, literally the 'highlands of the Kayan'). While out-migration began around the 18th century (Sellato 1993: 173), population movement was probably further catalyzed by the influx of warring Kenyah groups into the area (cf. Guerreiro 1993). On a quest for new territory, early Kayanic speakers began, first, by filling in river drainage areas, north of the Mahakam river. Speakers of what would become the Segai-Modang languages fought their way into the Kutai river basin. Groups that would later

speak Murik and Merap filled in the lower Kayan and Malinau river areas. In present-day Sarawak, Kayan language speakers migrated into the Rejang and Baram rivers, splitting linguistically into different dialects there (Smith 2017a: 409).

The Kayan first reached the upper Mahakam around 1760. Some Kayan (Mendalam Kayan) migrated down the Baleh river into the area, just northeast of the Kapuas wetlands. Approximately half a century later, the Long Gelat and the Uma' Suling (who would later become the more demographically pervasive Busang) would fill in the Mahakam region (Sellato & Soriente 2015: 342). According to Smith (2017a: 409), the advent of Kayanic speakers into the Mahakam area "completely leveled any linguistic diversity which might have existed there in the past."

Many former people groups have been displaced or absorbed during the expansion of Kayanic-speaking groups in Central Borneo. This has led to a wide variety of language contact situations which have shaped the languages of Kayanic people groups.

1.6.3 People: Dayak groups of the Interior

Below I provide a more detailed overview of the languages and cultures of Kayanic peoples. These 'people profiles' are divided by the subgroups within the Kayanic language family as outlined in Smith (2017a). First, I describe the speakers of various Kayan language varieties. Then, I approach the Modang tribes. Finally, I describe the Ngorek or Murik peoples that comprise the Murik-Merap subgroup, mentioned in Smith (2017a).

1.6.3.1 Kayan

Of the diversity that seems to characterize Borneo, what is most conspicuous about the Kayan languages is their thorough homogeneity (Rousseau 1990). Kayan forms a dialect continuum placed directly at the heart of the island. The name Kayanic or Kayan derives from

Kaya:n (sometimes written *Kayaan* in the literature) or *Kejin* meaning 'our place,' the name given for the Kayan river basin area by invading Kayanic groups. This name replaced the older place name for this region, *Bulungan* (Okushima 2006: 88). The Kayan form one of the most influential cultural groups in the interior of Borneo.

The Kayan point to the headwaters of the Kayan river in the Apo Kayan as their original homeland. Migrating out of this area (although some remained in pockets), the Kayan were a highly mobile and warlike group staking out territory and assimilating various groups (particularly, enslaving Muruts among others) down to the Mahakam river—one of the largest rivers in Borneo. They named this area, what was once the Kutai River, the *Mahakam* (or *Meka:m*, *Mekiam*, *Mahkam*), meaning 'ocean' or 'broad surface of water.' A tributary of the Segah river was also given this name—an area after which the Punan Mahkam are named (Okushima 2006: 88).

For the overall Kayan population size, various figures have been given, suggesting a population between 25-35,000 individuals.[5] The Baram Kayan and Rejang Kayan, located on the Malaysian side of Borneo in Sarawak, consist of 4,150 and 3,000 people respectively (Eberhard et al. 2021). Kayan River Kayan (sometimes referred to by various village names Uma Leken or Uma Laran) is located at the headwaters of the Kayan river and continues to various areas downstream. It has a population of around 2,000. The Mendalam Kayan on the far eastern reaches of the Kapuas number around 2,000 individuals in some estimations (quoted in Guerreiro 2002: 106; Ding Ngo & Lii' Long 1985). In others, they number around 1,500 (Eberhard et al. 2021). The Mahakam Kayan and Wahau Kayan, located on the Mahakam and the Wahau rivers, number 1,300 and 500 respectively (Eberhard et al. 2021). Note that these 'Wahau Kayan' may

[5] Rousseau (1990:15) estimates around 25,000. An unverifiable source suggests a figure of 27,000. After calculating the sum of all language groups mentioned in the Ethnologue (Eberhard et al. 2021), I arrived at a figure of 34,480 individuals. Given this variability it seemed wise to state the range.

be relatively new immigrants to the Wahau river, arriving within the last half century, perhaps due to a government resettlement program. Therefore, their most recent 'home territory' may be elsewhere.

The Bahau are, by far, the largest Kayan group of 19,000, located up and down the Mahakam river from Tring onward (Eberhard et al. 2021). The Busang, also along the Mahakam, numbering around 3,000 (Eberhard et al. 2021), are a "subcategory" of Bahau, although the groups make fairly stark distinctions between themselves (Rousseau 1990). A linguistic distinction is always mentioned (Espree-Conaway 2020; Rousseau 1990; Sellato 1980): the true Bahau word for 'no' is *bate* (therefore, they are the 'Bahau-bate') while the Busang word for 'no' is *jaan* (therefore, they are the 'Bahau-jaan'). This distinction is still made to this day (Espree-Conaway 2020). Busang villages appear to be interspersed throughout a sea of Bahau villages that line the Makaham river.

Kayan villages are generally located right along major rivers and many are only accessible by boat (particularly at certain times of the year). Traditionally, Kayan villages consist of a collection of ironwood longhouses, each up to 300 meters in size, with each containing around 200-300 individuals. Up to 500 individuals in a single longhouse has been documented as well. As is typical of Borneo, houses are built on stilts for flooding, and perhaps in the past, for defensive purposes. Village populations may range from a mere 30 individuals to over a thousand (Strouthes 1993). Note that in all of the Kayan villages I visited, it appears that individual family housing is now preferred.

Kayan societies are notoriously hierarchical (Strouthes 1993). Of the many kinds of social arrangements found in Borneo, the Kayan, like most Kayanic-speaking groups, including the Modang (but not the Punan groups or the Ngorek groups), are of what Sellato (2016) calls the

'stratified' type. Kayan societies are generally divided into four categories: ruling aristocrats (*maren*), [lower] aristocrats (*hipuy*), commoners (*panyin*), and, formally, slaves (*dipen*) who were prisoners of war from conquered tribes (Strouthes 1993). Aristocrats controlled villages and married with other villages across ethnic and linguistic lines to form alliances (Strouthes 1993: 134). This will become important when discussing language contact and ecological niche factors to which their languages may have adapted.

As is typical of Southeast Asia, rice plays an important role in the economic and symbolic life of the Kayan. The Kayan are what Rousseau (2001) calls 'middle-range' swidden agriculturalists. The Hudoq ritual, performed to encourage a fruitful rice harvest, still remains strong among the Kayan (especially among the groups of the Mahakam, in particular). This too will become important when discussing ecological niche factors and language adaptation.

The Kayan are well known in the ethnographic literature for their former headhunting practices. The *mamat* ritual (head feast) was the collection of a new head for "ritual purification" after the passing of a Kayan family member or the building of a new longhouse in the village (Strouthes 1993: 134). This ritual is now rare after the advent of Christianity and no new heads have been collected in over a hundred years. The vast majority of Kayan now adhere to the christian faith, despite being situated in predominately muslim countries.

1.6.3.2 Modang

The Modang, pronounced [mədaŋ], are a striking case of linguistic heterogeneity within the otherwise mostly homogeneous Kayanic subgroup. Although 'Modang' was originally a somewhat insulting exonym (Guerreiro 2002: 96), today the term seems to be embraced by some groups (perhaps, as part of an expanding or realigned, larger dayak ethnic identity in the face of Indonesian nationalism), while naming distinctions are stressed by others. Modang (or, as it is

sometimes known in the linguistic/anthropological literature, Segai-Modang or Modang-Menggaè) is a language-culture complex of five main agriculturalist people groups, along with two minor Punan (hunter-gatherer) groups, located exclusively on the Kalimatan side of Central Borneo. These include the Wehea, Long Gelat, Long Belah, Long Way and Menggaai along with the Punan Kelai and Punan Mahkam (Guerreiro 1993: 185). While Guerreiro (2002: 96) includes the Punan Merap as a culturally related group, their language is historically quite distinct (Smith 2017c) and therefore they will be dealt with elsewhere (Section 1.3.4.3).

It should be noted that there are a variety of names, attested for each of these groups. This is not incidental. The ethnonyms of these groups have changed over time as groups migrated to different areas. Various language varieties may possess slightly different names for any given group. Groups themselves may have often modified their endonyms to differentiate themselves from groups with whom they may have once been closely tied. Blanketed naming practices occurred from the Malayic civilizations on the coast (for example, the Bulungan Sultan referred to all Kayanic people groups as Segai [Segai'i/Segayi]) which adds to the confusion and colonial powers are said to have further stoked this confusion as well (Okushima 2006: 86-87; Guerreiro 1993: 185).

The Wehea (Wèhea) are sometimes known by the Indonesianized-version of their name, Wahau—the name, officially given to one of the main rivers where they live. They are also referred to as the Long Wahau, Wahau Modang, or Sawaw (the latter meaning something along the lines of ‘those people from the Wahau river’ (Guerreiro 2002: 96). The Long Gelat are sometimes known as Long Glat, Long Gelât, Long Gelaat, Long Gleaat, Long Glit, Long Ge’lat, Long Gliit, Long Geliit, or Lung Gelaat. These are probably just various spelling types, used throughout the literature. Long Gelat or Long Gelaat may be the Bahau/Busang pronunciation of

this group—a Kayanic group among whom the Long Gelat lives. The Long Wai are also referred to as Long Way, Long We, Kelingan, or, simply, Medang. The Long Belah are also known as Long Bleh or, simply, Medéang. The Gaai (Ga'ai/Gaay/Ga'ay/Gaè) are often referred to as the Menggaai (Mengga'ai/Mengaay/ Mengga'ay/Meng' Ge'ai), Menggaè, or Segai. Mengga'ai (or Menggaè) is the endonym while Ga'ai (Ga'è) is the Kayan-Kenyah exonym (Guerreiro 2002:96). Punun Kelai (Punan Kelay) and Punan Mahkam seem to be rather straightforward. Note that there are probably endless pronunciation and spelling permutations for many of these names, given the multilingual situation of the region.

For the convenience of understanding how these groups' language varieties converge or diverge, I will follow the more anglicized 'placename-historical ethnic category' format for most of these, as used above for the Kayan varieties. Those names are as follows: Wehea Modang, Long Gelat Modang, Kelijau Modang (Long Wai), Gaai Modang and Long Belah Modang. The Punan Kelai and Punan Mahkam remain as they are. Due to these groups having some rather marked socio-cultural distinctions, it would be potentially more confusing to give them the designation 'Modang'.

Each group is closely tied to specific riverine zones. The Wehea live on the Wahau and Telen rivers in the kecamatan (county) Muara Wahau. The Long Gelat inhabits the mid and upper Mahakam areas, situated among Bahau/Busang communities. The Long Belah people are located on the Belayan. The Long Wai people live on the Kelinjau river (thus, the dialect designation Kelijau Modang) while the Gaai (Gaai Modang speakers) live on the Kelai and Segah tributaries as well as along the lower Kayan river area. The Punan Kelai are known to live along the Kelai river while the Punan Mahkam live on the Segah river (Guerreiro 1993: 185).

On these rivers, Modang speakers generally opt to build their villages (*ekung*) on high river banks (getting from a boat on the river to the village generally requires an often steep, long ladder). They also prefer to build villages in locations where rivers and their constituent tributaries converge. These lowland areas tend to be flat, fertile flood zones where the shared village farmland (*lenih ekung*) is quite productive. Building here is also likely related to trade and travel ease as well as ease of defense in warfare. Here, they grow hill rice (*plaè*) in moderate-sized swiddens. Within a village, they may live in longhouses (*min*) consisting of three to six apartments each. Individual houses (*msow*) are also an option. These iron-wood constructed houses are often built on stilts to allow for flooding. This practice probably had defensive significance in the past as well. Two rows of houses with a roadway (*lan*) at its center are generally situated parallel with the river. Each row is a moiety. The row closests to the river is called the lower village/house (*dya' min)* and the one farthest from the river is the upper village/house (*lon min*). Most villages hold a population of around 200 to 600 people (Guerreiro 1993: 185-186).

In total, the population of Modang speakers numbers around 5,000. There are an additional thousand for Punan dialects. The speaker population is, by no means, divided equally among the various Modang language varieties. The Wehea group is, by far, the largest group with approximately 2,300-2,100 speakers (Guerreiro 1993: 185; Guerreiro 2002: 96). While disease, intermarriage, islamic conversion, and a low birthrate have contributed to a population reduction on aggregate, the Wehea group seems to have a trend of population growth (up 39% from 1,266 people in 1935) (Guerreiro 1993: 185). This suggests the possibility of ethnic (and, presumably, language) shift among the other Modang groups to Wehea and as well as other Dayak groups.

With the exception of the Punan (nomads) (Großmann 2017; Winzeler 2011), the Modang are considered 'middle-range' rice agriculturalists (Rousseau 2001, 2006) or 'rawah' cultures (Okushima 1999). This agricultural style means the village farming area resembles more a small collection of flat 'football fields', rather than the hugely expansive rice territories or terraces of industrial scale of, say, Java or Bali, (known as *sawah*), which often come to mind when thinking of Southeast Asia. Since the population density of the areas they inhabit are not very high, this kind of slash-and-burn agriculture works quite well without the threat of saturation or competition, even with allowing rather long (sometimes multi-decade) fallow periods (Guerreiro 1993: 186).

Rice plays an important role in the ritual life of the Modang—a ritual life which is rather rich. The 'Custom of Rice' (*Edat na' plaė*) is a very significant set of rituals and there are a great many taboos (*pli'*) to adhere to during the rice-growing season. Many of these rituals take place annually in early April. In addition to adat (custom) practices and its accompanying numenal belief, most Modang are Christian like most Dayak groups of Borneo (although there has been some Islamic conversion) (Guerreiro 1993: 185-186).

It should be noted that, historically, the Modang were the most violent and hegemonic groups of Central Borneo.

1.6.3.3 Ngorek/Murik

The Ngorek (or Ngurek, also sometimes referred to in the literature as 'Murik') are a group of at least 1,800[6], located on the upper Baram river area of Sarawak and minor areas in the adjacent North Kalimantan. Scholars believe that once there were potentially ten subgroups of Ngorek peoples (Sellato 2016; 1994; 1992). Today, there are five main Ngorek groups: Ngorek Lurah, Ngurek Bahau, Ngorek Pua', Ngorek Berap, and Nyibun (Jalong 1989; Soriente 2003;

[6] Population data for Ngorek tribes are really hard to come by and are very disjointed when one does.

Sellato 2016). In this study, these groups' languages are referred to as Ngorek/Murik, Hueng Bau, Pua', Merap, and Nyibun respectively. Culturally these groups are considered Kenyah, although their languages appear to be classified as Kayanic (cf. Smith 2017a; Soriente 2003; Sellato 2016). The Nyibun in particular, have almost completely assimilated into the Kenyahic culture of the Leppo' Ke' and Leppo' Ma'ut. While certain sound changes demonstrate its Kayanic linguistic roots, most of its lexicon has been supplanted by Kenyahic vocabulary (Soriente 2003: 276).

These groups were probably quite widely distributed around the upper Bahau river, having originally come from the Baram headwaters (Sellato 2016). Around 1700, these groups were forced to leave this area due to attacks from surrounding people groups. By 1750, they were probably in the Lurah and Bahau river areas (Sellato 2016).

Subsequently, further attacks by the various Kayanic-speaking groups in these new-found territories began. As a result of these attacks, many Ngorek groups were subsumed within the winning groups or scattered throughout the area. Some remained where they were, but were under some authority of the conquering groups (Sellato 2016).

As Kenyah tribes moved into the area, many of these Ngorek groups began to assimilate culturally with them. Note that the highly linguistically variegated Kenyah have cultural ties to the more linguistically homogeneous Kayan.

Early Ngorek groups seemed to have preferred swamplands (Sellato 2016: 120). These groups probably maintain so-called 'non-stratified' societies with group-internal, competitive status dynamics, predicated on individual wealth and prestige. Social structure including perhaps group identity or village affiliation within villages would have been quite loose along with there being no overarching structure among the various Ngorek villages and tribes. The lack of social

cohesion in addition to what appears to be a dearth of iron weapons and tools left these groups vulnerable in the face of other warring tribes (Sellato 2016).

Due to the dearth of iron tools, the Ngorek were probably not intensive rice farmers like their neighbors. Taro seems to take precedence over rice in many of these people groups, even today (Sellato 2016: 120). The Ngorek are also associated with funerary megaliths distributed throughout their areas of Central Borneo. These monuments are likely testaments to their internally status competitive way of life, with monument-building being signs of wealth and prestige (Sellato 2016).

1.7 Language varieties and genetic relationships

The Kayanic language family consists of three primary subgroups: (1) Kayan (or *Kayan Proper* as linguists have referred to it in the past), (2) Segai-Modang, and (3) Murik-Merap (Smith 2017a). The Kayan language varieties are currently divided into eight, following closely the river system upon which the speakers live: Baram Kayan, Rejang Kayan, Mahakam Kayan, Kayan River Kayan, Mendalam Kayan, Wahau Kayan, Bahau, and Busang. Segai-Modang consists of Wehea, Kelinjau Modang, Long Belah Modang, Long Gelat, and Gaai along with Punan Kelai, Punan Makham, and, probably also, Punan Segah. The Murik-Merap group consists of Ngorek, Hueng Bau, Pua' and Merap.

An early, preliminary attempt to organize the languages of Kalimantan, in this case by geography and reported similarity rather than comparative reconstruction, was Ray's *The Language of Borneo* (1913). Based on his own data and a review of previous literature related to various 'Kayan,' 'Bahau,' or 'Modang' related varieties, Ray's work organized what might later be termed 'Kayanic' languages this way. Again, note that Ray does not actually offer a 'Kayanic

subgroup' classification since his organization principles are not based on reconstruction methods.

Ray's Classification (1913)
Kayan
    Rejang Kayan, Uma Blubo', Uma Poh, Uma Baloi, Bahau, Bintulu Kayan
Murik
Modang
    Long Wai, Long Wahau, Long Bleh, Bahau
Pari

Ray offers data for the first for Kayan language varieties while citing the literature on connections with Bahau and Bintulu Kayan (probably a form of what we call here Rejang Kayan). He also believes that a group called the Pari might be related to the larger Bahau-Kayan-Mondang cultural group (Ray 1913: 12). The Pari are what are today supposedly considered the Maloh people (Wadley 2000), a group speaking the Tamanic 'Embaloh' language (Eberhard et al. 2021), although it should be noted that this group is geographically close to the Mendalam Kayan on the upper Kapuas river, near Putussibau. Perhaps, either Mendalam Kayan was mistaken in name for Maloh or there has been linguistic influence shared between the languages. We know at a baseline that there was contact between the Maloh and the Kayan (Wadley 2000: 97).

In Hudson's classification, with very limited data, Kayanic itself was subsumed within a larger Kayan-Kenyah family (1978: 29). Note that Hudson did not consider Merap a Kayanic language (or even a Kayan-Kenyah language), but placed it in another whole family altogether, the Rejang-Baram family (1978: 28). Some language names have been adjusted to match the naming conventions of this thesis. Here is the organization for Hudson's classification offered by Smith (2017a):

Hudson's Classification (1978)
Kayanic
- Western Kayanic
  - Baram Kayan, Rejang Kayan, Uma Poh[7], Busang, Ma'aging
- Lepak Aru Bahau-Muric
  - Lepak Aru Bahau, Long Hubung Bahau, Muric
- Long Gelat-Long Bentuk
  - Long Gelat, Segai, Long Wai Modang, Long Bentuk Modang
- Long Paka' Kayan-Penyabung
  - Long Paka' Kayan, Long Blu'u Kayan, Penihing, Seputan, Penyabung

Soriente is probably the first to offer a detailed classification, based on sound correspondence innovations. Below I have placed Soriente's classification (2008) with minor alterations, discussed during personal communication. Note that she also postulates a Kayan-Kenyah family.

Soriente's Classification (2008; with Personal Communication Update)
Kayanic
- Kayan
  - Uma Leken, Uma Laran, etc.
- Ngorek
  - Murik, Hueng Bau, Pua'
- Lebu' Kulit
  - Lebu' Kulit, Lebu' Timai, Uma' Kelep, Uma' Ujok, Nyibun

Blust (2010) offers the classification below. Blust seems skeptical of a Kayanic-Kenyah joint grouping, but seems fairly certain of a 'Kayan-Murik-Modang' group (2010: 52).

Blust's Classification (2010)
Kayanic
- Kayan
- Murik
- Modang

[7] Uma' Poh is sometimes called Uma' Po or Uma' Pu throughout the literature. Soriente (2013: 175) considers it a form of Baram Kayan.

Smith (2017a) offers the most up-to-date reconstructed classification of Kayanic, based on robust sound change correspondences. Below is Smith's classification; however, note that the language names have been altered to match the naming conventions used in this thesis.

Smith's Classification (2017a: 92)
Kayanic
- Kayan
  - Baram Kayan
  - Rejang Kayan, Busang
  - Bahau
  - Kayan River Kayan
  - Mendalam Kayan
- Murik-Merap
  - Murik
    - Murik (or Ngorek), Pua', Hueng Bau
  - Merap
    - [Punan] Merap (or Mbraa)
- Segai-Modang
  - Segai
    - Gaai, [Punan] Kelai
  - Modang
    - Kelinjau Modang, Wehea (or Wahau Modang), Long Gelat

Also, while the Ethnologue and other language inventories (Glottolog, etc.) offer classifications as well, because these classifications are more or less drawn from a combination of these cited scholars, their particular arrangements were left out of this section. Note how latter classifications appear to build upon earlier findings, despite early work having a dearth of data.

This study relies on Smith's classification, particularly for the subgroups within Kayanic. This is important because phonologically-informed lexicostatistics (what we will call here lexicostatistical/phonostatistics for short) will only be performed among the various language varieties within a given subgroup (i.e. within 'Kayan,' within 'Murik-Merap,' and within 'Segai-Modang') and not between them.

1.8 Motivations and methods

What does language diversification in Borneo indicate about the mechanisms of language diversification generally? This language-survey-as-case-study can be conceptualized as a kind of 'cartography of diversification,' uncovering fundamental aspects of the nature of how humans produce linguistic diversity. With a focus on intelligibility, language boundaries will emerge offering insight for theoretical as well as practical language development and policy concerns.

The study of dialectology is the study of boundaries. The natural ecology in which speakers use a language often creates boundaries between various language varieties. Rivers and steep hills or mountains may keep groups separate long enough to drive differentiation. However, most linguistic differentiation has social origins. The formation of separate group identities, social classes, castes and other phenomena of social differentiation becomes reflected in linguistic practice. As Romaine writes, all language variants are "part of a continuum in social and geographical space and time" (2000: 2).

Borneo, as a research area, has remained relatively untouched by serious linguistic investigation until quite recently. Very little is known about Kayanic in particular (Soriente & Inagaki 2012: 13). Apart from wordlists collected, here and there, throughout the colonial era, many parts of Borneo were seriously under-documented until about the 1980's (Adelaar 2010: 26). Even less scholarly work has been conducted in eastern Kalimantan (on the Indonesian side) where the majority of Kayanic speakers live. This area would hardly see any serious documentation work until the early 2000's (Soriente & Inagaki 2012: 3).

Undiscovered endangered languages abound in Borneo. In East Kalimantan's Bulungan Regency, Soriente has conducted documentation of Òma Lóngh (Uma' Lung), Lebu' Kulit,

Penan Benalui, among other (in this case, non-Kayanic) languages (2006), all languages unmentioned by major language inventories for endangered languages until relatively recently (UNESCO; Glottolog; Catalogue of Endangered Languages 2021; Eberhard et al. 2021; Sorosoro). Most of the Kayanic languages are endangered: Kayan River Kayan [2,000 speakers], Mahakam Kayan [1,300], Baram Kayan [4,150], Busang [3,000], Mendalam Kayan [1,500], Rejang Kayan [3,030], Segai [2,000], Punan Merap [200], Wahau Kayan [500], and Pua' [500] (Eberhard et al. 2021; Soriente & Inagaki 2012: 4). There are also signs of language shift among certain Modang dialects, although overall Modang is considered 'vigorous' (Eberhard et al. 2021). This is probably due to certain dialects remaining quite strong. For example, while a more systemic assessment is needed, just based on my observations, among the Wehea, intergenerational transmission is quite strong and the language seems relatively quite stable.

The research design follows a 'phase' structured procedure for collecting the various kinds of data. Considering that researchers of Kayanic collected the preponderance of this massively disjointed data on the language situation between an wide span 1980-2006, leaving an approximately 13 year gap, during which serious changes to the language ecology may have occurred (migration in or out, social organization, technological advancements, language policy changes, change in national political regimes, ethnic identity shifts, natural occurrences, economic reconfigurations, etc.), it is worth pursuing a 'pilot pre-phase' to get the proverbial 'lay of the land.' This pre-phase warrants the rapid appraisal pilot where a review of the literature (Chapter 2) and informal conversations are conducted in order to gain a general, updated picture of the language varieties in question and their language ecology.

Phase 1 involves quantitative-oriented wordlist collection methods. The wordlist data then undergoes quantitative analysis in the form of lexicostatistical/phonostatistical approaches.

Linguists using lexico-statistics usually consider language varieties that overlap by a percentage between 81-100%, variants of the same language. A 28-81% overlap usually constitutes distinct languages within a common family. Considerations of social propinquity are excluded here (Romaine 2000:5). More on the criteria used in this study in the methodology section below.

The results from these analyses will provide language boundaries, based purely on structural grounds (lexical composition and phonological structure). It should always be kept in mind, however, that linguists have long established that speech communities and the intelligibility among them is more than a matter of 'naturally' understanding words and sounds. Ecological phenomena can bring about changes in intelligibility. An example being that of *acquired intelligibility* whereby speakers, through contact, learn the differences among their language varieties and through this experience maintain a certain level of intelligibility that the lexical composition and phonological structures would not inherently allow (Anderbeck 2018; Blair 1990; Grimes 1995). In Borneo in particular, the language varieties of upstream groups are readily intelligible by downstream communities, but not the other way around due to acquired intelligibility, probably engendered from the social/societal (and geographic) ecologies (cf. Wood 2000, quoted in Anderbeck 2018: 248). Statistical methods would never reveal this. This constraint must always be kept in mind.

After data collection, the quantitative methodology, having produced a language intelligibility percentage table, is then used to establish a dialect map. Using Anderbeck's (2018) 'clustering approach,' language varieties are sorted into 'superclusters,' 'clusters,' and 'subclusters' in order to reach our practical goals of predicting intelligibility. Note that these clusters are also useful for one of our practical aims—language planning (language policy and MTB-MLE programs).

The map interpretation involves *convergence* with themes, identified as ecological niche factors in the ethnographic record. This 'ethnology' (Chapter 3) is a 'comparative investigation of people groups, observing their cultural patterns as they converge and diverge.' All the ethnographic material is compared and contrasted among the different groups, yielding the factors, leading to language contact, segregation or isolation, et cetera that impact language use, shaping the intelligibility outcomes. The map is interpreted from the ecological niche factor themes delineated in the ethnological description.

"Explaining" intelligibility constitutes the language-survey-as-case-study in the present 'cartography of diversification'—a case study enlightening the theoretical understanding of the exact processes and mechanism of language differentiation. Adaptational tendencies in these processes and mechanisms support the linguistic niche hypothesis—that language variation (and implied change) are not exclusively neutral, but are shaped by the larger ecological environment in which languages are spoken. More on the methodology in Chapter 4 below.

1.9 Goals of this study

The three main goals of this study center around providing new and up-to-date information on the linguistic situation in an under-described area of the Austronesian world. These goals are as follows:

1) To map the language varieties within the Kayanic subgroup.
2) To establish language boundaries, based on intelligibility.
3) To provide new linguistic data and initial-stage documentation of the Kayanic languages.

Reaching these goals allows us to address other practical and theoretical objectives, concerning the particular outcomes of this study. The practical objectives are:

1) To propose modifications to ISO 639-3 codes, rendering them more accurate.

2) To propose language and educational policy suggestions, based on intelligibility.
3) To provide geo-spatial data, locating the position of these languages.
4) To update language inventory maps (Ethnologue; Glottolog; Endangered Languages Project; UNESCO Atlas).

Four major theoretical objectives are accomplished upon the results of this study. This objectives are:

1) To reify the established integrity of the Kayanic family.
2) To provide a description of the evolution of Kayanic, including the problematic Modang subgroup.
3) To demonstrate the potential for synchronic dialect mapping/atlases to empirically inform theories of language variation and language evolution.
4) To put forth a case, demonstrating the dynamics of language diversification.

Accomplishing these goals are significant not only for building a deeper understanding of the languages of Borneo (and, therefore, Austronesian languages and linguistics), but also for demonstrating the contribution sociolinguistic survey work and field linguistics, more generally, to studies in the evolution of linguistic diversity.

1.10 Structure of this text

Chapter 1 'Evolutionary Cartographies of Language Diversification' introduces the topic of the evolution of language diversification from a synchronic perspective, that is, arguing how examining language maps and dialectology data can lead to important insights into the mechanisms that drive language diversity. This chapter also outlines the motivations and goals, both theoretical and practical, for pursuing a language survey of the Kayanic languages of Borneo. In accordance with Mufwene's (2001: 2) 'internal' and 'external' factors that motivate

language change, rooted in ethnography, a brief overview of the history, social structure, and culture of the Kayanic people is given, exposing the environment factors that are at play among the language groups. These ethnological notes, outlining details about the physical (natural and geographic), socio-cultural, and economic ecosystem within which Kayanic languages developed and diversified, are at the analytical center of this study. It provides the very environmental factors which would bias certain language varieties to change in one direction as opposed to another. Simply put, it outlines the ecological conditions to which these language varieties have adapted, nudging intelligibility to mutually converge with other dialects or diverge from them.

Chapter 2 'Ecology of Diversity' provides a review of the literature on the evolution of language diversity in the context of the theory of 'language ecology'. With an eye toward environmental niche factors in language diversification, the theory of 'language ecology' provides a solid framework for grounding the structure of the Linguistic Niche Hypothesis claim.

"Methodology & Fieldwork: Research on the Language Ecology of Borneo" (Chapter 3) describes the methodology, employed to capture an understanding of the situations of intelligibility as well as a moment of reflexivity on the ethical and experiential aspects of fieldwork with the indigenous peoples of Borneo.

Chapter 4 lays out the data and the results of the sociolinguistic survey. The data and analytical framework are brought together in the lexical similarity percentages and the linguistic map of the Kayanic languages, shedding light on the more general processes and mechanisms of language differentiation from the perspective of Borneo. A linguistic map with commentary serves as the more concrete 'finished product' of the language survey. Language similarity percentages and various visual formats demonstrating those relationships are also included.

Chapter 5 ‘Conclusions: Progress and Prospects’ concludes the language survey by reflecting on the achievements and limitations of this study all while making suggestions for future research. Implications for education are discussed, describing the applied linguistic implications of this study and the practical implications of the map for language policy, endangered languages and their documentation, and mother-tongue based multilingual education (MTB-MLE). These topics are discussed as per their potential benefits for the indigenous language communities of Borneo.

1.11 Summary: Mapping Kayanic and mapping language diversification

This exercise in the ‘cartography of language diversification’ records the synchronic relations of language contact as suggested by patterns of intelligibility in order to begin to respond to that greater, more fundamental question that linguistics seeks to answer: Why are there so many languages? The results of this sociolinguistic survey reveal to linguists key insights in how humans produce language diversity (and probably have done so throughout [pre-]human history). The Kayanic languages of Borneo, in addition to its practical utility for language conservation and development purposes, provide a robust case for demonstrating diatopic relations of language contact and differentiation from the everyday linguistic practice of intelligibility and multilingualism. The ecological perspective on contact-induced diversification delineates the socio-economic, political, and environmental factors at play in the nexus of these contacts and serves as the theoretical conduit through which I describe these relations. But what is the theory of ‘language ecology’? I now turn to describing this framework.

CHAPTER 2

Ecology of Diversity: A Review of the Literature

> For however many things have a plurality of parts and are not merely a complete aggregate but instead some kind of a whole beyond its parts, there is some cause of it since even in bodies, for some the fact that the there is *contact* is the cause of a unity/oneness while for others there is *viscosity* or some other characteristic of this sort.
>
> —Aristotle

2.1 Why are there so many languages? The language diversification question

Language diversification is a matter of *diffusion* and *divergence*—that is, language innovations emerge and spread by being acquired as new features of certain language varieties and the emergence and spread of those features cause those language varieties to develop on different trajectories. This divergence and maintenance of recognizably discrete entities (i.e. "they speak a little differently over there") is the creation of language variation. The question, then, must be, how do languages diverge (or converge, for that matter)? What are the mechanisms that drive linguistic divergence (and, therefore, language diversification)?

Investigating the mechanisms that drive language diversification starts with the realization that not all structural variation in language occurs randomly. The role of randomness is a well recognized component of language change (Karjus et al 2020, Newberry et al. 2017, McWhorter 2011, McWhorter 2016, Reali & Griffiths 2010, Bentley et al. 2004). After all, "chance is one of the keystones of language change" (McWhorter 2016: 74). Certainly, there are random innovations that, once propagated throughout a language community, become full-blown changes in a given language variety. For example, a random change might be one due to linguistic 'reanalysis'. The original word for 'a protective cloth worn while cooking' was 'a napron' which, the determiner and the onset of the word being homophonous to another form of

that determiner ‘an,’ was reanalyzed as the words ‘an apron’. In Haitian creole, the word ‘friend’ *zami* derives from setting the word break before the phonological liaison on the article *les* instead of after in the often said *les amis* [lɛzami] in spoken French. Another example would be the co-occurrence of variant forms in the English past tense [learnt/learned, burnt/burned, dreamt/dreamed, etc.] (Newberry et al 2017).

The fact of the very formation of these innovations and the fact that these innovations were propagated and have become so thoroughly integrated into the English and Haitian creole speech communities are all happenstance—examples of random change. Upon analogy of random genetic drift (‘stochastic drift’ or ‘neutral mutation’) in population genetics (Kimura 1985), these random ‘copy errors’ in some communities and their propagation to subsequent speaker generations leads to language diversification. Many linguists commonly hold that much of language diversification derives from the “gradual accumulation of these random changes” (Lupyan and Dale 2016), seeing that “chance is baked into a great many transformational processes (of which language change is one), such that to refuse the role of randomness is to not truly understand them” (McWhorter 2016: 74). While it is certain they play a role in language variation, linguists are still debating the extent to which these random changes account for much of language variation.

What is clear, however, is that many significant contributions to language variation derive from non-random factors—some linguistic, some extra-linguistic (Sapir 1921, Jespersen 1922, Andersen 1990, McMahon 1994, Croft 2000, Mufwene 2001, Baxter et al. 2006, Mufwene 2008, Lupyan & Dale 2010, 2016, Steels & Szathmáry 2018, Hua et al. 2019, Karjus et al. 2020). In addition to language variation being motivated from processes within a given variety or from contact among or between varieties, factors beyond language such as the physical environment,

social environment, and the technological setting among others have all served as sources, driving language diversification (Lupyan and Dale 2016). Everything from the phonetics and phonology to the morphology, syntax, and information structure of the language seems to be susceptible to environmental influence and shaping (Everett et al. 2016, Lupyan & Dale 2016, Bentz 2018). These environmental influences shape new language varieties, creating the diversity seen in the world's 7,000 languages today.

2.2 The ecology of language

The framework most apt to describe and explain the environmentally-driven aspects of language change is that of the language ecology. This framework is also known as the 'ecology of language,' 'linguistic ecology,' 'language and ecology' studies and is sometimes lumped into the category of 'ecolinguistics' as well (Stanlaw 2021, Fill & Mühlhäusler 2001). It should be stated, however, that ecolinguistics often takes a divergent meaning, referring to studies in a kind of linguistic environmentalism, examining the ways which the natural environment and social perspectives on that natural environment impact language use as well as how, in turn, that language use and its social discourses impact the natural environment (Stanlaw 2021). This study abandons this aspect of the nebulous definition of this "still emerging" (Chen 2016) field of language ecology studies, in favor of descriptive precision.

The origins of this framework begin with Haugen's ecosystem analogy (1970).[8] This helped describe how languages were connected with each other and the locations in which they were spoken. One of the central concerns of language ecology studies is the emergence of language diversity. Since then, many have written on language ecology from a wide variety of perspectives (Bastardas-Boada 1996, 2002, 2002, 2014, 2018, Calvet 1999, Fill & Mühlhäusler

[8] Precursors to Haugen's work include Sapir (1912) and Voegelin, Voegelin & Schutz (1967).

2001, Hornberger & Hult 2008, Hult 2009, 2010, 2012, Mühlhäusler 1992, 1995, Mufwene 2001, 2008, Sánchez Carrión 1985, Skutnabb-Kangas & Harmon 2018, Maffi 2005, Steffensen & Nash 2007).

To describe language diversification, one of the most salient models of employing the language ecology framework takes an evolutionary approach (Mufwene 2008, Mufwene 2001). Modeled on the principles of population genetics (Mufwene 2008, 2001), this language ecology framework allows linguists to identify which aspects of the environment are interacting with language change and in what ways. This evolutionary approach's central claim is that adaptation to certain ecological niches or environments drives the systematic ways in which different features emerge and their persistence in certain lines of descent. The adaptive process involves (1) the emergence of variant features, (2) competition among those variant features, and (3) the inheritance of those variant features. If a given set of features 'win' the competition among language varieties as well as within one of them, this 'winning feature bundle' comes to constitute a language variety as an entity, perceivably discrete from others where it has prevailed. This tripartite process is inherent to any evolving system (Lewontin 1970) and is a key component of language as a complex adaptive system (Mufwene 2001).

2.2.1 The Linguistic niche hypothesis

While the linguistic niche hypothesis is at a nascent stage in scientific inquiry, several studies have been conducted on the topic (Lupyan & Dale 2010, 2015, 2016, Dale & Lupyan 2012, Lewis & Frank 2016, Wray & Grace 2007, Koplenig 2019, Tinits et al. 2017, Benítez-Burraco & Kempe 2018, Antunes et al. 2020, Hill et al. 2017, Bentz & Winter 2013, Kempe & Brooks 2018, Sinnemäki 2020; and 'language adaptation' more generally, Reali et al. 2018, Everett et al. 2015, 2016, Maddieson 2018, Bentz 2018) To reiterate, the linguistic niche hypothesis is the

proposition that languages adapt to their environments and that this adaptation molds language variation and change.

Another term for ecological niches or linguistic niches is *Eco-Linguistic Niche (ELN)* (Antunes et al. 2020)—although this term is most often applied in the context of what we call the physical niche here or the physical geographic environment (climate, rainfall, elevation, humidity, etc). This term is defined as "the range of environmental conditions present in the territory of a population speaking a specific language or group of languages characterized by common language traits" (Antunes et al. 2020: 1).

2.2.2 Ecology and evolution: Linguistic niche adaptations and language diversity

Mufwene (2001) divides the language ecology into the 'internal ecology'—the linguistic features in interaction in a given person's language ability—and 'external ecology'—the linguistic surroundings of that person. Steffenson and Fill (2014) along with Lupyan and Dale (2010, 2016) delineate several principle language environments. In addition to these, scholarly work on a 'developmental niche' is in its initial stages (Pleyer 2020, Hartmann & Pleyer 2020, Benítez-Burraco & Kempe 2018, Sinha 2009). Although not explicitly named thus, what I will call here a 'bio-corporeal niche' and a 'genetic niche' appear to arise from the literature as well (Dediu et al. 2017, Dediu & Ladd 2007, Maddieson & Coupé 2015). Therefore, I identify ten main language environments or ecological niches: (1) physical niche, (2) social niche (3) cultural niche, (4) symbolic niche, (5) cognitive niche, (6) linguistic niche, (7) technological niche, (8) developmental niche, (9) bio-corporeal niche, and (10) genetic niche. This innovation in synthesizing key linguistic niches will be useful for future studies in language ecology. While this study will focus primarily on the physical, social, cultural, and linguistic niches, the ten ecological niches proposed here are described below.

2.2.2.1 Physical niche

The physical niche includes the physical geography and its barriers such as mountains, rivers, swamps, dense forests, oceans, and deserts among others (Lupyan & Dale 2016; Steffenson & Fill 2014). The transportation system (or lack therefore) might be considered part of the physical niche as well. This is probably the most well-attested niche and certainly the most widely researched in the context of evolutionary linguistics and language ecology studies. While studies on the link between the physical environment and language in modern linguistics date all the way back to Sapir (1912), the most recent studies aim to provide empirical evidence for the connection.

The physical niche can be said to mold language phonetically through acoustic adaptation (Lupyan & Dale 2016, Maddieson & Coupé 2015). For example, although prosodic registers of the spoken languages within those speech communities (30 documented in all), whistle languages are all systematically located where the rough terrain (and the inevitable lengthy distance among people due to it) constitutes a biasing force in their creation. These languages appear in many, diverse areas such as Mexico, the Canary Islands, New Guinea, Ethiopia, Nepal, and Turkey among others (Meyer 2021). In all cases, these languages arose in difficult-to-traverse landscapes which necessitated long-distance communication. As Lupyan and Dale write, "[the] sound system has adapted to its environment" (Lupyan & Dale 2016: 653).

In addition to adaptation to the physical environment, languages are shaped by the socio-cultural setting in which they are used. While the 'social' and 'cultural' are often described together, I have divided these into separate niches since the effects 'cultural' as a 'systems of learned, intergenerational behaviors and traditions' may differ from those due to a given group's

social organization. Inevitably one affects the other--an assumption of language as a complex adaptive system--however, here I will treat them separately.

2.2.2.2 Social niche

The social niche includes the type of society and social structure of a given speech community (Lupyan & Dale 2016; Steffenson & Fill 2014). There are strong, patterned differences between societies that are small and isolated with limited communicative domain specialization compared with those that are large with many L2 speakers and many specialized internal subgroups.

Wray and Grace (2007) refer to the languages of these two types of social structures as *esoteric* (or intra-group) languages and *exoteric* (or inter-group) languages. In investigating the grammatical correlations between each language type, in a study of over 2,000 languages, Lupyan and Dale (2010) discovered that exoteric languages overall had simpler morphology than esoteric languages. Exoteric languages tend to be more isolating, avoiding the agglutination found in small-scale society languages. Word order seems to take precedence since, if there is case marking, these case systems are generally of a more reduced in form. Verb conjugates are simpler and grammatical gender distinctions tend to be reduced. The influence exerted by social structure (predicated on the number of L2 speakers and isolation) is strong and apparent at multiple levels of grammatical analysis.

Included within the social niche, or rather a subtype within the social niche, is what one might call the economic niche. This includes economic motivations for the arrangement of social structures within a given society.

2.2.2.3 Cultural niche

Cultural niche refers to the effect of cultural patterns—material goods, but moreover a system of socially shared ideations and beliefs—can have on the structure of a language. While this example is very controversial and the evidence of some of its larger claims (i.e., the lack of recursion) is still hotly debated [pro: Futrell et al. (2016)[9], Oliveira & Everett (2010), Everett & Oliveira (2010) and on 'numbers,' Gordon (2004); Frank et al. (2008) & con: Nevins et al. (2009), Sauerland (2010)], Everett's (2005) work with the Pirahã and their cultural construct of the 'immediacy of experience' [exemplified in the use of *'ibipío* 'arrive in/leave out of the immediate experience'] clearly lead to a unique structuration of the language. The lack of color terms, kinship terms, and numbers but more importantly the lack of tense, reduction of pronouns, and perhaps the lack of (or reduction of) syntactic embedding shows the power of culture on language.[10]

2.2.2.4 Symbolic niche

The symbolic niche includes the adaptive environment of the symbols and values of a given people group (Steffenson & Fill 2014). As Haugen (2001: 57) mentions it, language ecology not only consists of language systems in interaction, but also that "part of [the] ecology is sociological: its interaction with the society in which it functions as a medium of communication." One of the most conspicuous scenarios demonstrating how languages adapt to

[9] This computational analysis of Dan Everett and Steve Sheldon's Pirahã corpora, after having searched for center-embedding, sentential complements, adverbials, complementizers, embedded possessors, along with basic conjunctions and disjunctions structures, found that a grammatical analysis of Pirahã as having no recursive processes to be quite plausible.

[10] I should state here that although I find this argument compelling in terms of the power of culture on the structuration of language, even if the tall claim that the language does not have recursion is valid, this does not 'disprove' the recursion is a requirement for the formation of language in the first place. As part of the Mura language family, if other varieties possess recursion, it could easily be that recursion was required as part of the capacity that made Pirahã possible, even if it is subsequently abandoned as a daily communicative tool. Thus, recursion in this way could easily still be an inherent feature of the human language faculty even if the Pirahã do not use it.

the symbols and values of a given society are how various societies maintain religious registers and taboo languages. These languages capture notions of 'the sacred' and 'the profane' imbued within each society. In many Australian aboriginal languages, there are so-called 'mother-in-law languages' whereby speakers adopt a special set of vocabulary to engage with in-laws of the opposite sex (Dixon 2010: 58-59). These languages are clearly wrapped up in the symbolic system of language communities which promote the value of distance among these relatives, if not outright complete avoidance as a social standard. In addition to this, death and naming practices as linguistic practices of symbolic value are a worldwide language ecology phenomenon.

2.2.2.5 Cognitive niche

The cognitive niche refers to the greater system of knowledge and mental abilities beyond language and how language is integrated within them such as perception, spatial knowledge, and affect or emotion (cf. Steffenson & Fill 2014). Language, in many ways, is "an extension of numerous domain-general cognitive capacities such as shared attention, imitation, sequential learning, chunking, and categorization" (Beckner et al 2009: 17). This is probably because language has been molded to fit within the domain-general cognitive powers of the human brain, rather than the brain adapting or being remolded to integrate a newly emergent language structure within it, as has been often assumed (Christiansen & Chater 2008). Put more precisely, language has adapted to the brain, not the other way around, and in doing so language is shaped by the human brain's other mental states and capacities.

In the sound system, evidence of this appears in 1970s studies on the McGurk effect (McGurk & MacDonald 1976). When the syllable [ba] is played while watching lips articulate

[ga], test subjects perceive the more articulatorily medial syllable [da], showing that language relies on multiple channels of perception for hearers to arrive at their perceived conclusions.

In the morpho-syntax of the Cha'palaa language, olfaction is grammatically encoded (Floyd et al. 2018). A closed class of 'smell quality' roots are obligatorily combined with certain smell (*-dyu*) or taste (*-mbu*) classifiers, creating 15 documented smell terms. For example, the underivable root sen- is combined with the *-dyu* classifier, arriving at '*sendyu*,' meaning "fishy or metallic smell[ing]," used to refer to smells, exuding from raw fish or blood (Floyd et al. 2018: 180-181). This is a clear case of language adapting to the olfactory perception component of human cognition.

Semantically, 'metaphorical abstraction' seems to be supported by a more perceptually concrete schema (Pinker 2010, Jackendoff 1978, Lakoff & Johnson 1980, Talmy 2000). Situations such as '*John comes from school everyday at the same time*' appear to scaffold more abstract scenarios such as '*The word 'déjà-vu' comes from French*.' This metaphorical motion verb phenomenon (here specifically related to the construction '*come+from*') "reflect[s] an ability of the human mind to readily connect abstract ideas with concrete scenarios" (Pinker 2010: 8997).

2.2.2.6 Linguistic niche

The linguistic niche, as part of the internal ecology, includes the number of languages in interaction within an individual's cognition. Individual bilingualism or multilingualism plays a role in the structuration of entire communities, particularly if entire communities are multilingual. In this way, languages are essentially in contact, embodied within individuals. This includes their mother tongue in addition to any other learned or acquired languages. Community level multilingualism (or lack therefore) also plays an important role, but as an external factor in

the language ecology. The number of languages spoken in the surrounding community, can influence practices in any one of the given languages present (cf. Mufwene 2001; Steffenson & Fill 2014).

2.2.2.7 Technological niche

The technological niche refers to tools that use language (writing, texting, etc.) (Lupyan & Dale 2016). Languages adapt to different technological niches. The restructuring of language to the technology of writing (as opposed to speech) provides a good example of adaptation to 'linguistic technologies.' At a baseline, there is generally the development of a formal register for writing, not found in (or normally used in) everyday oral speech (Lupyan & Dale 2016: 654). Not only are there systematic distinctions between the register types (and/or its often accompanying stratification) of language in primarily oral as opposed to literate speech communities, the substance of those developed registers is altogether different as well.

The fleeting nature of orality seems to create a 'not or never bottleneck' for the vocabulary and grammatical structures used in a given language (Christiansen & Chater 2016: 2). Because oral speech is processed in real time, there seems to be "strong selection pressure against words and grammatical constructions that cannot be easily parsed in real time" (Lupyan & Dale 2016: 654). Information processing of oral speech fades significantly within 50-100 ms after the utterance is made (Elliot 1962, Remez et al. 2010). For sign languages, information processing similarly deteriorates significantly around 60-70 ms after utterance (Pashler 1988).

Due to the impermanence of oral language, the differing communicative domains (captured notably in registers and jargons), and perhaps that writing affords a kind of set of mental "training wheels" accounts for the fact that non-chirographic languages tend to avoid complex hypotactic syntactic constructions (Dąbrowska 2015: 229-231). This is supported by

documentation of languages, either rarely employing or altogether lacking subordination (Aronoff et al. 2008, Everett 2005, Kalmár 1985, Mithun 1984; Szczepaniak 2015).

Apart from discourse/pragmatic and morphosyntactic adaptation, Castro-Caldes' et al. (1998) PET scan study reveals that literacy affects the oral phonological capabilities of speakers to produce novel lexical items and that that these capabilities rely on different neurological structures that are unengaged in illiterate speakers.

2.2.2.8 Developmental niche

Language may also be said to adapt to a developmental niche. This niche in addition to the genetic and bio-corporeal niches are at a relatively nascent stage in scientific investigation. Pleyer (2020: 346) argues that "pretend play served as [a] developmental niche which scaffolded the emergence of complex forms of language, namely the development of complex constructions for negotiating and sharing perspectives." More research is needed to further clarify how language adapts to the ontogenetic environment of early child development.

2.2.2.9 Genetic niche

The adaptive relationship between genetics or a genetic niche and language has especially been fruitful from the perspective of speech and the sound system as well as linguistic modality. Since the work of Cavalli-Sforza in the 1980s and 90s (1997, Cavalli-Sforza et al 1988), language families have been shown to correlate with human genetic populations following what would be expected from prehistoric migrations. Chomsky (2005: 6) proposed a "genetic endowment" for language as one of his 'three factors of language design' which would be "nearly uniform for the species." Tying language, and especially language variability, to genetics directly still eluded science at this point. In 2001 and subsequent work for decades to come, the groundbreaking discovery of the FOXP2 gene (Lai et al. 2001: 519) provided a glimpse into the

genetic origins of language. In the KE family, members with normal range IQs struggled with specific language impairment (SLI), causing severe speech and language pathological deficits. Since the early 2000s, many genes such as FOXP1, CNTNAP2, TECTA, MTRNR1, MYO15A, GNPTAB, GNPTG, NAGPA, ROBO1 among over 30 others have been implicated in some aspect of speech, hearing, or linguistic behavior (e.g. stuttering, hearing, dyslexia, etc.) (Dediu 2015, Fisher & Vernes 2015, Graham & Fisher 2015).

The work of Dediu & Ladd (2007, 2011) endeavored to finally tie genetic frequencies in various human populations to variation in language structure. Two brain growth genes ASPM (MCPH5, 1q31) and Microcephalin (MCPH1, 8p23), are theorized to cause a learning bias in certain human populations whereby under the right linguistic conditions they might adopt tonal languages. By languages adapting to the genetic variability in the human population, we might see the rise in linguistic diversity present today. Wong et al. (2020) provides direct evidence for tonal adaptation in languages where populations carry the ASPM gene. More research on language structure and genetics are currently underway (Hernandez et al. 2015, Progovac et al. 2018, Wexler 2017).

It should be noted that none of the work in language diversity and genetics implies a deterministic relation between language and genes.

2.2.2.10 Bio-corporeal niche

While I probably could have considered the genetic niche under the bio-corporeal niche, it seemed best that they were described apart due to the complexity of genetic expression. The bio-corporeal niche is the effect of the human body—often exemplified through changes to and/or variation in it—on the structure of language. There is significant evidence that the human bite as molded by food production and consumption in a given society (i.e., agricultural foods

versus foods in hunter-gather societies) biases the appearance of certain phones within a language. Blasi's et al. (2019) experimental models show that 30% less muscular effort is required in the production of labiodental phones when speakers have an overbite or overjet bite and these bite structures reduced the anatomical distance between the teeth and lips by 24-70%, increasing the chance of accidental labiodental phone production. Overbite and overjet configures are more prevalent in agricultural societies because of the food they eat compared to the more prevalent edge-to-edge bite found in hunter-gather societies. This leads us to conclude that the human body, in its anatomy, can bias the structure of a given language, leading to linguistic diversity.

2.2.3 The dynamic interplay of ecological niches

It should be noted that while I have described these niches as discrete categories, as Steffensen and Fill (2014: 7) writes, these distinctions mentioned here are "not be taken as a rigid either/or-separation of different ecologies" and that each "distinction captures different perspectives on the environment of language." Zooming in on a single niche within the "constellation" of the overall 'ecosystem' at any given point in time should not suggest that the others are somehow insignificant or epiphenomenal (cf. Steffensen & Fill 2014: 7).

Care must be taken when entering this line of inquiry. The idea that language adapts to a host of community-specific niches should not imply anything fixed about the nature of those communities or their languages. Overly deterministic approaches to human diversity (with linguistic diversity beginning a necessary part of it) once led to theories promoting scientific racism. Spurious fields like eugenics aimed to reduce human diversity to genes and evaluate the status of people groups accordingly. The Boasian revolution in the social sciences corrected this error with adaptation at the heart of its explanation—people modify themselves to more

successfully fit within the environments they inhabit. I suggest a similar approach to language. Avoiding all deterministic tendencies, the state of these niches at any particular moment in time are but 'nudges' that bias language structure probabilistically toward certain pathways over others.

I conceptualize the sum of these niches as synthesized in the figure below:

Figure 11. The Niche Structure of the Language Ecology

Internal Ecology (I-Language)		External Ecology (E-Language)
Linguistic Niche (Mother Tongue Internal)		Linguistic Niche (Communal Multilingualism)
		Physical Niche
		Social Niche
		Cultural Niche
		Symbolic Niche
Linguistic Niche (Individual Multilingualism)		Cognitive Niche
		Technological Niche
		Developmental Niche
		Genetic Niche
		Bio-Corporeal Niche

These ecological environments guide the evolution of language diversification, leading to greater linguistic diversity in the world. It should be noted, however, that while this has led to greater diversity on a global scale (as is its tendency), these same processes can lead to reduced diversity in particular corners of the globe at particular times (e.g. Koineization or sprachbund). The factors that compose each environmental niche drive language divergence. Taking a bio-evolutionary approach to describing these processes at the level of the "divergence of dialects" reveals the greater mechanisms at work in the "initial stages of linguistic divergence" (Honkola et al. 2018: 2). Dialectology, especially from work with language distances/similarity percentages (Honkola et al. 2018), but also language maps (Haas 2010) can be used to observe these mechanisms. This study will focus on how the niche factors from these language environments drive the divergence and diversification of the Kayanic languages of Borneo as

measured through calculated language similarity expressed through the phonological structure and lexical composition of these languages and the resultant levels and geographic patterns of intelligibility. Indeed, I am proposing that a 'cartography of diversification' of these little studied Kayanic languages will provide insight into the mechanisms of language diversification more generally.

2.3 Dialectology: A Synchronic Laboratory of Language Evolution

As previously mentioned, one way to gain an understanding of the mechanisms of language diversification is to observe language divergence 'in vivo.' This evokes the field of dialectology. Upon analogy with how biology describes speciation, dialectologists aim to take a "microevolutionary approach to studying the initial stages of linguistic divergence, i.e. the divergence of dialects" (Honkola et al. 2018: 2). Dialectology (and particularly its more quantitative manifestation, *dialectometry*) as a field can be thought of as a synchronic laboratory, with apt tools for describing and explaining language adaptation at the glossogenetic level.

Dialectometry has a history (at least, half a century long) of demonstrating how quantitative linguistic distances are shaped by the physical environment in which speakers live and interact. Jean Séguy (who coined the term 'dialectométrie') revealed in his early work on the dialects of Gascogne in southern France that geography can have a clear effect on language variation, as represented by lexical differentiation. It should be noted that Séguy too relied on language map data in his trailblazing work (1971: 337-338). Today's dialectometric studies do the same, although there have been methodological and technological improvements (Heeringa & Nerbonne 2001; Wieling et al. 2011).

In their study of Dutch, Wieling et al. (2011) show that the linguistic distances are molded by both geographical and social niche factors. Geographical distance from the Dutch

center of political and economic power showed successive patterns of lexical diffusion with the most peripheral language varieties deviating the most from standard Dutch. Geography appeared to be the most impactful factor in this study; however, social factors seemed to also play a role. The population size and average age of the population were the social niche factors at play among Dutch dialects (Wieling et al. 2011: 11). Population size and structure will play an important role in understanding the Kayanic language ecology. As was mentioned earlier in this review, although patterns of diffusion are indispensable as they establish geographical and implicative coherence (See Chapter 1, Section 1.5), demonstrating the mechanism of divergence is also essential to explaining linguistic niche adaptation, specifically, and language diversification in general.

In a study of Finnish, Honkola et al. (2018) aim to formally test hypotheses of language adaptation by examining how ecological niche factors bias the process of language differentiation at the dialectal level. Here, they take a "biological microevolutionary approach in order to study and understand the first steps of linguistic divergence" (Honkola et al. 2018: 9). In addition to sheer geographic distance and local history effects, physical environmental differences (subtle as they may be between different municipalities in Finland) and the socio-cultural practices that arose from people adjusting to those differences had a heavy impact on the differentiation and divergence of various dialects. Note that the physical niche and cultural niche factors appear to have an effect independently of the other effects mentioned (Honkola et al. 2018: 6). In one particularly important point, Honkola et al. suggest that "differences in cultural adaptations, such as subsistence strategies arising from variation in the environmental conditions, could have acted as non-physical barriers and limited contacts between groups" with language divergence arising from this isolation (2018: 8). This point reminds us that while

language diversification is a matter of diffusion and divergence, diffusion and divergence (or convergence, for that matter) among language varieties is a matter of language contact. Language contact is at the heart of this study's description of language adaptation and variation.

In a study of Japonic language varieties, Lee & Hasegawa (2014) show that the ocean as a natural barrier forms part of the physical niche that has driven language diversification. They refer to this process as *allopatric speciation*. The concept of allopatric speciation ('allopatric' from Greek, meaning 'other homeland') in evolutionary biology refers to a model whereby "physical isolation creates an effective barrier to gene flow" (Freeman & Herron 2007: 612). In linguistic terms, allopatric speciation might be taken to mean communicative isolation leading to ever decreasing intelligibility as the flow of lexical innovations (or repurposing) and phonological feature innovations are impeded. This, over time, would lead to discrete unintelligible clusters or what we informally refer to as 'different languages'. These findings apparently also hold for Ainu language varieties—the indigenous languages of Japan (Lee & Hasegawa 2014).

In addition to these studies (and the German study (Haas 2010) mentioned in the introduction showing how an investigation such as this is even possible), Montemagni et al.'s Tuscan *Gorgia* study (2013), demonstrates how focusing on *synchronic* dialectometric and language mapping data can index diachronic change, without ever touching diachrony per se (i.e., through historical-comparative reconstruction, real-time studies, etc.).

Dialectology has much to offer to studies in language adaptation and language ecology. Quantitative lexical distances as well as the accompanying data visualization, language maps, offer the opportunity to identify language varieties and show patterns of diffusion and divergence, predicated on local ecological conditions.

2.4 Language mapping and language identification

One of the most important tools in the 'dialectology laboratory' is language mapping. Language maps (also referred to as 'language atlases' when maps are more extensive) are the substance of language inventories which form the main source of data for studies in *diversity linguistics*. This cover term, provided by Drude (2018: 122), can be defined:

> [in] a broad sense, … [as] those branches of linguistics that show interest in the diversity of languages, their structure and relationship: descriptive linguistics (especially of previously understudied languages, often in a fieldwork setting), language typology, and comparative linguistics. Language documentation is included in or at least a close neighbor to this group. … In a narrower sense, ['diversity linguistics' also includes]… those studies that are interested in the diversity in itself, aiming in a first step at creating comprehensive catalogues of the world's linguistic diversity, where all languages are recorded with the information necessary to identify them…

This 'cartography of language diversification' for Kayanic languages encompasses within its goals the foci of diversity linguistics in both 'broad' and 'narrow' senses. Here, language and dialects are identified, both by name and by language boundaries, established based on calculated intelligibility. Their geographic distribution is then visualized graphically on a map. The patterns of this distribution will speak to the mechanisms (in this case, linguistic ecological niche adaptation mechanisms) that created the language diversity, observed in the region. This will, of course, serve to reify the integrity of and describe the evolution of the Kayanic language family. It will also demonstrate the potential role of synchronic dialectology in studies of language evolution while providing a specific case in point. The practical goals of this study include proposing adjustments to ISO 639-3 codes while providing initial-stage documentation (new lexical data) for some Kayanic varieties that have yet to or rarely appear in print.

Language identification, followed by language mapping and language inventorying is an essential way in which dialectological methods can be employed to answer important questions in the evolution of language diversity. Nowhere is there more needed than in Borneo. In giving a

brief annotation of Mosley and Asher's *Atlas of the World's Languages* (1994), Blust and Smith (2014: 35) call the work:

> [a]n ambitious attempt to map the world's languages. Like most works of global scope, detail is sacrificed in the name of comprehensiveness, and this is perhaps more true of Borneo than many other locations for which better descriptive data is available.

Borneo needs good language maps and starting with Kayanic might be a good place.

Language inventory has a long tradition in the history of language studies, despite its fluctuations in interests over time. Gesner's *Mithridates* (1555) from the Renaissance is an early work in language inventorying (Drude 2018: 123). During the Age of European colonialism, we see the *Catálogo de Lenguas de las naciones conocidas* by Hervás y Panduro (1800) and Friedrich Adelung and Johann Severin Vater's work also named *Mithridates* (1806). Wilhelm von Humboldt also investigated diversity linguistics (Drude 2018: 123). He believed that language diversity was the result of an active *Geisteskraft* 'mental power' exercised by various groups (1999 [1836]: xi). Although he viewed language as emanating primarily from the inner workings of the mind, he nevertheless reminded us to:

> take note of the quality of the transmitted material and the historical milieu in which a nation finds itself at the time of a significant reshaping of language, between a prehistory that works upon it, and the seeds of further development that lie within itself (1999 [1836]: 26).

The internal (or mental) and external ecologies of language interact, indeed, for "[s]peech demands elements adapted to the possibility of its limitless use" (Humboldt 1999 [1836]: 109). Vico was also an early theorist on language adaptation along these lines, suggesting that languages reflect the collective learned experiences of its speakers. In short, language reflects culture, and therefore can be probed to "seek out the ancient wisdom ... from the very wisdom of their words" (Vico 1988 [1711]: 40).

Throughout the late 19th and 20th centuries, dialectology and dialect mapping gained interests, visualizing and theorizing on the mechanisms of language change and therefore language diversity creation. Voegelin and Voegelin (1977) among few others brought language inventorying into the modern age (cf. Drudge 2018). Wurm and Hattori (1981) produced a stunning language inventory and atlas for the South Pacific region (including Borneo). Many of today's polygons that represent the distribution of Kayanic language varieties are based on Wurm and Hattori's work.

The Ethnologue (begun in the 1950's but would rise to ascendancy by the 1990's) became the standard reference for language inventory data and language maps (cf. Drude 2018). The Ethnologue, to date (August 2021), has information for 7,139 languages (Eberhard et al. 2021). The massive amount of data in the Ethnologue soon warranted a system for unambiguously identifying language varieties. An example of the problems encountered without such a system are the fact that 'Kayan' is the name of a language in Borneo, but also the name of a language in Myanmar and Papua New Guinea. 'Rejang' as in 'Rejang Kayan' is also a language in Sumatra. To deal with this issue, the International Standardization Organization (ISO) in conjunction with the Ethnologue established the three-letter ISO 639-3 code, for uniquely identifying languages (Drude 2018: 124). To reiterate, one of the goals of this study is to suggest new ISO codes for mutually unintelligible varieties and therefore update modern language inventories.

It should be noted that other language inventories exist as well for various purposes. Some are more generalized language inventories (Glottolog, Sorosoro, etc.) while others may focus on endangered languages (UNESCO, Endangered Languages) or language typology (WALS, etc.). Language inventories and their maps help capture "landscapes of multilingualism"

which helps us better understand the language ecology which is particularly important for endangered languages (Drude et al. 2018: 2).

What is important here is that language maps, as part of the dialectology toolkit, can be drawn upon to "illustrate some facet of information about language, or about communities of language speakers" (Gawne & Ring 2016: 190). For our purposes, an interpretive language map seems to be the most appropriate way to visualize and model the synchronic variation of language groups in order to understand the mechanisms that lead to that variation.

2.5 Identification and mapping of the Kayanic languages

Research on Kayanic languages is quite scant. Therefore, these varieties, from the standpoint of language mapping and language identification, are in need of serious work. In passing, brief notes on the sociolinguistic and ethnographic milieu appear here and there. A few wordlists were collected during the Dutch and British colonial periods (Barth 1910; Burns 1849; Crawfurd 1852; Douglas 1911). Dutch military doctor Anton Willem Nieuwenhuis also collected a wordlist (Sellato 2002: 31); however, his materials are unavailable.

In contemporary times, a few scholars have written brief works on Kayanic languages (Blust 1972, 1974b, 1998, 2000, 2010; Clayre & Cubit 1974; Guerreiro 1996, 2002; Rousseau 1974). Blust (1977) wrote a basic sketch grammar of the Kayanic language, Uma Juman (what we call here Rejang Kayan). Southwell (1990) wrote a dictionary for one Kayan variety. The Pusat Bahasa—the government ministry for the languages of Indonesia—has recently published some wordlists from this area, although matching the wordlists to specific ethnic groups is challenging (Aritonang et al. 2002; Kurniawati et al. 2002).

Smith (2015, 2017, 2019) and Soriente (2003) provide considerably more data for Kayanic languages. This study aims to add to the documentation available on Kayanic languages

while providing an in-depth analysis on the structure of intelligibility, speaking to the larger question of the evolution of language diversification.

It should be noted that various naming conventions complicate the language identification process and, therefore, also, that language mapping process. Sometimes this was due to scholarly inconsistency, but it is mostly due to the complex nature of 'ethnicity' and 'ethnic categorization' in Borneo (See Rousseau 1990).

Several maps exist for Kayanic languages (Wurm & Hattori 1983, Eberhard et al. 2021) including a chart outlining the villages by ethnic group (Okushima 2006). These maps require dialectometric analysis in the context of a language survey so that the language groups outlined are more than impressionistic.

2.6 Sociolinguistic survey and survey tools

The practical instruments of carrying out linguistic fieldwork for the purposes of language mapping and identification are subsumed within the realm of language survey research. Language surveys (also known as Sociolinguistic surveys) are typical of survey research in the social sciences in some ways, but are quite unique in others. While they rely on many of the same research instruments as other social science (questionnaires, interviews, tests, rating or attitude scales, content analyses, systematic literature reviews, and [participant-]observation) (Blair 1990, Grimes 1995, Decker & Grummitt 2017, Cooper 1980), language surveys differ from other sciences in that the product is considered a substantial work in itself. While, in other sciences, surveys are generally "characterized as exploratory in nature, as being especially useful at the preliminary stages of investigation, when the investigator is often ignorant of the relevant variables[,] more often [sociolinguistic surveys] are not" (Cooper 1980: 113). Extensive fieldwork and correlational designs often characterize survey research in linguistics; however,

experimental designs are also used (Cooper 1980: 114). From this perspective, language surveys can be generally informative or reportative in nature, but they can be hypothesis-driven as well (Cooper 1980: 120).

The language survey is a collection of research techniques that have literally been field tested all over the world in both pre- and post-industrial societies (Labov 1966 [North America], Anonby & Holbrook 2010 [South America], Bender 1971 [Africa], Ladefoged et al. 1971 [Africa], Karan 1996, 2001 [Africa], Grierson 2005 [1928] [South Asia], Clifton 2002, 2005; Malyshev et al. 2019 [Central Asia], MacNamara 1966 [Europe], Badia i Margarit 1969 [Europe], Pelkey 2011 [East Asia], Anderbeck 2008, 2018 [Southeast Asia], Anderbeck & Aprilani 2013 [Southeast Asia], Meyers et al. 2003 [Southeast Asia], Z'graggen 1975, 1980 [Oceania], Charpentier & François 2015 [Oceania], Bickford 1991, 2005 [Sign Languages], also see SIL International's *Journal of Language Survey Reports*, formerly known as *Electronic Survey Reports*, containing over 500 survey articles and monographs, representing 98 countries and territories). Language surveys capture new data, not only from the language itself, but also, about *language behavior* and *behavior toward language* (Agheyisi & Fishman 1970; Cooper 1980). Language surveys tell the state of the language interaction in a given linguistic community.

The typical language survey (again see SIL International's *Journal of Language Survey Reports*), provides an overview of the places and people groups involved in the speech community under question. They usually rely on a battery of research instruments (such as wordlist elicitation, morphosyntactic elicitation, text collection, bilingual/multilingual proficiency tests, non-reactive procedures, participatory folk dialectology focus groups, etc). The relationship between language surveys and ethnographic methods from anthropology is also

quite tight, employing a number of ethnographic tools (Bernard 2000; Bernard 2011). Language surveys, particularly, in their 'people profiles' may engage in *ethnology*—analytically comparing various people groups, revealing the distinctions and relationships among them in language practice. Language contact and attitudes gains a fruitful source of data in ethnological analyses. Methods are usually detailed with results presented in a manner consistent with any social science study. Then all of this information is brought together and interpreted to provide a picture of the language ecology of a given speech community.

Because language surveys usually collect language data to support its arguments, in addition to data *about* the language, and, because an plethora of new data is usually provided in the appendices (wordlists/lexicons, grammatical data, and annotated texts), I consider language survey research a form of language documentation. After all, language surveys may sometimes provide the only language data and grammatical description we have for certain language varieties (Mahakam Kayan and Wehea are examples of this in this language survey).

In addition to the typical Boasian 'trilogy' that forms the gold standard of language documentation (dictionary/lexicon, grammar [sketch], and a collection of texts), 'sociolinguistic language documentation' (Childs et al. 2014) and 'ethnographically informed language documentation' (Harrison 2005) is essential to capture the uniqueness of linguistic practices and culturally dependent grammatical features that would not otherwise be available to the researcher. As Childs et al. (2014: 169) write "in highly multilingual and fluid linguistic contexts where language use is organized around multilingual repertoires rather than 'native' languages," documentation needs to be about more than language structure.

If language documentation is to accomplish its superlative goal of recording and preserving (or even revitalizing) small, indigenous, and minority languages (SIM)[11] (particularly endangered ones), it must focus on preserving not just single languages, but entire language ecologies (Mühlhäusler 1992). Documenting and preserving language ecologies must be at the heart of language documentation and maintenance work, as Mühlhäusler (1992: 178) writes:

> The aim of language maintenance must be to preserve active use, discourse, functioning patterns of transmission, and other supporting ecological factors. Of particular importance is the maintenance of dialectal and linguistic variety, for such diversity is the basis of language regeneration. … It is very much an ecology that can guarantee to sustain discourse among the speakers of a language and speakers of languages.

From this perspective, language survey research is indispensable to the language documentation and conservation enterprise.

The information gathered in language survey research also helps contribute to language policy and language program planning (MTB-MLE programs to sustain SIM languages or language revitalization programs to revive endangered or moribund languages). Surveys are also often used to evaluate language policy measures and language programs. And, of course, surveys can contribute to language promotion, creating positive attitudes towards one's language and lending a kind of "symbolic value" since the language is worthy of having scientific study, conducted on it or even in it (Cooper 1980). See *La Llengua dels Barcelonins: Resultats d'una enquesta sociològico-lingüística* (Badia i Margarit 1969) for an example of this kind of language promotion in Catalan (Cooper 1980: 121). Note that this language promotion can be viewed as part of the symbolic niche that shapes language structure since language shift from negative linguistic attitudes can have an effect on the evolution of all levels of linguistic analysis

[11] I first introduced the term, 'Small, indigenous and minority' languages or 'SIM languages' for short, in an academic setting at the Canada Institute of Linguistics (CanIL) Colloquium, during my talk, *Do Languages Adapt to their Environments? The Role of Language Documentation and Linguistic Fieldwork in Evolutionary Linguistics (Notes from Indonesian Borneo)* in April of 2021. I also used this term in my talk *Field-based evolutionary linguistics* at Cog Fest (The NY-St. Petersburg Institute of Linguistics, Cognition and Culture) in July of 2021.

(phonology, morphology, syntax, information structure, etc.). This, of course, also supports the language documentation and conservation enterprise.

Language surveys such as this hypothesis-driven 'cartography of diversification' aim to capture the dynamics of language diversification, through wordlist collection and lexicostatistical/phonostatistical analysis of intelligibility, interpreting those dynamics through linguistic ecological niche factors identified in Kayanic ethnology.

2.7 Conclusions and summary: Defining terminology

To remind us of our ultimate guiding inquiry, this study is concerned with whether the lexical composition of these Kayanic language varieties, as dictated by the phonological structure of their lexemes, show adaptation to the environment (*Linguistic Niche Hypothesis*) in which speakers live and interact. It is proposed that language adaptation in this way, is a driving force in the creation of *linguistic diversity*. If so, this would be demonstrable by showing that *ecological niche factors*—specific elements of the environment that shape an outcome—bias patterns of intelligibility to diverge (or converge) with other language varieties. Remember that *diffusion* and *divergence* form the main substance of *language diversification*.

While some variation is *neutral* (i.e *neutral mutations*), occurring under the conditions of *random* or *stochastic drift* (e.g. regular sound change), which itself can be a complex *concerted evolutionary process* (e.g. chain shifts, like the Great Vowel Shift in English), some variation is due to the aforementioned adaptive biases, emerging from being shaped by the dynamic web of interactions among various ecological niches—specific 'compartments' of the environment that are open to influencing each other as well as language. The fact that this linguistic behavior, in this case, mutual intelligibility, emerges from, not just being shaping by the 'compartments' themselves, but also the interaction of the 'compartments,' shows that language (and the inherent

variation and change that constitute it) is a *complex adaptive system*. This dynamic web connects at least ten ecological niches: (1) physical niche, (2) social niche, (3) cultural niche, (4) symbolic niche, (5) cognitive niche, (6) linguistic niche, (7) technological niche, (8) developmental niche, (9) bio-corporeal niche, and (10) genetic niche. These linguistic niches mold language at multiple levels of linguistic analysis (although our focus here is on the lexicon and phonology). Again, while described discretely for clarity purposes, in reality, they are not, forming that dynamic web of interactions that constitute the complex adaptive system. Simply put, 'the whole is not the sum of its parts.'

Of course, this study will focus on a subset of the ecological niches at play—physical, social, cultural, symbolic, and linguistic niches—within the overall *language ecology* of Kayanic speakers.

In order to observe and understand how linguistic niches create language diversity, we must borrow methods from a field that observes language divergence 'in vivo'—*dialectology* or *dialect geography* (and more specifically still, its quantitative subsection, *dialectometry*). Dialectology has always been concerned with the adaptive interplay between language variation and the social and environmental conditions of speech communities. Dialectology from a micro evolutionary perspective will show how ecological niches lead to *allopatric speciation* by creating either *segregative* or *integrative* conditions for speakers, limiting interactions and, thus, *intelligibility flow* (analogized with 'gene flow' in biology)—the flow of phonological structure and lexical composition features through language contact which would maintain intelligibility.

This, naturally, leads us to ask, what kinds of dialectology tools can do this? The answer? Language mapping and identification tools as part of language survey fieldwork techniques. *Calculated intelligibility* data will be collected through language survey wordlist elicitation (and

archival wordlists) which will then be analyzed using *lexicostatistics/phonostatistics*. This will provide *language distances* which are then clustered and plotted cartographically—forming an *interpretive map*. This in itself will be a great contribution to the field since very little work has been done on Kayanic. The results of the map will be interpreted further through the ecological niches, found in the ethnographic recording and presented in the *ethnology* (the comparative study of Kayanic cultures and their ecological conditions) in the next chapter. This will demonstrate the linguistic niche adaptive mechanisms of language diversification from the case study of Kayanic languages in Central Borneo—a 'cartography of diversification.' In addition to this, practical goals of refining *language inventory* data (adjusting *ISO codes*) and providing new *language documentation* for previous undescribed or underdescribed SIM languages allow this study to contribute to *language conservation* and potentially *mother tongue-based multilingual education (MTB-MLE)* or *language revitalization* in the region. This is an existential concern of the entire field of linguistics and hypothesis-driven, sociolinguistic-sensitive language survey research such as this is at the heart of the enterprise.

Let us not turn to identifying the ecological niches at play in Kayanic language communities.

CHAPTER 3

Methodology & Fieldwork:

Research on the Language Ecology of Borneo

> [T]echnique exists for the service of ideas. Without content, technical virtuosity is barren. Sociolinguistic surveys can be useful not only to the extent that their techniques are appropriate and sound but, more importantly, to the extent that the information they are designed to collect is worth gathering…
>
> —Robert L. Cooper[12]

3.1 Introduction

In this section, I describe the methodology and its constitute tools, used to address whether the phonological structure and lexical composition of the Kayanic languages as reflected in their patterns of calculated intelligibility show signs of adaptation to their language ecology—the claim purported by the Linguistic Niche Hypothesis. I began by describing my fieldwork—indeed the methodological center of all language survey research. In addition to this, I take a moment of reflexivity on the ethical and experiential aspects of fieldwork among the Kayanic peoples of Central Borneo. From there, I describe the quantitative orientation of this study, followed by a detailed outline of specific survey tools employed. Finally, I lay out the method and procedures for doing lexicostatistics/phonostatistics. I conclude with a section on data processing, analysis, and management.

3.2 Finding the field

The first step in any fieldwork research is to tightly circumscribe one's area of inquiry. Regionally speaking, this study focuses on the Kayanic language varieties of Central Borneo.

[12] Quoted from Cooper (1980).

Kayanic languages share this area with many different languages across many different language subgroups within Austronesian. This has naturally led to many scenarios of language and cultural contact (which will be important to our discussion of language adaptation) which, of course, has created many continuities among the people groups there. Despite these continuities, a strict focus will be maintained on language varieties, established as Kayanic, as was previously attested in the historical-comparative reconstruction literature (Smith 2017a; Soriente 2003).

3.3 Life in the field

The vicissitudes of life in the field are well known. Going to new places and forming new relationships is always rewarding, but fieldwork is, of course, not without its challenges.

My fieldwork was conducted in Indonesia from August 2019-March 2020. One of the biggest adjustments required a great shift in plans due to the infamous Covid-19 pandemic—a point to which I will return presently. When I arrived in Indonesia, language training was the first matter of order. Having not spoken Indonesian in years, I began classes in Yogyakarta, refreshing my Indonesian language skills to be used in the field. After a little over a month of intensive language training, I was able to contact a member of the Wehea community, to visit East Kalimantan, to get the lay of the land. I took the opportunity to explore the cultural landscape of the region. I visited East Kalimantan two times in 2019, traveling from village to village over land and upriver by boat, getting necessary permissions and forging connections for later research.

Due to needing a visa status change, survey research did not take place until March 2020. Unfortunately data collection was cut short due to a required evacuation of all foreign researchers, therefore I had to leave with the data I had already collected. In a frenzy, I managed to collect four wordlists (Wehea, Kelinjau Modang, Mahakam Kayan, and Bahau) within the city

of Samarinda. From there, I sought out wordlists for the other language varieties that were published in the linguistic literature. Note that my Wehea and Mahakam Kayan lists were completely new additions to the literature at that point. My Bahau list was incomplete and therefore not incorporated into this study.

3.4 Ethical concerns: On reflexivity and practice

Ethics in human subjects research is always a question of responsibilities and actions. What are my responsibilities to the communities and the individuals that comprise them? In what ways do I ensure their physical, psychological, socio-economic, or legal protection?

As field linguists, we work with people. This is the most important of all considerations to remember when doing any linguistic work in the field.

While conducting a sociolinguistic survey for the Kayanic languages, I made sure to conduct myself in a friendly and professional manner, always seeking the proper permission and consent to ensure my consultants protection and well-being.

Another ethical concern has to do with the racism-soiled history of studies in the evolution of language and of human cognitive capacities in general. While linguists (as cognitive scientists) mostly abandoned the topic of language evolution in the 19th century (*Société de Linguistique de Paris* in 1866) until quite recently (Pinker & Bloom 1990), Anthropology and other social sciences continued work on the subject. Alexander Francis Chamberlain in his *The Child: A Study in the Evolution of Man* (1900) describes an 'baby-cry origin of language' theory wherein he draws corollaries between enfant linguistic features, nature-imitating onomatopoeia and reduplicative syntax in the, so called, "languages of savages and barbarians" and the evolution of language. Speaking of native Australian, American, and African languages (so called "primitive peoples" or "primitive races"), he writes (1900: 115-116):

> The first speech of mankind consisted of natural cries--shouts, grunts, and hisses...That there is a great variety in the onomatopoeic or imitative words of the lowest races of man--a much greater variety than can be said to exist in the early speech of the human child--is evident from a careful study of their language. In the different Australian dialects, e.g., we find the following words (among others) for 'laugh': waler, krambalwert, kangalla, gooryman, kinka, tirrikeblin, munka, yie, munjur, kindi pillia, karibok, ginthinthintha, wathiman, yathin, etc.; and among American Indian tribes the following names for the 'butterfly': tlētlū, lōlēnū, kōlílū, walwilékash, képkap, wékwak, etc.; and in Australia: billybyleukka, coolumbria, booroo booroo, balumbir, etc. So, also, there is immense variety in the words for 'yes' and 'no' among the Australian and other primative languages, complexity being often found where least expected, and simplicity where it might not be looked for.
>
> It is fair, however, to say that, with respect to human noises and movements especially, the Australians (and some other primitive races as well, like the Fanti) evidence great skill in onomatopoeic imitation. The Dieyerie language of South Australia, e.g., has many very expressive words of this sort.

One must proceed with caution not to imply that what is special about these languages is some snapshot of a primitive or exoticized state. That the principles of language evolution are universal and that this is just one case study that is *inclusive* of the languages of Borneo. Any claims like those made by Chamberlain (1900) are clearly and blatantly false; yet, we must aim to enlighten our understanding of how the ‘external’ environment in which people are situated interacts on the structuration of ‘e-language’ which is turn, ‘seeps into’ the structure and restructures the ‘i-language’ or linguistic competence, leading to the birth of linguistic diversity.

3.5 Quantitative approaches to language diversification and evolution

This thesis relies primarily on quantitative methods in calculations of intelligibility. Lexicostatistics and phonostatistics are the methodologies by which I will analyze the lexical data for Kayanic languages. Lexicostatistics measures the word-level similarity between a given set of lexical items. Phonostatistics examines the specific phones within those lexemes, measuring language distance as a function of differences in phonetic features. These methods provide percentages that represent the likelihood that there is no structural barrier to communication between given sets of language variants.

3.6 Connecting methods with tools

Quantitative methods call for the use of specific tools in the data collection process. First, I will describe how data were collected (through original fieldwork) and amassed (through the linguistic literature). Then, I will describe, in the next section, the procedures involved in data management, preparation, and analysis, yielding the results outlined in the next chapter.

3.6.1 Selecting sample populations

As is typical of language mapping studies (cf. Anderbeck 2008), this study relies on 'judgement' sampling or what is also known as 'purposive' sampling (Bernard 2011). Both the sampling of villages and the consultants within them rely on this sampling method. Bernard (2011) explains that with judgement sampling "you decide the purpose you want informants (or communities) to serve, and you go out to find some." This sampling method was selected, because it meets two main criteria for which non-probabilistic sampling methods such as this one are particularly apt: (1) large survey studies for hard-to-reach populations and (2) studies that focus on socially shared or cultural knowledge (Bernard 2011). This study meets both of those criteria as it is a very largely encompassing survey, based on garnering socially shared knowledge of lexical items in a given language community.

While the original plan was to visit and collect new data for the villages representing each language group, due to Covid-19, consultants, representing several language groups (Wehea Modang, Kelinjau Modang, Bahau, and Mahakam Kayan) were purposefully sought. All other language varieties were obtained through the linguistic literature.

As most of the information about language communities and their exact locations are hard to come by, villages were chosen by scouring the available linguistic, anthropological, and conservation biology literature. Of the 20 Kayanic language variants attested, three—namely

Long Belah (Guerreiro 1993: 185), Mendalam Kayan (Eberhard et al. 2021), and Punan Mahkam (Guerreiro 1993: 185)) were completely unidentifiable using the aforementioned sources.

3.6.2 Sociolinguistic survey instruments

Below are the survey instruments used to collect the data, required for accessing the lexical composition and phonological structure of the Kayanic language varieties under study.

3.6.2.1 Informal interviews (with 'knowledgeable insiders')

Interviews with knowledgeable insiders was extreme helpful in order to seek the proper permissions to conduct fieldwork, make the connections and form relationships with those who would be consultants, lend credibility, and get the 'lay of the land' in terms of the locations and distributions of languages and their locations. Most were truly informal, but one in particular took the for of a semi-structured interview, typically of ethnographic/sociolinguistic interviewing (Emerson et al. 2011; Blair 1990). One really useful interview in particular yielded an extensive list of languages and the villages where they were spoken in addition to a basic 'insider-view' mapping of the Mahakam river.

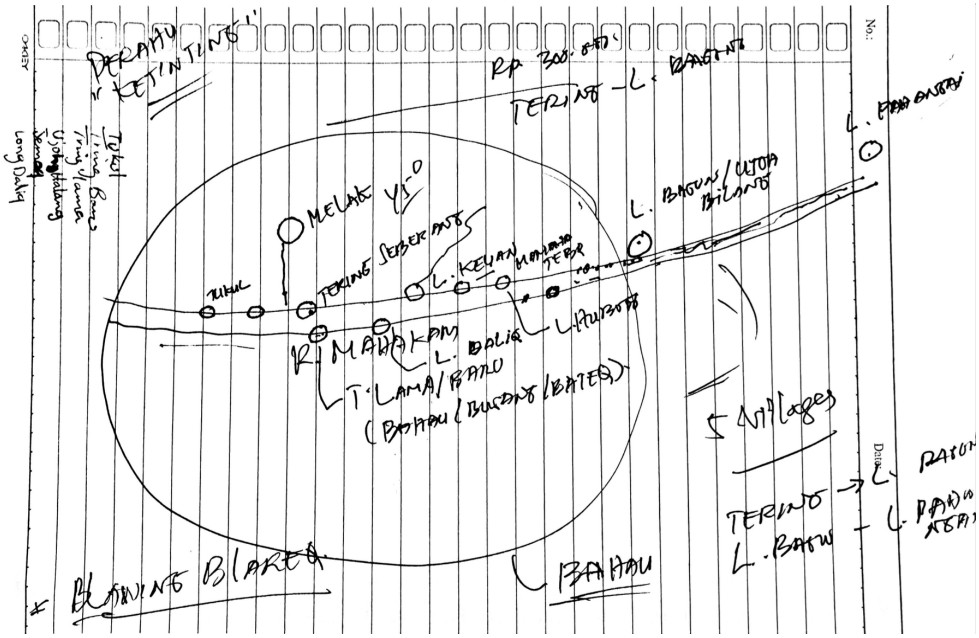


Figure 12. Ethnolinguistic map produced in one of the interviews

3.6.2.2 Wordlists: Collection and comparison

The data corpus for this study includes wordlists from 17 Kayanic language variants.The wordlist used for this study includes the 210-word *Basic Austronesian Wordlist* (Greenhill et al 2008). For the lists, I collected, the Basic Austronesian Wordlist was augmented by the rest of the Swadesh wordlist upon which it was based (Swadesh 1955). A few items, specific to the Kayanic socio-cultural and natural environment were added to my lists as well. The wordlists in this study were collected both from fieldwork by me and through the linguistic literature. Only the words that figured in the Basic Austronesian Wordlist were analyzed using lexicostatistics/ phonostatistics. Due to Covid-19, one speaker followed up with additional data for a wordlist with recordings over Whatsapp. The additional lists were found throughout the linguistic literature and incorporated into the analysis. Here is the table of wordlist sources:

Table 3. Wordlists used in the lexicostastical/phonostatistical analysis of this study

Wordlists	
Espree-Conaway (2020)	Kelinjau Modang [Long Tesak], Wehea, Bahau, Mahakam Kayan
Smith (2017a)	Ngorek, Merap [Mpraa], Baram Kayan [Long Naah], Data Dian Kayan, Rejang Kayan [Balui Liko], Busang, Bahau, Long Gelat, Modang, Gaai, Punan Kelai.
Soriente (2003)	Pua', Hueng Bau, Nyibun, Kayan River Kayan [Uma Laran Kayan].
Blust (1974a)	Ngorek (Murik)

Data from Pusat Bahasa and other archival sources were consulted but not incorporated into the lexicostatistical/phonostastical analysis of this study (Aritonang et al. 2002; Astar et al. 2002; Kurniawati et al. 2002; Blust 1974a; Rousseau 1974a; Smith 2019; Southwell 1990; Douglas 1911, 1912; Burns 1849; Brooke Low 1896; Barth 1910).

3.6.2.3 Participant observation and fieldnotes

While in the field, I kept a field notebook where I collected information and reflected upon that information and my role in its creation (the kind of reflectivity, typical of fieldwork-based sciences). I also actively participated in social and cultural activities that further deepen my knowledge of the contextual environment in which these languages are used. This aided, not only in building the kind of trust, necessary to collect good linguistic data (in this case, primarily lexical data), but also, to have the experience to better interpret the ethnographic data, surveyed in this study.

3.7 Data management, preparation, and analysis

The analysis process began with collaborating the Ethnologue Project Indonesia (EPI) survey team and the databases they had amassed based on Pusat Bahasa data (Aritonang et al. 2002; Astar et al. 2002; Kurniawati et al. 2002) and other data sources in the literature (Smith 2017; Soriente 2003). In a desire to get a 'first pass' at the language varieties their were likely to cluster and to observe the most likely clustering patterns, Gabmap (Leinonen et al. 2016)—a computational dialectology tool was used. This tool automates the integration of precise geographic data, Levenshtein distance statistical analysis[13], and cartographic visualization. Karl Anderbeck (one of the committee members on this thesis) and I ran an initial pass on this data, arriving at the figure below.

[13] Levenshtein distances are considered phonostatistics in more contemporary dialectological studies (Anderbeck 2018); however, this study takes a more traditional approach to defining phonostatistics as outlined in Grimes & Agard (1959) and Mckaughan (1964). More on this below.

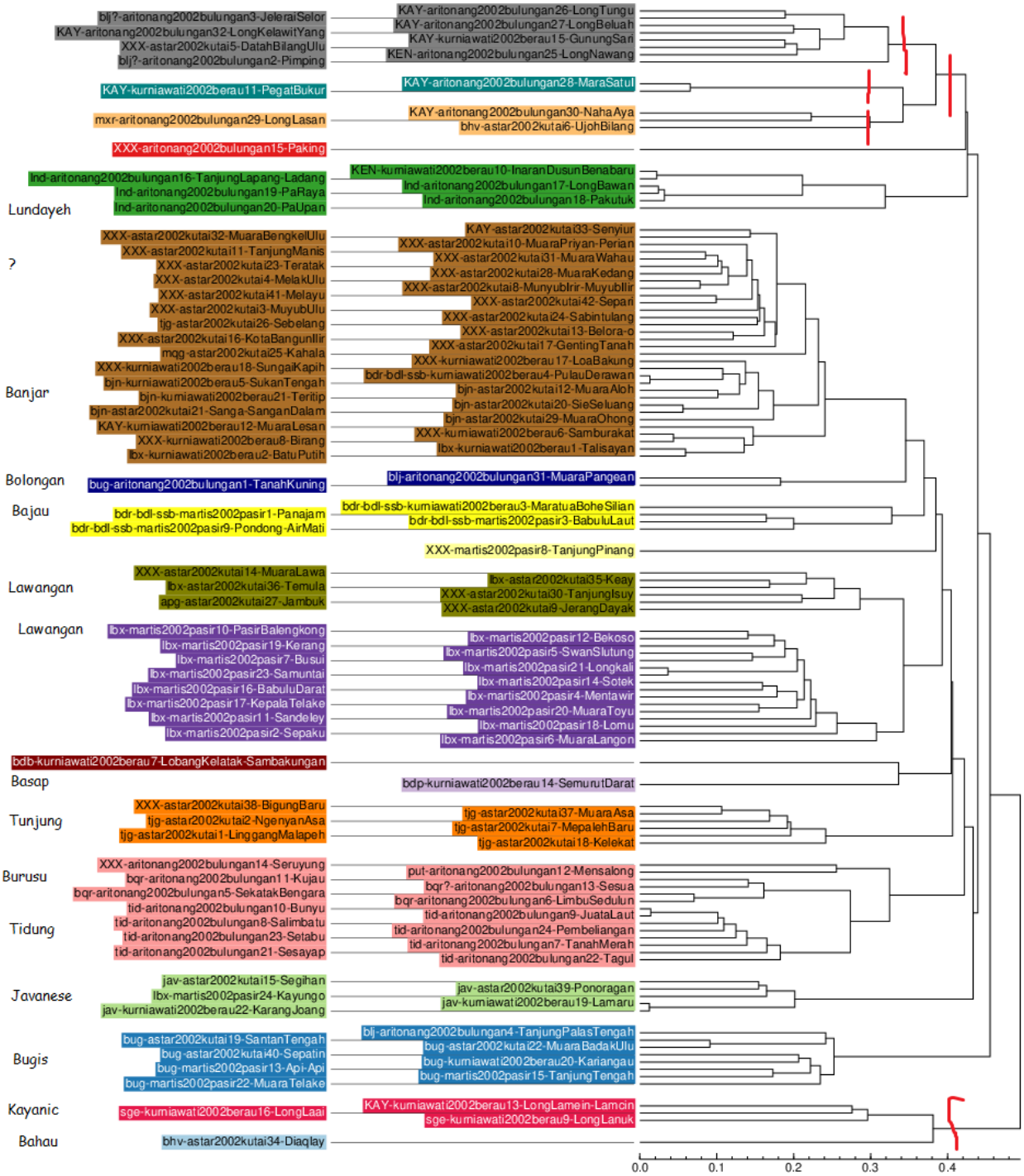
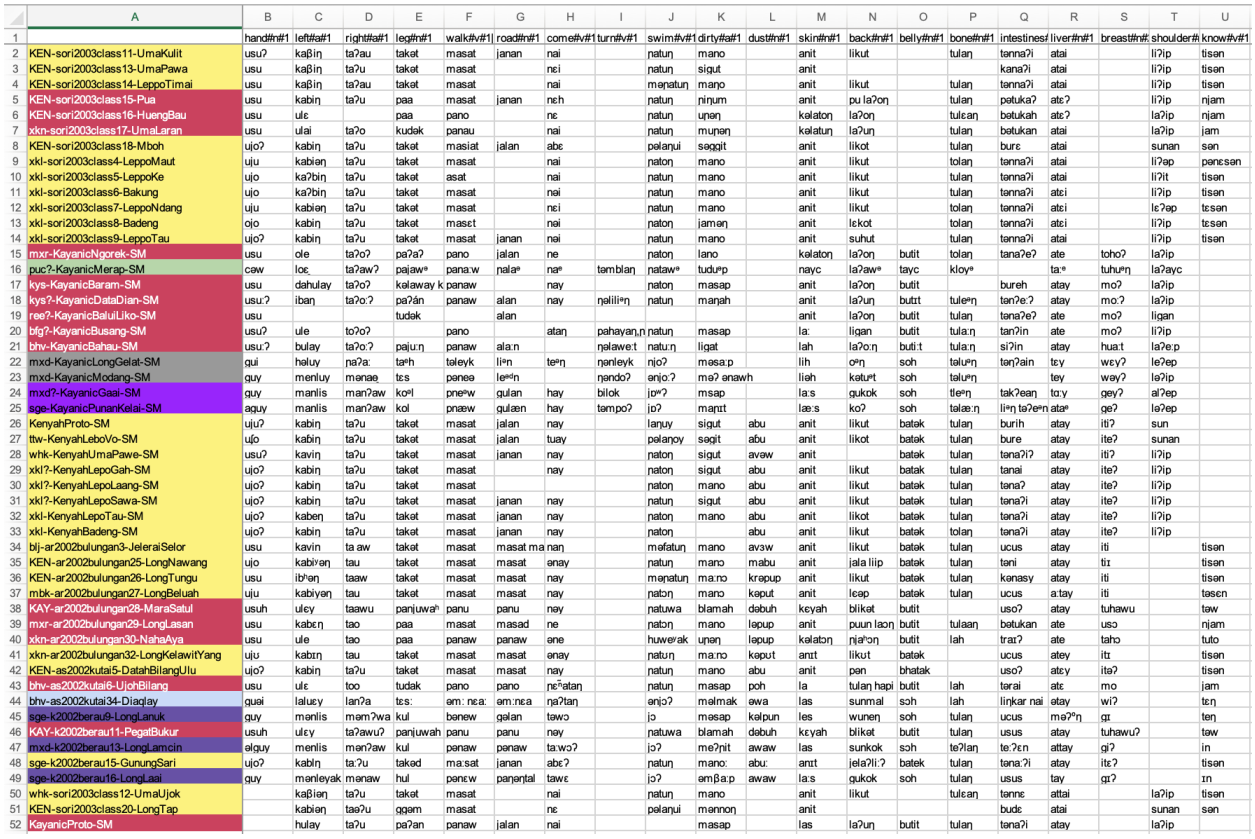


Figure 13. Gabmap Computational Classification of Language Data from Linguistic Scholarship

Below is a snapshot of the coded EPI database used in the Gabmap program.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		hand#n#1	left#a#1	right#a#1	leg#n#1	walk#v#1	road#n#1	come#v#1	turn#v#1	swim#v#1	dirty#a#1	dust#n#1	skin#n#1	back#n#1	belly#n#1	bone#n#1	intestines#	liver#n#1	breast#n#	shoulder#	know#v#1
2	KEN-sori2003class11-UmaKulit	usu?	kaβiŋ	ta?au	takət	masat	janan	nai		natuŋ	mano		anit	likut		tulaŋ	tənna?i	atai		li?ip	tisən
3	KEN-sori2003class13-UmaPawa	usu	kaβiŋ	ta?u	takət	masat		nɛi		natuŋ	siqut		anit				kana?i	atai		li?ip	tisən
4	KEN-sori2003class14-LeppoTimai	usu	kaβiŋ	ta?au	takət	masat		nai		mənatuŋ	manɔ		anit	likut		tulaŋ	tənna?i	atai		li?ip	tisən
5	KEN-sori2003class15-Pua	usu	kabiŋ	ta?u	paa	masat	janan	nɔ:h		natuŋ	ŋiŋum		anit	pu la?oŋ		tulaŋ	pətuka?	atɛ?		li?ip	njam
6	KEN-sori2003class16-HuengBau	usu	ulɛ		paa	pano		nɛ		natuŋ	uŋəŋ		kəlatoŋ	la?oŋ		tulɛaŋ	bətukah	atɛ?		la?ip	njam
7	xkn-sori2003class17-UmaLaran	usu	ulai	ta?o	kudək	panau		nai		natuŋ	muŋəŋ		kəlatuŋ	la?uŋ		tulaŋ	bətukan	atai		la?ip	jam
8	KEN-sori2003class18-Mboh	ujo?	kabiŋ	ta?u	takət	masiat	jalan	abɛ		pəlaŋui	səgqit		anit	likot		tulaŋ	burɛ	atai		sunan	sən
9	xkl-sori2003class4-LeppoMaut	uju	kabiəŋ	ta?u	takət	masat		nai		natoŋ	mano		anit	likut		tolaŋ	tənna?i	atai		li?əp	pənɛsən
10	xkl-sori2003class5-LeppoKe	ujo	ka?biŋ	ta?u	takət	asat		nai		natuŋ	mano		anit	likut		tulaŋ	tənna?i	atai		li?it	tisən
11	xkl-sori2003class6-Bakung	ujo	ka?biŋ	ta?u	takət	masat		nəi		natuŋ	mano		anit	likut		tulaŋ	tənna?i	atɛi		li?ip	tisən
12	xkl-sori2003class7-LeppoNdang	uju	kabiəŋ	ta?u	takət	masat		nɛi		natuŋ	mano		anit	likut		tulaŋ	tənna?i	atɛi		lɛ?əp	tɛsən
13	xkl-sori2003class8-Badeng	ojo	kabiŋ	ta?u	takət	masɛt		nai		natoŋ	jamaŋ		anit	lɛkot		tolaŋ	tənna?i	atɛi		li?ip	tɛsən
14	xkl-sori2003class9-LeppoTau	ujo?	kabiŋ	ta?u	takət	masat	janan	nai		natuŋ	mano		anit	suhut		tulaŋ	tənna?i	atai		li?ip	tisən
15	mxr-KayanicNgorek-SM	usu	ole	ta?o?	pa?a?	pano	jalan	ne		natoŋ	lano		kəlatoŋ	la?oŋ	butit	tolaŋ	tana?e?	ate	toho?	la?ip	
16	puc?-KayanicMerap-SM	cəw	loɛ	ta?aw?	pajawᵉ	pana:w	nəlaᵉ	naᵉ	təmblaŋ	natawᵉ	tuduᵉp		nayc	la?awᵉ	tayc	kloyᵉ		taːᵉ	tuhuᵉŋ	la?ayc	
17	kys-KayanicBaram-SM	usu	dahulay	ta?o?	kəlaway k	panaw		nay		natoŋ	masap		anit	la?oŋ	butit		bureh	atay	mo?	la?ip	
18	kys?-KayanicDataDian-SM	usu:?	iban	ta?o:?	pa?án	panaw	alan	nay	nəliliᵉŋ	natuŋ	manah		anit	la?uŋ	butɪt	tuleᵉŋ	tən?e:?	atay	mo:?	la?ip	
19	ree?-KayanicBaluiLiko-SM	usu			tudək		alan						anit	la?oŋ	butit	tulaŋ	tana?e?	ate	mo?	liqan	
20	bfg?-KayanicBusang-SM	usu?	ule	to?o?		pano		ataŋ	pahayaŋ,ŋ	natuŋ	masap		la:	ligan	butit	tula:ŋ	tan?in	ate	mo?	li?ip	
21	bhv-KayanicBahau-SM	usu:?	bulay	ta?o:?	paju:ŋ	panaw	ala:n		ŋəlawe:t	natu:ŋ	ligat		lah	la?o:ŋ	buti:t	tula:ŋ	si?in	atay	hua:t	la?e:p	
22	mxd-KayanicLongGelat-SM	qui	həluy	ŋa?a:	taʰh	təleyk	liᵉn	teᵉŋ	ŋənleyk	ŋio?	məsa:p		lih	oᵉŋ	soh	təluᵉŋ	təŋ?ain	tɛy	wɛy?	le?ep	
23	mxd-KayanicModang-SM	quy	mənluy	mənae	tɛs	panea	leᵉᵈn		ŋəndo?	anjo:?	mə? ənawh		liah	kətuʳt	soh	təluᵉŋ		tey	way?	la?ip	
24	mxd?-KayanicGaai-SM	quy	mənlis	man?aw	koᵉl	pneᵉw	gulan	hay	bilok	jᵊw?	msap		la:s	qukok	soh	tleᵉŋ	tak?eaŋ	tɑ:y	gey?	al?ep	
25	sge-KayanicPunanKelai-SM	aquy	manlis	man?aw	kol	pnæw	gulæn	hay	təmpo?	jɒ?	manɪt		læ:s	ko?	soh	talæ:ŋ	liᵉᵉŋ tə?eᵉᵉn	ataᵉ	ge?	la?ep	
26	KenyahProto-SM	uju?	kabiŋ	ta?u	takət	masat	jalan	nay		laŋuy	siqut	abu	anit	likut	batək	tulaŋ	burih	atay	iti?	sun	
27	ttw-KenyahLeboVo-SM	uʃo	kabiŋ	ta?u	takət	masat	jalan	tuay		pəlaŋoy	səgit	abu	anit	likot	batək	tulaŋ	bure	atay	iteʔ	sunan	
28	whk-KenyahUmaPawe-SM	usu?	kaviŋ	ta?u	takət	masat	janan	nay		natoŋ	siqut	avəw	anit		batək	tulaŋ	təna?i?	atay	iti?	li?ip	
29	xkl?-KenyahLepoGah-SM	ujo?	kabiŋ	ta?u	takət	masat		nay		natoŋ	siqut	abu	anit	likut	batak	tulaŋ	tanai	atay	iteʔ	li?ip	
30	xkl?-KenyahLepoLaang-SM	ujo?	kabiŋ	ta?u	takət	masat				natoŋ	mano	abu	anit	likut	batək	tulaŋ	təna?	atay	iteʔ	li?ip	
31	xkl?-KenyahLepoSawa-SM	ujo?	kabiŋ	ta?u	takət	masat	janan	nay		natuŋ	siqut	abu	anit	likut	batək	tulaŋ	təna?i	atay	iteʔ	li?ip	
32	xkl-KenyahLepoTau-SM	ujo?	kabeŋ	ta?u	takət	masat	janan	nay		natoŋ	mano	abu	anit	likot	batək	tulaŋ	təna?i	atay	iteʔ	li?ip	
33	xkl-KenyahBadeng-SM	ujo?	kabiŋ	ta?u	takət	masat	janan	nay		natoŋ		abu	anit	likut	batək	tolaŋ	təna?i	atay	iteʔ	li?ip	
34	blj-ar2002bulungan3-JeleraiSelor	usu	kavin	ta aw	takət	masat	masat ma	naŋ		məfatuŋ	mano	avəw	anit	likut	batək	tulaŋ	ucus	atay	iti		tisən
35	KEN-ar2002bulungan25-LongNawang	ujo	kabiᵛəŋ	tau	takət	masat	masat	ənay		natuŋ	manɔ	mabu	anit	jala liip	batək	tulaŋ	təni	atay	tiɪ		tisən
36	KEN-ar2002bulungan26-LongTungu	usu	ibʰəŋ	taəw	takət	masat	masat	nay		mənatuŋ	ma:nɔ	krəpup	anit	likut	batək	tulaŋ	kənasy	atay	iti		tisən
37	mbk-ar2002bulungan27-LongBeluah	uju	kabiyəŋ	tau	takət	masat	masat	nay		natoŋ	manɔ	kəput	anit	lɛəp	batək	tulaŋ	ucus	a:tay	iti		təsɛn
38	KAY-ar2002bulungan28-MaraSatul	usuh	uɛy	taawu	panjuwaʰ	panu	panu	nəy		natuwa	blamah	dəbuh	kɛyah	blikət	butit		uso?	atay	tuhawu		taw
39	mxr-ar2002bulungan29-LongLasan	usu	kabeŋ	tao	paa	masat	masad	ne		natoŋ	mano	ləpup	anit	puun laoŋ	butit	tulaaŋ	bətukan	ate	usɔ		njam
40	xkn-ar2002bulungan30-NahaAya	usu	ule	tao	paa	panaw	panaw	ene		huwɐak	uŋəŋ	ləpup	kəlatoŋ	ŋjaʰoŋ	butit	lah	tratʔ	ate	taho		tuto
41	xkn-ar2002bulungan32-LongKelawitYang	uju	kabɪŋ	tau	takət	masat	masat	ənay		natuŋ	ma:nɔ	kəput	anɪt	likut	batək		ucus	atey	itɪ		tisən
42	KEN-as2002kutai5-DatahBilangUlu	ujo?	kabiŋ	ta?u	takət	masat	masat	nay		natuŋ	mano	abu	anit	pən	bhatak		uso?	atɛy	itəʔ		tisən
43	bhv-as2002kutai6-UjohBilang	usu	ulɛ	too	tudək	pano	pano	nɛᵉatan		natuŋ	masap	poh	la	tulaŋ hapi	butit	lah	tərai	atɛ	mo		jam
44	bhv-as2002kutai34-Diaqlay	quai	laluɛy	lan?a	tɛs:	əm: nɛa:	əm:nɛa	ŋa?tan		anjo?	məlmak	awa	las	sunmal	soh	lah	liŋkar nai	atay	wi?		tɛŋ
45	sge-k2002berau9-LongLanuk	quy	mənlis	məm?wa	kul	bənew	qəlan	təwɔ		jɔ	məsap	kəlpun	les	wuneŋ	soh	tulaŋ	ucus	məʔᵒŋ	qɪ		teŋ
46	KAY-k2002berau11-PegatBukur	usuh	uɛy	ta?awu?	panjuwah	panu	panu	nəy		natuwa	blamah	dəbuh	kɛyah	blikət	butit	tulaŋ	usus	atay	tuhawu?		taw
47	mxd-k2002berau13-LongLamcin	əlquy	mənlis	mən?aw	kul	pənaw	panaw	ta:wɔ?		jɔ?	meʔnit	awaw	las	sunkok	soh	teʔɛlaŋ	te:ʔɛn	attay	qiʔ		in
48	sge-k2002berau15-GunungSari	ujo?	kabiŋ	ta:?u	takəd	ma:sat	janan	abɛ?		natuŋ	mano:	abu:	anɪt	jelaʔli:?	batek	tulaŋ	təna:?i	atay	itɛ?		tisən
49	sge-k2002berau16-LongLaai	quy	mənleyak	mənaw	hul	pənɛw	panəntal	tawɛ		jɔ?	əmβa:p	awaw	la:s	qukok	soh	tulaŋ	usus	tay	qɪ?		ɪn
50	whk-sori2003class12-UmaUjok		kaβiəŋ	ta?u	takət	masat		nai		natuŋ	mano		anit	likut		tulɛaŋ	tənnɛ	attai		la?ip	tisən
51	KEN-sori2003class20-LongTap		kabiəŋ	taəʔu	qqəm	masat		nɛ		pəlaŋui	mənnoŋ		anit				budɛ	atai		sunan	sən
52	KayanicProto-SM		hulay	ta?u	pa?an	panaw	jalan	nai			masap		las	la?uŋ	butit	tulaŋ	təna?i	atay		la?ip	

Figure 14. Snapshot from the EPI survey team database 'epi2019northborneo4gabmap'

This information would inform the language varieties chosen as part of the lexicostatistical/ phonostatistical analysis later in the process. More specifically, this data showed a signal overlap between Kayan and Kenyah languages which was corroborated by Soriente (2003) which made it sensible to test Nyibun and Pua' as part of the Kayanic group.

The data were then coded and consolidated by data scientists on the EPI team. From there, I added my lists to database. I then compile a new analytical database where each sheet placed the two varieties in question side by side. The data were normalized to characters, following the pattern, employed in Smith (2017a, 2019) during the analysis. Lexicostatistics/ phonostatistics analysis of the 17 language varieties chosen was performed for each pair. Language varieties were not compared across subgroups as, based on qualitative observation, the likelihood of intelligibility across subgroups is quite low and such an analysis would be beyond

the scope of this thesis. This comes out to 41 language comparisons or approximately 8,200 apparent cognate pairs analyzed for a dataset of nearly 3,000 lexical items collected. Here is a link to the lexicostatistical/phonostatistical analysis datasheets: https://bit.ly/3uZzgF7

3.7.1 Lexicostatistics/Phonostatistics: Statistical methods for analysis

Lexicostatistics/phonostatistics is a statistical method for analyzing the phonological structure and the lexical composition of a given set of language varieties. In many ways, these methods are not separate, phonostatistics is a deeper level of lexicostatistics as the lexical similarity is judged in the assembly of apparent cognates and phonostatistics more precisely calculates their similarity distance. As mentioned before, some use Levenshtein distance to calculate phonostatistics; however, because the Gapmap system—a common tool of measuring Levenshtein distance—was not operating for most of the year of the writing of this thesis, this type of analysis was not particularly feasible. Mckaughan (1964: 98) refers to *phonostatistics* as:

> …[A] statistical measure of language divergence,…a measurement of degrees of distance,…a statistical expression of the logical difference between pairs of related languages. In practice lends support to the lexicostatistical conclusions, and also gives further comment on language relations.

This is how lexicostatistics/phonostatistics is used here. Nahhas & Mann (2006; and also in Nahhas 2007) outline the procedures for performing lexicostatistics/phonostatistics analysis. While they do not explicitly states this, their statistical method is known as a Sørensen-Dice index. Here are the equations of this method, first to calculate the lexical structure and phonological similarity within a pair of apparent cognates (A) and then to calculate the lexical similarity across the entire language variety (B).

(A)

$$c = \frac{2p_{x \cap y}}{p_x + p_y}$$

(B)

$$ID = \frac{c\{2\ |A \cap B|\}}{|A|\ +\ |B|}$$

Allow me to explain these calculations. In the first equation, c is the score for any two apparent cognates. Variable x represents the data point of one set (or language variety wordlist) which 'intersects' with y point in another set (or wordlist) phonetically and semantically—that is, by being possible apparent cognates. p is the unity of those intersections which is divided by the number of points in those sets that intersect. If these intersecting points adhere to parameters set before the analysis they 'pass' and are apparent cognates. The parameters I set were those suggested in Nahhas & Mann (2006: 38-39).

Consonants must be exactly the same to be considered category 1 in similarity. Consonants were category 2 if they differed by 1 place or manner of articulation. They were considered category 3 if they differed by 2 or more places or manners of articulation. For vowels, vowels that were exactly the same or one place of articulation away were considered category 1. Two places were considered category 2. Very different vowels of three or more places away were considered category 3. Here is a fictitious example of the labeling process.

Table 4. An example of alignment and the analysis of one apparent cognate

Wordlist A **Wordlist B**

kmafeʔeʔi ‘food’ knagɛʔæʔɑ ‘food’

k	m	a	f	e	ʔ	e	ʔ	i
k	n	a	g	ɛ	ʔ	æ	ʔ	ɑ
1	2	1	3	1	1	2	1	3

If the number of category 1 designations is other 50% and the number of category 1 and 2 designations is over 75% then the apparent cognate ‘passes’ and is assigned a ‘1’ (i.e., c = 1). If the apparent cognate fails then it is assigned a ‘0’ (i.e., c = 0). The apparent cognates in the table above passes since over 50% of the phones are category 1 and over 75% of the phones are either category 1 or category 2.

The next procedure (represented by the second equation or the Sørensen-Dice index). To arrive at the ID or the ‘intelligibility distance,’ the intersection of set A and set B (or wordlist A and wordlist B) which, of course, is the sum of all ‘passes’ or all points where c=1 is calculated and multiple times two. This is then divided by the number of sets which is again two. From there a percentage is given which represents the ID or what is referred to in full matrix presented in the next chapter as the ‘Intelligibility Similarity Percentages Based on Lexicostatistical/ Phonostatistical Comparison.’

While the SIL standard is usually set at 60% and others have suggested a 70% (Nahhas & Mann 2006), here we use Romaine’s (2000:5) more restricted criteria of 81%. Romaine’s 81% threshold is used as the ‘cut-off’ for language varieties being considered dialects of the same language. Qualitative evidence for the validity of his threshold comes from Southwell's description of Kayanic intelligibility. While he mentions that Kayan speakers can communicate with ease over the vastly expansive territory in which they inhabit, for what we label as Baram

Kayan and Rejang Kayan, “the variants in pronunciation, intonation and vocabulary render conversation during initial contact more difficult, but communication is quite possible” (Southwell 1990: ix). Note that the findings below put these two varieties to 90% which suggests that if there is minor difficulty at this level and that comprehension difficulty would increase successively with a decrease in the intelligibility percentage.

After arriving at the intelligibility percentages, clustering took place in the form of geometric figures, plotted in conceptual space. Clustering seems to be an important component of phonostatistics (Grimes 2001: 88). Clusters (or subclusters, clusters, and superclusters in Anderbeck’s (2018) terminology) are presented as dialect chains, language chains—specially defined here as successive wave patters of unintelligible but closely clustering languages (i.e., language varieties at 79% cluster closely but are still considered separate languages), and projected related languages of very distant similarity in lexical composition or phonological structure. More on this in Section 4.5.

The percentages and clusters are considered the ‘first component’ of our cartography of diversification—that is the conceptual component showing implicative coherence—while the geolinguistic mapping is considered the second component. This component establishes geographic coherence which is then interpreted in light of environmental niche factors related to Kayanic ethnology, cultural geography and language contact.

3.7.2 Data Archiving

All original language documentation from this study will be archived among various documentary databases. All documents and recordings (audio and video) will be filed with *Kaipuleohone*—an archive for language documentation for the Asia-Pacific area at the University of Hawaii. Copies will be archived at the Bloomfield Language Institute (BLI)

(https://bit.ly/3iqoPaj). Copies will also be filed with the Universitas Negeri Malang and the Universitas Borneo Tarakan—as per the requirements of my research visa—so that local scholars and community members have access to the data. Language 'sample' videos have also been housed at *WikiTongues* (a YouTube-based archival platform) providing general public exposure for language communities—a boon for local communities aiming to protect their traditional lands.

> Mahakam Kayan (WikiTongues): https://bit.ly/3uZyY0Z
> Bahau (WikiTongues): https://bit.ly/3PsIVvQ

Copies of the wordlists will be submitted to the *Austronesian Basic Vocabulary Database*, to further the comparative work conducted on the Austronesian language family as a whole. The comprehensive database for this study will be filed on *Zenodo*—a scientific database platform—for the sake of scientific rigor (verifiability and replicability).

3.8 Summary: The purpose of the procedures?

The purpose of the procedures—the enactment of the methodology—as Cooper (1980) mentioned in the quote at the beginning of this chapter is "for the service of ideas." Indeed, the "information they are designed to collect [must be] worth gathering." In this methodology chapter I have outlined the methods, techniques and procedures required to shed light on the diversification of language from the perspective of Borneo. Let us now turn to the analysis to understand the nature of language diversity within the Kayanic language family and the environmental 'nudges' that provide clues for how it came to be this way.

CHAPTER 4

Kayanic Languages and Kayanic Dialects

Consuetudo loquendi est in motu.
'Everyday language is always in motion.'
—Marcus Terentius Varro

4.1 Introduction

The first component of our 'cartography of diversification' is the calculation and mapping of intelligibility distances based on lexicostatistical/phonostatistical comparison. These lexical similarity results allow us to map, in conceptual space (clusters), the *in vivo* divergence or convergence of these languages as features diffuse (or not) from one language variety to another. First comes a conceptual mapping, followed by geographic mapping—the second component of our 'cartography of diversification.'

4.2 Previous research

As mentioned in Chapter 2, very little scholarship has focused on Kayanic languages (Smith 2015). Within this work, very few scholars have worked on Kayanic dialectology, particularly from a lexicostatistics perspective (Blust 1974a, Hudson 1967; 1978, Soriente 2003, Kroeger 1998, Masfufah 2018, Omar & Norahim 2020). No scholarship, to my knowledge, has taken a phonostatistics approach to these languages. Blust (1974a, 1974b) offers some lexicostatistical analyses for various Kayan varieties (spoken at villages, Long Atip and Uma Juman) as well as Ngorek or Murik (from Long Semiang).[14] I have labeled these varieties for our purposes: Long Atip Kayan [lak], Uma Juman Kayan [ujk], and Ngorek [ngo]. The first comparison is based on wordlists of 100 words while the second is based on a 180-item wordlist (Kroeger 1998: 42).

[14] Blust (1974) uses the term 'Murik' in his paper.

LAK
| 72 UJK
| 56 58 NGO

LAK
| 84 UJK
| 65 66 NGO

The relative distance between Blust's calculations is what is important (cf. Kroeger 1998) here.

In an overview of the languages of Sarawak, Kroeger (1998) discusses some Kayanic lexicostatistics data in comparison to other languages of Borneo. Based on the language varieties from Long Dunin (Kenyah) and Uma Juman (Kayan), Kroeger presents data from Blust (1974a) showing a lexicostatistical score of 41%. This is not a definitive relationship between Kenyah and Kayan because Kroeger (1998: 48) finds a closer relationship between Long Dunin and another Kayan language variety, Uma Bawang (51.5%). He goes on to present Blust's Kayan and Murik comparisons (outlined above). Altogether, he uses this evidence to suggest that Kayanic is closer to Kenyah than any other language groups, yet that there is not enough evidence to consider them as one Kayan-Kenyah family or that Kayanic should be included in Blust's larger, overarching North Sarawak subgroup.[15] The propinquity between Kayanic languages and Kenyah languages is probably a function of linguistic and cultural contact over centuries.

Soriente (2003: 34) offers some lexicostatistical data for various Kayan and Kenyah languages. Her focus here is, of course, on Kenyah. Pua' clusters with Nyibun at 67%. While they are all considered culturally Ngorek due to their ethnic histories, Nyibun clusters more closely with Kenyah varieties ranging from 78-85%. Smith (2017) does not include Nyibun in his Kayanic classification, although Soriente (2008: 59) classifies it under Kayanic. Soriente mentions, that Nyibun, a language under severe language shift to Leppo' Ke' (itself an

[15] Work from Smith (2015) also supports this conclusion.

endangered language),[16] has replaced most of its vocabulary with borrowings from Leppo' Ke,' Kayanic phonological features[17] still persist and therefore it should be classified as Kayanic.

With numbers similar to what I calculated for Pua' and Nyibun, further investigation is needed to understand the relationship between Nyibun and other Murik-Merap languages.

Pua'
| 67 Nyi

Soriente (2003: 35) also mentions that "Pua' [stands] somewhat isolated being at an intermediate level between Kayan and Kenyah."

Although these languages are not closely related, Masfufah (2018: 35) compares Bahau with Benuaq and Tonyooi (or Tunjung), in order to measure the distance among these disparate Bornean languages.

BHV
| 27 BEN
| 31 46 TJG

Masfufah (2018: 35) offers this tree to show, not the family or subgrouping relatedness, but the distances between the languages.

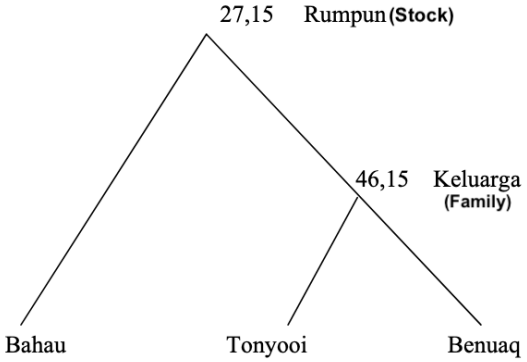


Figure 15. Masfufah's Distance Tree. Adapted by DeAndré A. Espree-Conaway. Accessed 14 September 2021.

[16] Nyibun and Leppo'Ke' share a lexicostatistical score of 82%.
[17] Nyibun shares phonological traits with Pua', in particular, but also with Kayan varieties as well.

Using more nuance relatedness criteria,[18] Masfufah (2018: 37) concluded that all of the languages are related on a larger scale. Their data is probably picking up the phylogenetic signal of their being Austronesian languages. Musfufah further details this relatedness by explaining that Tonyooi and Benuaq are of the same family while Bahau are related to these languages by being of the same 'stock.' Again, the phylogenetic signal between Bahau and the others is probably Austronesian while the signal between Tonyooi and Benuaq is probably the fact that they belong to the Greater Barito subgroup (Eberhard et al. 2021). Naturally, Masfufah (2018: 37) concludes that all the languages of their study are unintelligible.

Omar & Norahim (2020) has also offered some lexicostatistics analyses. They compare Lepo' Tau, a dialect of so-called Mainstream Kenyah [xkl] (Eberhard et al. 2021), with Kayan from the "tributaries of the Upper Baram" (Omar & Norahim 2020: 94) or Baram Kayan [kys].

KYS
| 49 XKL

While in a footnote, they offer that they have "also conducted a lexicostatistic comparison of Kayan related languages" (Omar & Norahim 2020: 97), no such data is given in the paper.

4.3 Lexical similarity results

It should be clearly noted that questions of 'relatedness' or 'true cognicity' are not dealt with here since the comparative method was not part of this study. These data, however, do provide reifying evidence, 'backing up' the conclusions of historical-comparative studies conducted elsewhere (Smith 2017; Soriente 2003).

To reiterate from this previous chapter, the similarity between the phonological structure

[18] Masfufah (2018: 32) uses a very nuanced set of criteria to derive conclusion about relatedness from lexicostatistics: percentages between 81-100% constitute a language [the same as our criteria]; 36-81% are in the same family; 12-36%, the same 'stock;' 4-12%, the same microphylum; 1-4%, the same mesophylum; and <1%, the same macrophylum.

of a pair of lexemes is calculated in the following manner:

$$c = \frac{2p_{x \cap y}}{p_x + p_y}$$

If c = 0.75 or higher[19], the pair 'passes' and is considered an apparent cognate. All apparent cognates which comprise the set of c are divided by the total number pairs in the wordlist comparison space. This produces the *ID* or *intelligibility distance* between two language varieties.

$$ID = \frac{c\{2\ |A \cap B|\}}{|A|\ +\ |B|}$$

The intelligibility percentages are presented in table 5 below. Lexicostatistical/phonostatistical comparisons were run for the various language varieties within their respective subgroups (as established in Smith 2017). The data is presented to three decimal places in accordance with the tradition in quantitative dialectology.

[19] Note the caveat that 50% or more of the phonemes must be exactly the same for consonants and exactly the same or one articulatory place away for vowels.

Table 5. Similarity Percentages Based on Lexicostatistical/Phonostatistical Comparison

KAYAN VARIETIES

BAH						
|0.767	BUS					
|0.789	0.821	KRK				
|0.451	0.445	0.407	MK			
|0.707	0.815	0.829	0.327	BK		
|0.755	0.878	0.866	0.377	0.926	RK	
|0.795	0.815	0.843	0.395	0.859	0.913	DDK

NGOREK (MURIK-MERAP) VARIETIES

NGO				
|0.734	HUE			
|0.608	0.525	MER		
|0.649	0.577	0.649	NYI	
|0.707	0.833	0.481	0.594	PUA

MODANG VARIETIES

WEH				
|0.614	KM			
|0.545	0.876	LGA		
|0.573	0.639	0.615	SEG	
|0.530	0.618	0.562	0.751	PKE

4.4 Kayanic languages and Kayanic dialects

The criteria for intelligibility or the 'cut-off thresholds' described here are more complex than usual. The SIL rule of thumb for lexicostatistics is to place the threshold below which structural factors are highly likely to prevent intelligibility at 60% (Grimes 1995; Blair 1990). Nahhas & Mann (2006) suggest a 70% threshold. Romaine (2000:5) and others (Jackson 1983; Swadesh 1954; Ellis 2007) have used the higher threshold of 81% for marking the difference between 'language' and 'dialect.' Ellis reminds us, however, that the threshold indicating that there is no structural barrier to communication is the point, not that the percentage actually indicates differentially how well speakers understand each other, as he mentions that two Micronesian languages in his data had a intelligibility score of 95%, yet were still unintelligible for other reasons (Ellis 2007: 13). There is some evidence that ordinarily if a language variety scores at 85% or above, that languages are likely to be intelligible enough to share written literature, but, in the case of education programs, no action to create a joint orthography/ literature between any two groups should ever precede without further testing (Grimes 1992: 32).

Anecdotally speaking, while Spanish and Italian share a lexical similarity of 82%, deep communication is hardly possible. My spanish-speaking friends have been able to watch Italian movies, getting the gist of chunks of text (rather than individual sentences) enough to follow the story well. However, when in Italy, I saw a Spaniard ask for directions to a random person on the street. While they both spoke Spanish and Italian respectively, it was not without slowing the rate of speech significantly and perhaps other linguistic accommodations. Since asking for directions was quite strained, one can imagine that deeper communication would be precluded.

Instead of relying on one threshold, we will use them in various ways. The thresholds used in this study are organized thus:

Level 1	–	No Structural Barriers to Intelligibility ≥81%
Level 2	–	Highly Variable Possibility of Intelligibility 61-80%
Level 3	–	Low Possibility of Intelligibility 60%
Level 4	–	No Inherent Intelligibility <60%

Figure 16. Predicted Intelligibility Thresholds

Logically speaking, there is full consensus that anything 60% and below is completely unintelligible. So Level 4 is the 'No Intelligibility' level. Level 3 scores imply both that the possibility of intelligibility is low and that if there is intelligibility, it too is indeed low. Level 2 is the most ambiguous level. While the possibility of a language variety being intelligible is medium to low, if there is intelligibility it is unlikely to be high—meaning complex linguistic interactions are probably precluded in this scenario. Level 1 implies that there are probabilistically no structural barriers (phonological or lexical) to intelligibility, but further testing is needed to truly elucidate whether there is intelligibility and the quality of that intelligibility.

Here only Level 1 language varieties are considered 'dialects' of a common language, while all other levels are considered 'separate languages.' If the highest threshold is not met, then I would recommend labeling them as different linguistic systems that are likely to only be crossed by linguistic accommodation and/or language learning.

Among the Kayan language varieties, it is clear that Rejang Kayan, Data Dian Kayan, and Baram Kayan are all dialects within the same language, despite there being slightly more distance between Baram Kayan and Data Dian Kayan (86%). Rejang Kayan and Data Dian Kayan scored 91% and Rejang Kayan and Baram Kayan scored 93% (the highest percentage among all Kayanic languages). A dialect chain is in formation here. It also includes Busang through its Level 1 connections with Rejang Kayan and Kayan River Kayan through its closeness to Rejang Kayan. Kayan River Kayan and Data Dian appear to be mutually intelligible

as well. A moderate distance of 84% is indicated for these dialects. Busang and Kayan River Kayan, Busang and Data Dian Kayan, Busang and Baram Kayan, and Kayan River Kayan and Baram Kayan all lean toward the edge of intelligibility at 81 or 82%.

Bahau seems to be slightly divergent from the other language varieties, falling below the 81% threshold. This suggests that the possibility of *inherent* intelligibility is variable, although learning or linguistic accommodation may overcome this. It is curious that the relationship between Bahau and all other language varieties seems to coherently range between 70-80%. According to this data, I would consider Bahau a separate language from the Rejang-Data Dian-Kayan River-Busang-Baram Kayan dialect chain. Mahakam Kayan[20] is, of course, completely unintelligible with the other language varieties. In sum, there are three languages of Kayan: Bahau, Mahakam Kayan, and, what I will call, the Central Kayan Dialect Chain.

In the Ngorek category, the distribution of the languages is much more spread out in terms of intelligibility. Also, the quality of intelligibility is much lower overall compared to the Kayan language varieties. What is clear, however, is that the relationship between these languages form a language/dialect chain. Huang Bau and Pua' are the only varieties that can be said to be highly likely to be dialects of a common language among the Ngorek data. At 83%, it should be noted that they learn toward the bottom edge of intelligibility. Ngorek and Hueng Bau and Ngorek and Pua' have scores of 73% and 71% respectively. This places these varieties within the Level 2 'highly variable category.' Ngorek and Merap (61%), Ngorek and Nyibun (65%), and Merap and Nyibun (65%) all fall within the low probability of intelligibility category. These varieties, if they are indeed intelligible, the quality of the intelligibility is also probably

[20] In addition to the classification of Mahakam Kayan within Kayanic in previous works (Guerreiro 1996), despite the low percentages not typical of related languages, several basic sound changes provide evidence for the placement of Mahakam Kayan within Kayanic. Mahakam Kayan appears to have undergone the Kayanic innovation of word-final glottal stops after final vowels (cf. Smith 2017: 60). There is also a vowel deletion and subsequent word-initial place assimilation resulting in voiceless nasals (e.g., *ku > m̥, as in *kuman > m̥man 'eat').

low. All other intelligibility comparisons are unambiguously unintelligible [Hueng Bau and Merap (53%); Hueng Bau and Nyibun (58%); Merap and Pua' (48%); Nyibun and Pua' (59%)]. These data suggest that there are four Ngorek languages: Ngorek 'Proper' or Murik, Hueng Bau-Pua', Merap, and Nyibun.[21]

Modang is extensively diverse internally. Only one comparison meets the criteria for possible intelligibility. Kelinjau Modang and Long Gelat have a score of 88%. All other varieties have to be considered at this point completely separate languages, although it should be noted that a kind of language chain does emerge among these language varieties.

The next pair that comes closest to the threshold of intelligible, but is still significantly below it is Gaai and Punan Kelai at 75%. Within the 60s range, we have Wehea and Kelijau Modang (61%), Kelijau Modang and Gaai (64%), Kelijau Modang and Punan Kelai (62%), and Long Gelat and Gaai (62%). All other comparisons fall unambiguously below intelligibility: Wehea and Long Gelat (55%), Wehea and Gaai (57%), Wehea and Punan Kelai (53%), and Long Gelat and Punan Kelai (56%). I am sure that you have noticed at this point that Wehea appears to be the most divergent of all Modang language varieties. Based on this data, there are four Modang languages, Wehea, Kelinjau Modang-Long Gelat, Gaai, and Punan Kelai.

It should be reasserted at this point that all proposals of 'dialects' of the same language are tentative. While percentages above 81% probabilistically have a good chance of being dialects of a common language (cf. Shackelford 2019), this can only be elucidated through true intelligibility testing such as recorded text tests (RTT), recorded story playing (RSP), sentence repetition tests (SRT), oral proficiency tests, Picture-Pointing Task (PiPoT), etc. (Blair 1990; Shackelford 2019; Radloff 1991; Anderbeck & Yuditha 2020). This means that further study is

[21] Again, note that Nyibun's status as a Kayanic language (versus a Kenyah language) and thus a Ngorek language is still contested, even though it appears to cluster well with Ngorek varieties.

needed to confirm the Rejang-Data Dian-Kayan River-Busang-Baram Kayan, Hueng Bau-Pua', and Kelijau Modang-Long Gelat language designations that I have proposed here.

Because sociological factors can also impact intelligibility (or the willingness to engage with intelligibility), the fact of intelligibility must also be evaluated through qualitative social dialectology tools, such as interviews, questionnaires, perceptual dialect mapping focus groups, etc. Bilingualism must be ruled out as well.

4.5 Patterns of intelligibility

As was mentioned in Chapter 1, a *dialectal-geographical radiation* approach is most appropriate for dealing with the interpretation of patterns of intelligibility. The wave theoretic approach requires two demonstrable patterns of coherence: *implicative coherence* and *geographical coherence*. Implicative coherence is a successive pattern of unity, establishing a kind of 'linguistic whole' among various language varieties. This looks like a kind of fading pattern from one language variety into another, rather than broken into discrete parts. Variety A overlaps with variety B which overlaps with variety C and so on and so forth, in conceptual space. Geographical coherence involves observing this implicational coherence on a geographic dimension. So language variety A is in a location near variety B which overlaps with variety C and so on and so forth.

Establishing these patterns of coherence is important in order to ensure that the features that support intelligibility are successively acquired from one language variety to another and did not happen to arise within each language variety independently of each other. This can also be thought of as *intelligibility flow*, as the features of intelligibility flow from one variety to another, maintaining their divergent or convergent identity. As Haas (2010: 651) explains, the distribution of intelligibility features must show a "geographic constellation" in "wave formation." We would

expect to see language or dialect chains, where, in concentric 'ripples,' different language variety comparisons go from 90+% to 80+% to 70+% to 60+% to finally no possible chance of intelligibility at all. Here is the idealized model from Chapter 1 that we expect to observe:

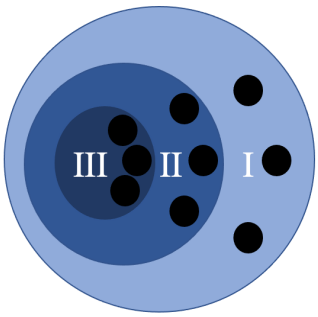


Figure 17. Idealized Model Map of Intelligibility Diffusion and Dialect Distribution

Note that we only deal with implicative coherence, and, therefore, only conceptual space, here. Geographical coherence will be dealt with in next section.

Several dialect and language chains establish the implicative coherence among the language varieties of Kayanic. The most well established dialect chain is the Central Kayan Dialect Chain which includes Rejang Kayan, Data Dian Kayan, Kayan River Kayan, Busang, and Baram Kayan. Many have noticed the relative homogeneity of the Kayan languages (Hudson 1967, 1978, Smith 2015, Smith 2017). Smith (2017: 92) appears to have similar results from his comparative reconstruction study, although he still divided the chain in three: Rejang Kayan-Busang, Data Dian Kayan-Kayan River Kayan, and Baram Kayan. Bahau, although not part of the dialect chains, still clusters with the other varieties in a kind of language chain. Mahakam Kayan is the only outlier that is historically related, but isolated in terms of intelligibility.

Within the Ngorek subgroup, Hueng Bau and Pua' form a dialect chain while all other varieties form a language chain, maintaining implicative coherence rather evenly throughout the subgroup. The Modang languages (Wehea, Kelinjau Modang, Long Gelat, Gaai, and Punan

Kelai) form a language chain within which Kelinjau Modang-Long Gelat remain close enough to form a dialect chain. Here is a mapping of language names and abbreviations.

Bahau	BAH	BHV
Busang	BUS	BFG
Kayan River Kayan	KRK	XKN
Mahakam Kayan	MK	XAY
Baram Kayan	BK	KYS
Rejang Kayan	RK	REE
Data Dian Kayan	DDK	--
Ngorek Proper or Murik	NGO	--
Hueng Bau	HUE	--
Merap	MER	--
Nyibun	NYI	--
Pua'	PUA	--
Wehea	WEH	--
Kelinjau Modang	KM	MXD
Long Gelat	LGA	--
Segai or Gaai	SEG	SGE
Punan Kelai	PKE	PUK

Below is a conceptual space representation of the language and dialect chains within Kayanic.

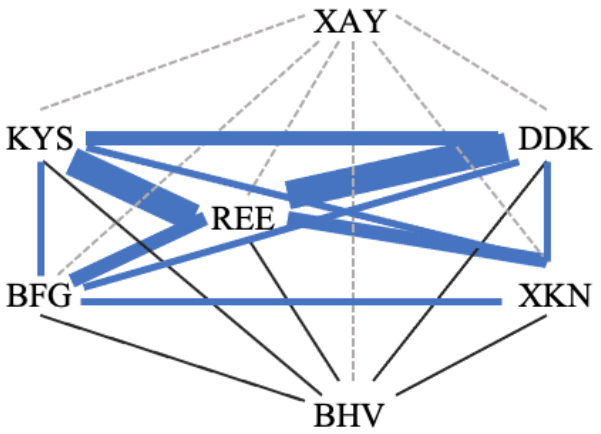


Figure 18. Conceptual Space Map of the Language/Dialect Chain Structure of Kayan. The blue forms the dialect chain with the line thickness indicating intensivity of intelligibility. The dark gray indicates the language chain. The dashed, light-gray indicate phylogenetically related languages outside the language/dialect chain.

Rejang Kayan appears to form a kind of center with other dialects linked to it to form the chain. It is the center while the other dialects emanate out in successive waves of reduced intelligibility at the periphery.

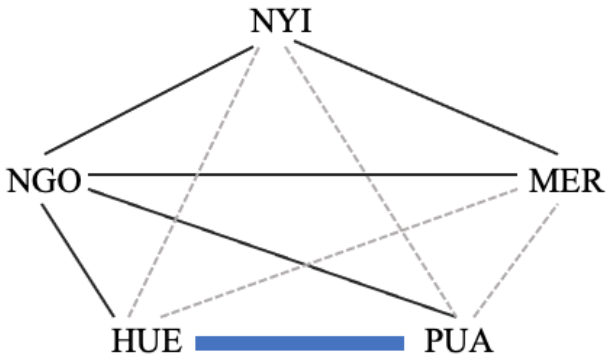


Figure 19. Conceptual Space Map of the Language/Dialect Chain Structure of Ngorek (Murik-Merap) subgroup. The blue forms the dialect chain with the line thickness indicating intensivity of intelligibility. The dark gray indicates the language chain. The dashed, light-gray indicate phylogenetically related languages outside the language/dialect chain.

Here, Hueng Bau and Pua' form the only dialect chain with all other languages forming a chain of unintelligible varieties or languages. Note that Nyibun and Merap only show a relationship to Hueng Bau and Pua' through Ngorek which shows a clear relationship to both.

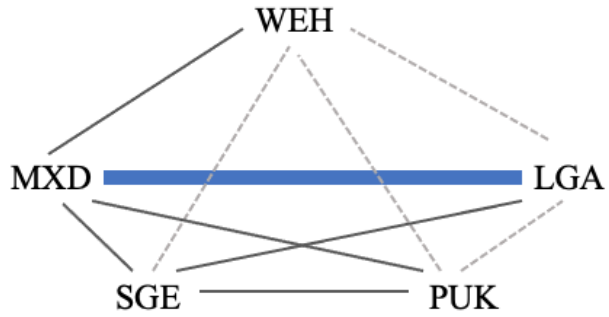


Figure 20. Conceptual Space Map of the Language/Dialect Chain Structure of Modang. The blue forms the dialect chain with the line thickness indicating intensivity of intelligibility. The dark gray indicates the language chain. The dashed, light-gray indicate phylogenetically related languages outside the language/dialect chain.

In this conceptual space map, only Kelinjau Modang and Long Gelat are close enough, among the Modang language varieties, to form a dialect chain. All others form a language chain with Wehea as the outlier, only barely connected to the chain through Kelinjau Modang.

Now that we have established the interrelationships among the various languages, we turn our attention to demonstrating implicative coherence. We can use conceptual maps similar to those in Haas (2010), to visualize the concentric waves of intelligibility among these languages. Below I present the maps for the Kayan, Ngorek, and Modang subgroups.

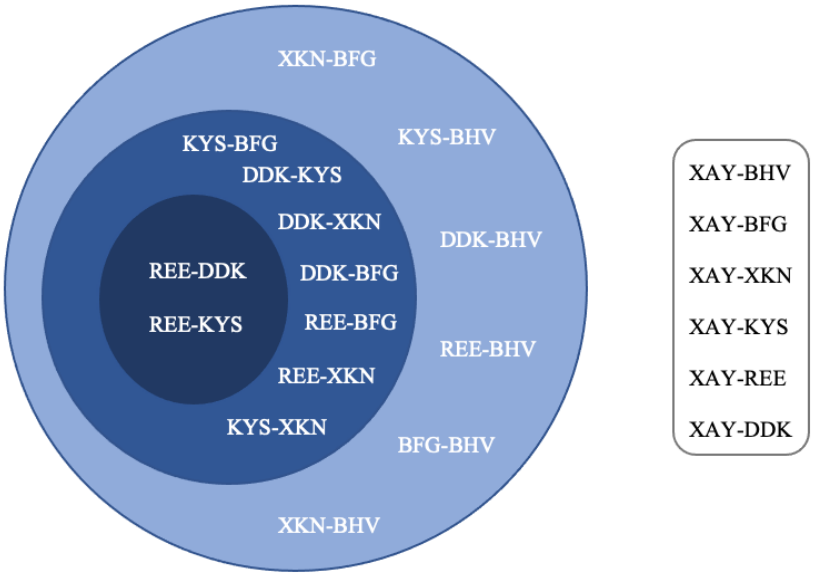


Figure 21. Idealized Model Map of Intelligibility Diffusion and Dialect Distribution for the Kayan subgroup.

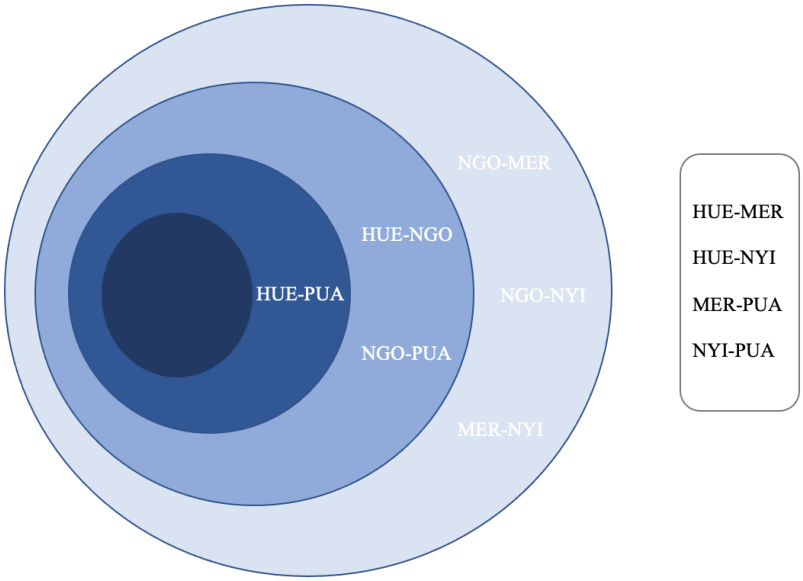


Figure 22. Idealized Model Map of Intelligibility Diffusion and Dialect Distribution for the Ngorek subgroup.

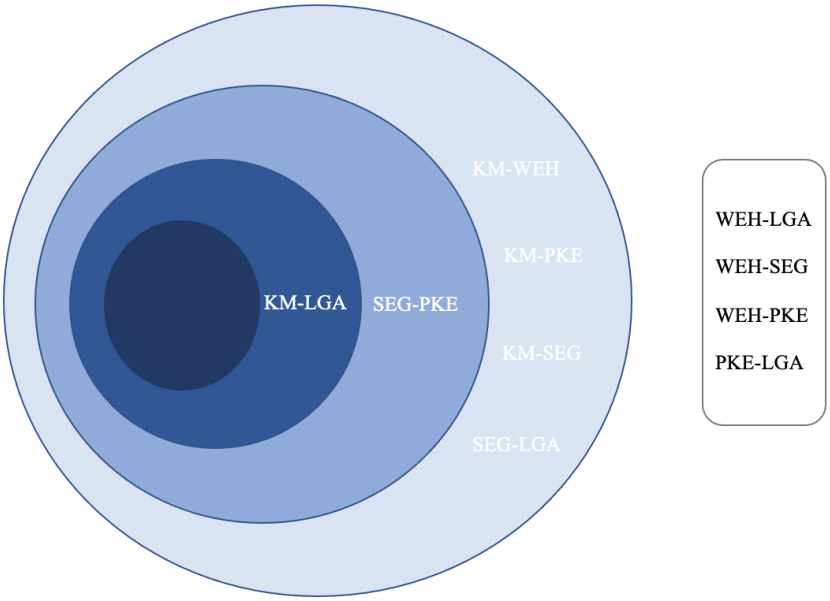


Figure 23. Idealized Model Map of Intelligibility Diffusion and Dialect Distribution for the Modang subgroup.

Now that we have observed the implicative coherence in its wave pattern (*wellenbild*), we must address *geographical coherence*, plotting the patterns we see conceptually into 'real' space within the physical and social geography of these speech communities.

4.6 Linguistic Map of Kayanic Languages

Here is a 'bird's-eye view' map of the Kayanic languages of Borneo (Figure 24). The very first impression is that shear distance probably plays a role in diversification in terms of being a physical niche factor. Distance as a niche factor has been found in other parts of the world (cf. Beyer et al. 2019). The area is highly riverine and the Central Borneo region coincides almost perfectly with the mountainous center outlined in green. Central Borneo becomes even more mountainous as you approach the Kayanic homeland—the Apo Kayan—as shown in the map in figure 25.

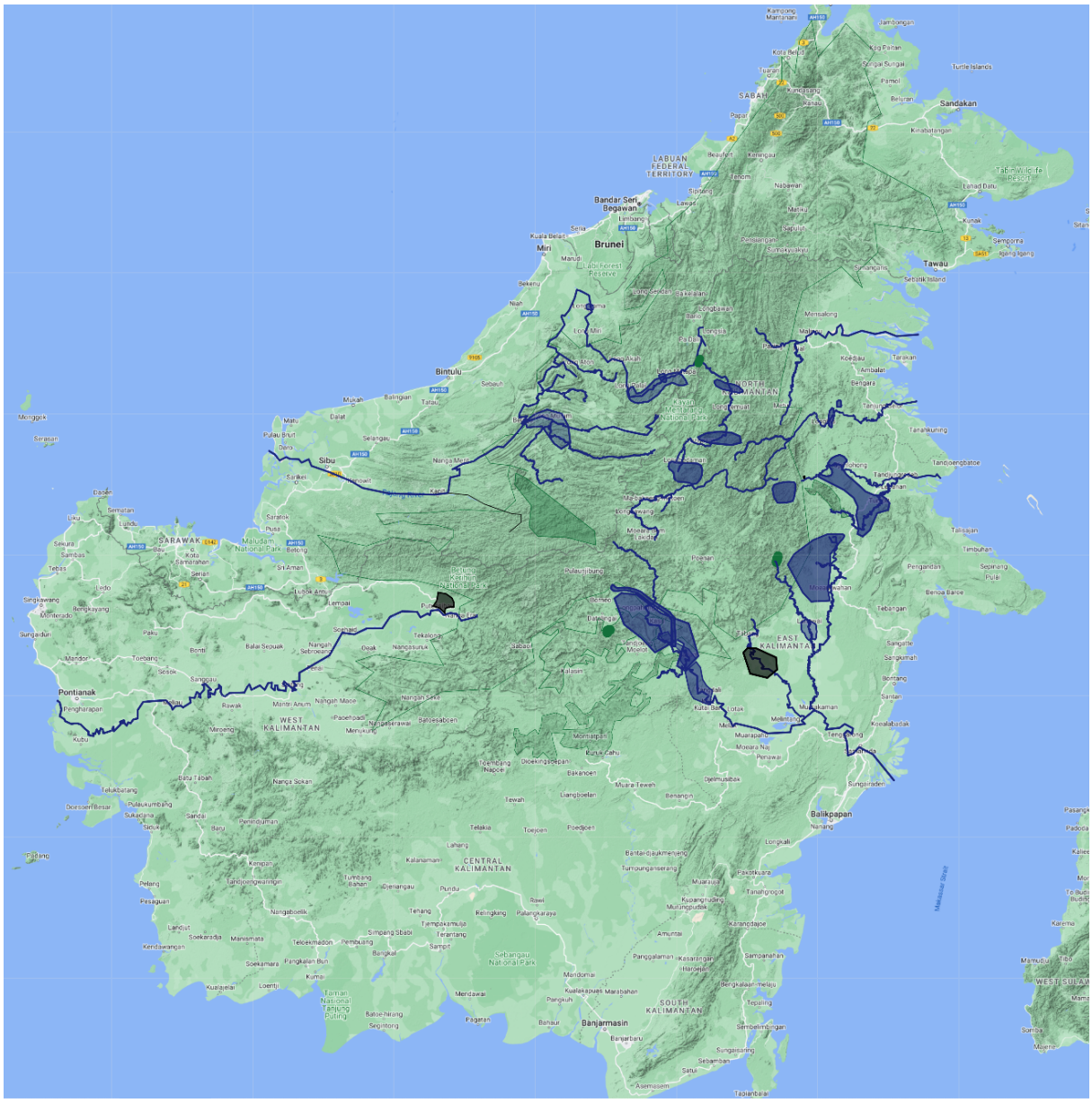


Figure 24. Bird's Eye View Map

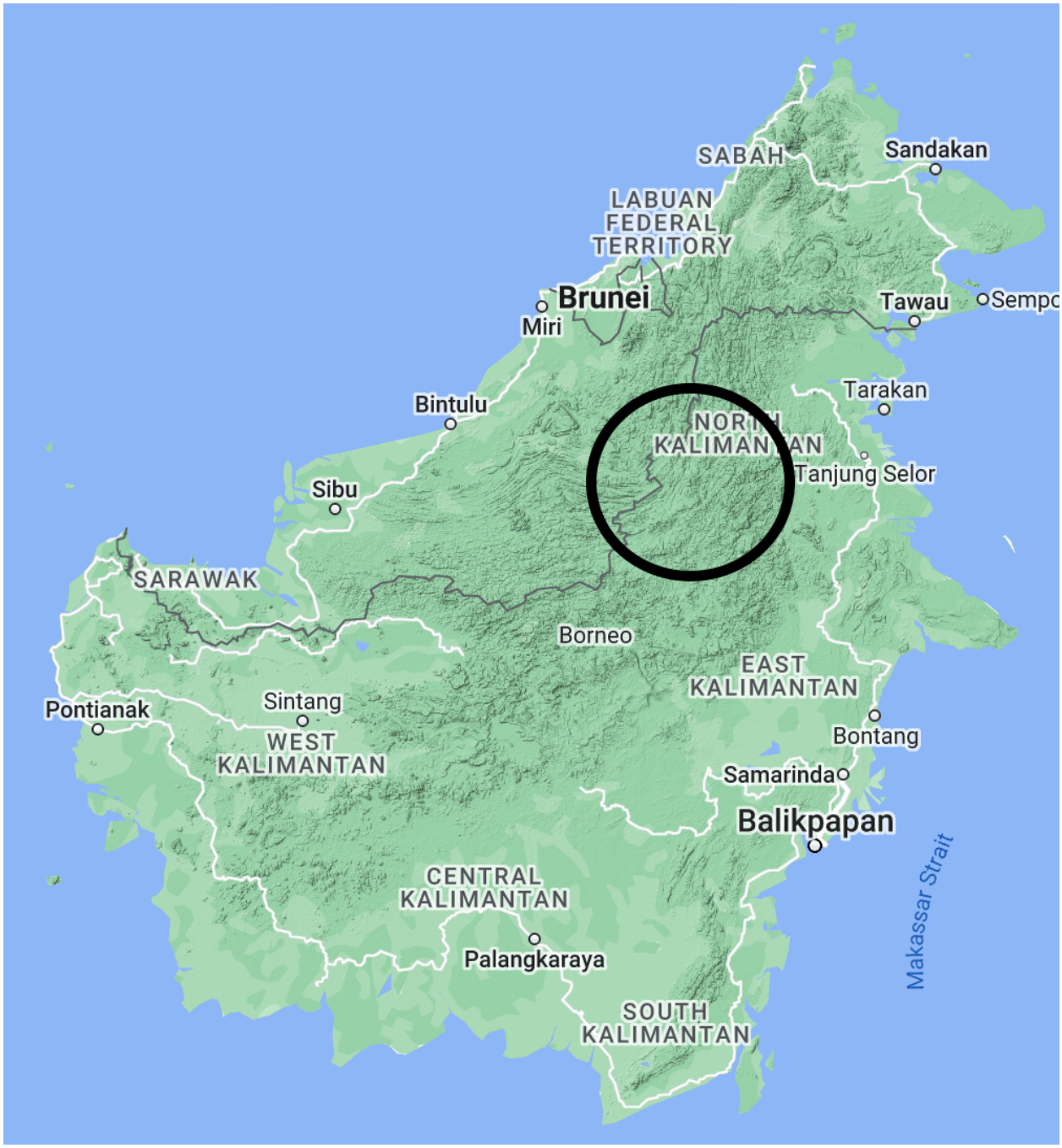


Figure 25. Apo Kayan

The isolation in the hyper-mountainous areas in and around the Apo Kayan may account for the faded signal that prevent linguists from connecting Kayanic language with other subgroups or within larger families on the island.

On these maps (Figures 27-33), the yellow marks represent rapids, the dark red marks

represent very dangerous or impenetrable rapids, and the green dots are very high altitude points on the island. The blue polygons represent language varieties studied in the study. The black polygons are those that are Kayanic for which there is no data. Figure 26 below labels all of the language varieties in this study.

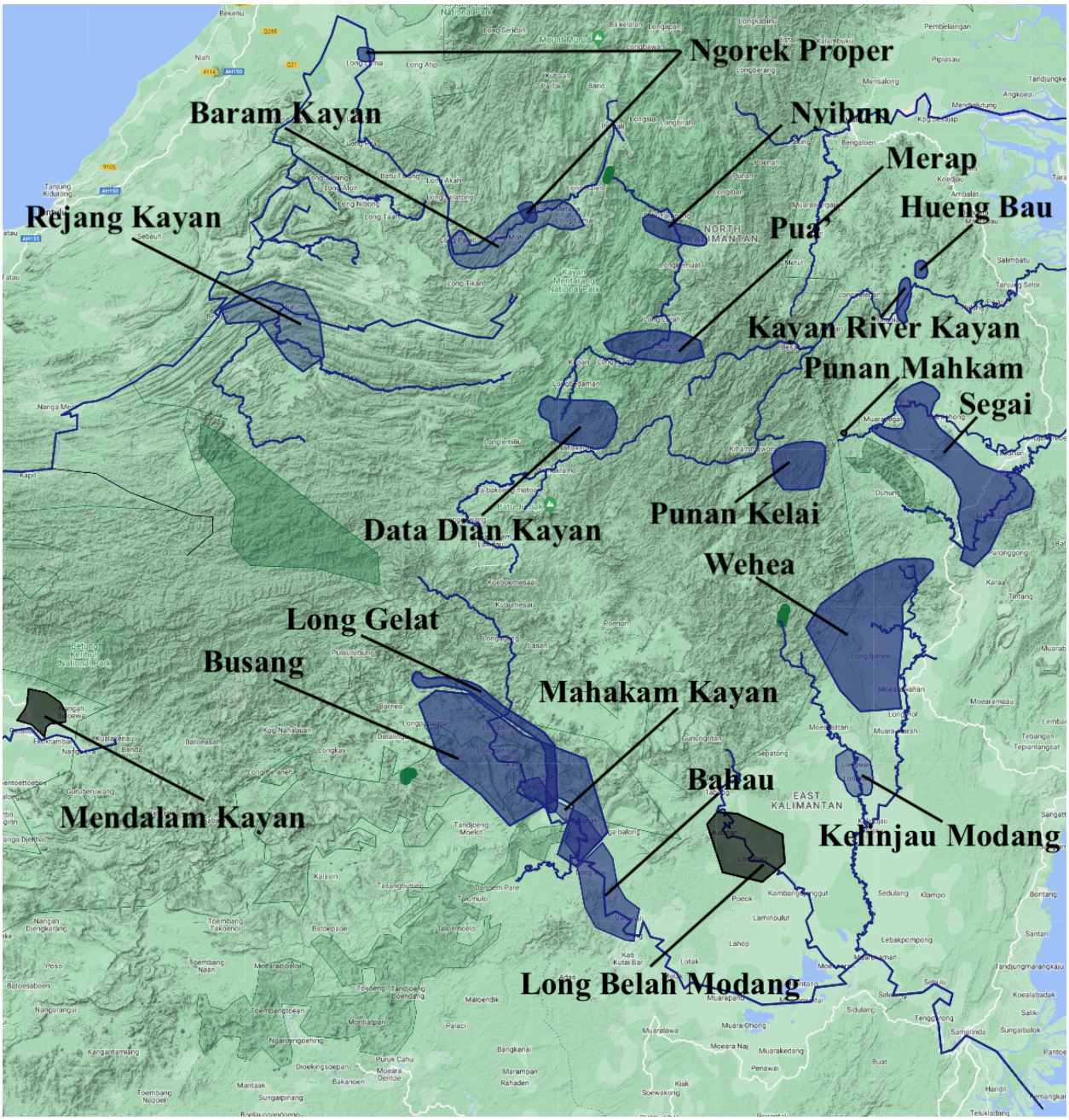


Figure 26. Zoomed-In Bird's Eye View Map

The mountainous center spine of Borneo, near the Apo Kayan (Figure 25) and above (politically dividing Indonesia and Malaysia) separate Baram Kayan, Rejang Kayan, and Ngorek from the other language varieties on the Indonesian side.

When we look at Rejang Kayan, the dialect area is capped on both sides by rapids (Figure 27). It is also surrounded by mountainous terrain. Note that the river, depicted in the middle of the right hand side of the image goes off into the mountains with no further connections to other populated waterways.

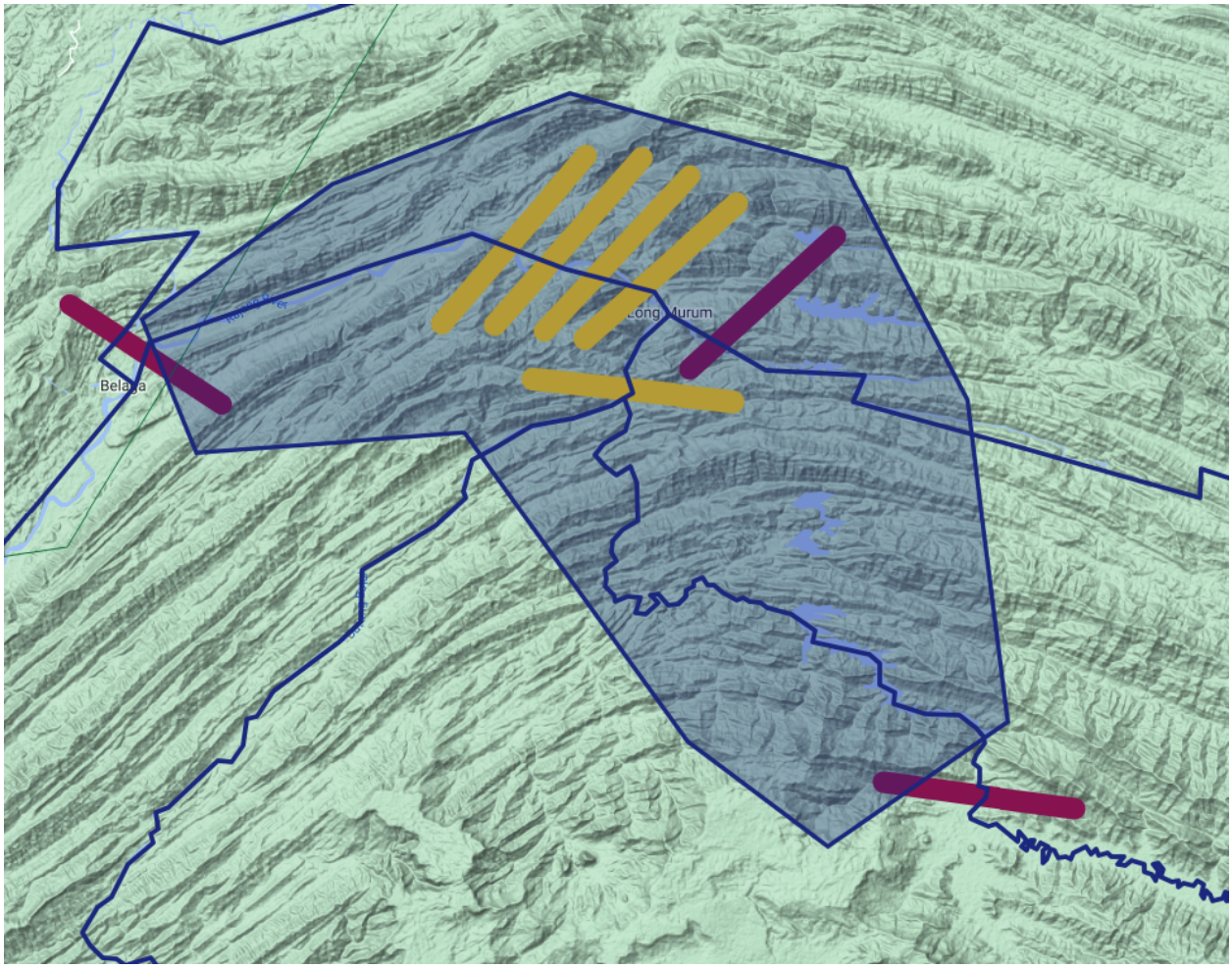


Figure 27. Rejang Kayan among the Rapids

Despite being capped by rapids, this cannot be the whole story as there are many rapids internal

to the dialect. This implies that rapids may be more social borders than physical in nature which the ethnographic record supports. Future studies should look into diversity within this community to see if it is as coherent as it seems despite the dangerous, impassable rapids located in the middle of it.

The waterways by which the Rejang Kayan might communicate with the Baram Kayan are blocked at several points (Figure 28). These rapids as physical-cum-social barriers, along with the mountains, would limit interaction in each other's daily communications (*praxis of language variation*), leading to dialect divergence.

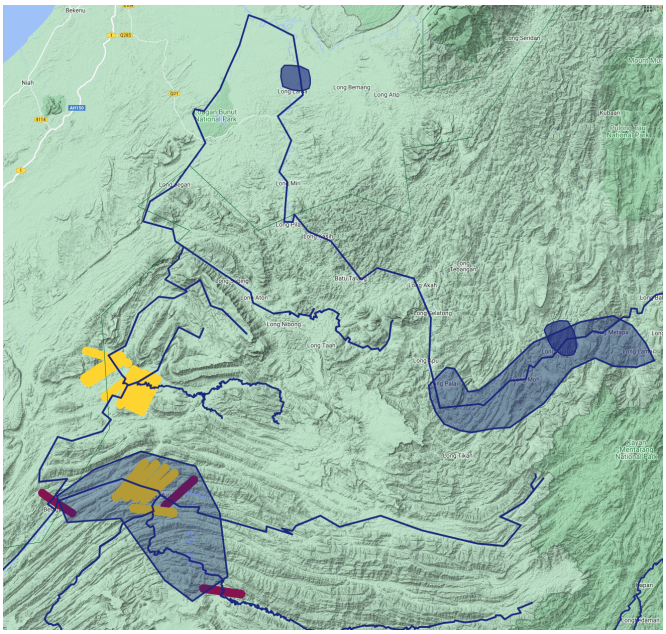


Figure 28. Rejang Kayan to Baram Kayan Waterways

It should be noted here that the physical niche factor's role is relatively weak here, since intelligibility flow among these dialects is still far from high. This may reflect a more recent migration from each other and/or sustained contact.

Ngorek proper or Murik is located in two distant spots, one of which is surrounded by Baram Kayan (Figure 29). This also attests to the reduced, but present role of the physical niche and plays up the role of the social niche (i.e., ethnic distinction). However, the Ngorek,

historically, did remain longer in the mountainous Apo Kayan, perhaps giving them more time to differentiate before migrating out.

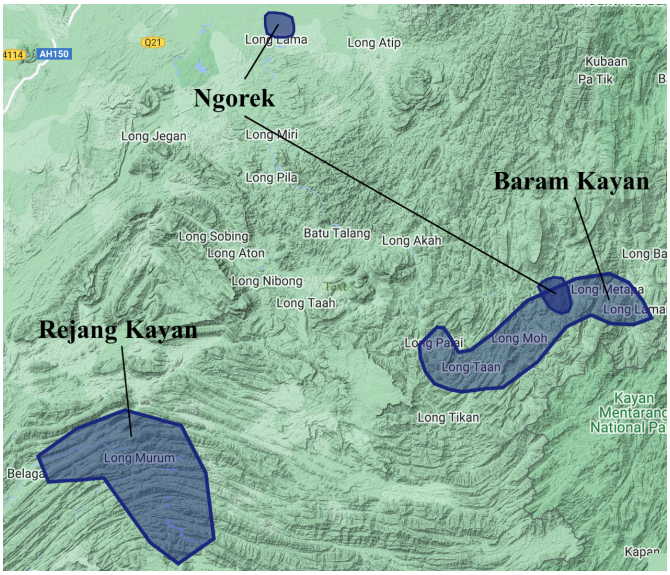


Figure 29. Map of Ngorek Proper

Now, let us examine some of the other language varieties: Kayan River Kayan, Hueng Bau, Punan Kelai, Nyibun, Pua', Data Dian Kayan, Punan Mahkam, and Segai. The Ngorek varieties, Hueng Bau, Nyibun, and Pua' are all separated by great distances and mountains (Figure 30).

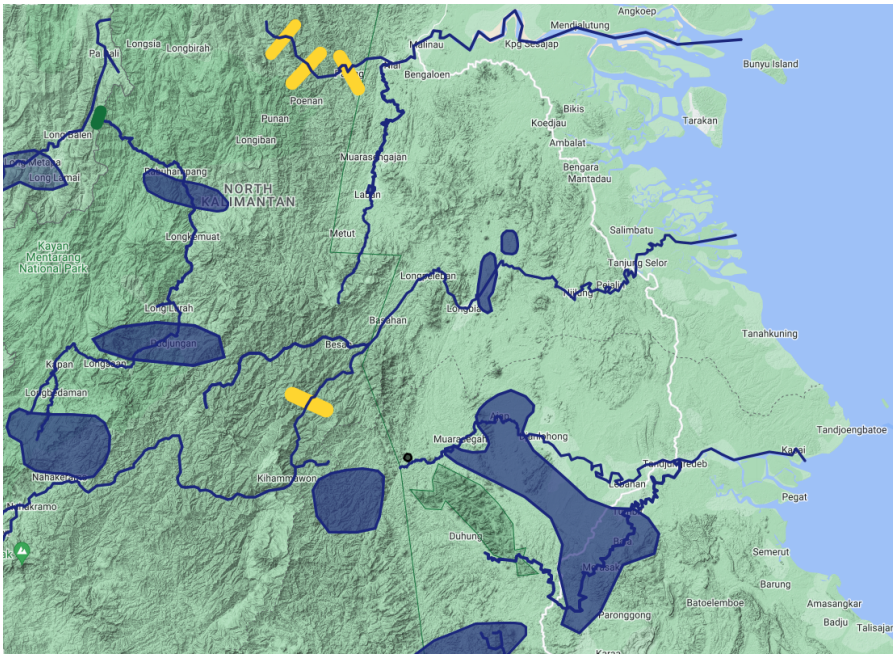


Figure 30. Map of North and Upper East Kalimantan Languages

The Kayan River Kayan is very isolated from other varieties, by distance, mountains, and rapids on the Kayan river (Figure 30). It accordingly diverges from other dialects of Kayan, there is still probable intelligibility flow. Data Dian Kayan remains ‘intermediate’ to all three other Kayan dialects. Despite rapids in connected waterways in Kayan River Kayan and steep mountains of over 5,000 feet, Data Dian Kayan maintains intelligibility flow with the other varieties. There is a waterway and a possible lower land crossing that may account for the high levels of intelligibility flow. Future studies should look further into this waterway as a source of trade and other linguistic and cultural contact.

Let us now turn to the Modang languages (Figure 30). Segai is separated by distance from all other Modang varieties. It is separated from the Punan Kelai by there being no connected waterways and mountains after 3,000 feet. The area inhabited around the Punan Kelai is quite rugged as well.

The Wehea are far upriver on the Wahau and rapids are met along the way. Landways between the Wehea and the Kelinjau Modang are relatively low which may account for why Wehea clusters closer to Kelinjau Modang compared to any other Modang language (Figure 31). Again, there is no data for Long Belah Modang (Figure 31).

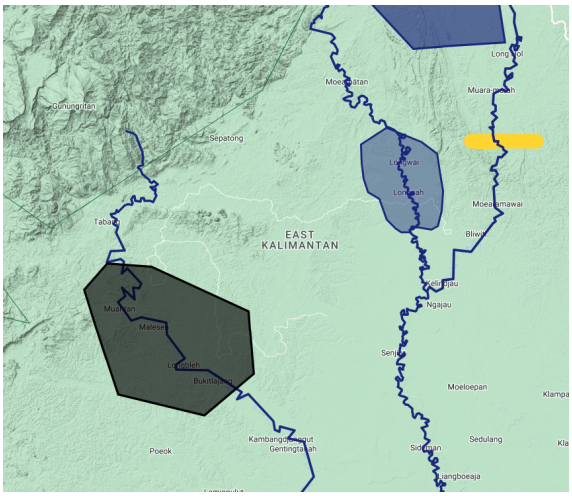


Figure 31. Map of Long Belah Modang

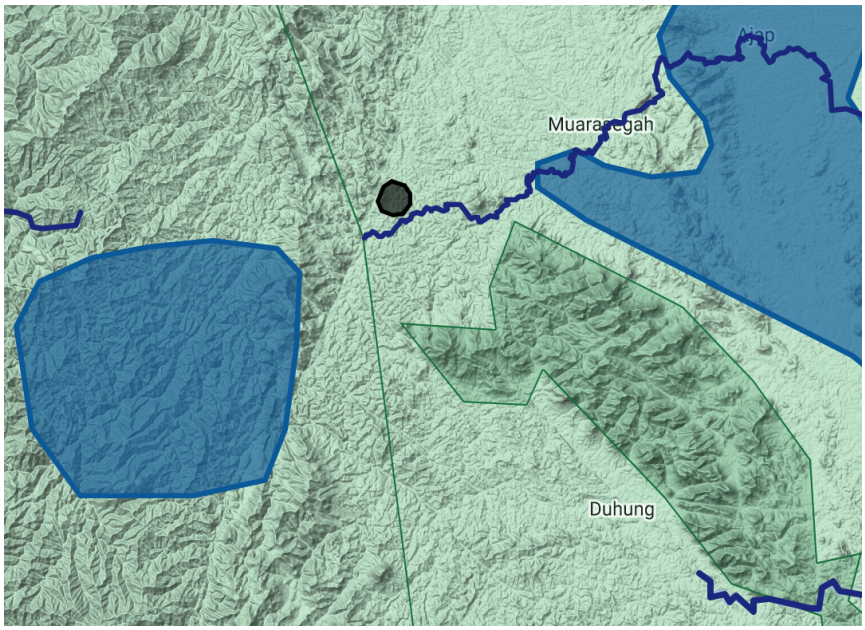


Figure 32. Map of Punan Mahkam

Not too far from the Punan Kelai community is the Punan Mahkam community—(black polygon) a Modang variety—for which we have no data (Figure 32). If a decision had to be made for these varieties, Punan Mahkam and Long Belah Modang should be free to be their own 'languages' before being proven 'guilty' of being a 'dialect' of a wider language.

The Modang variety, Long Gelat, is situated on the Mahakam, separated from the others by distance and mountains (as well as rapids for one group of Long Gelat) (Figure 33). Perhaps due to relatively open waterways (although distant ones), intelligibility flow is maintained between the Long Gelat and the Kelinjau Modang.

Note how the Long Gelat live among the Busang, even living in the same longhouse apartments. The social hierarchy of the Long Gelat toward their Busang subjects probably accounts more for the persistent distinctions that exist between it and other Modang varieties as well as perhaps between Busang and Bahau.

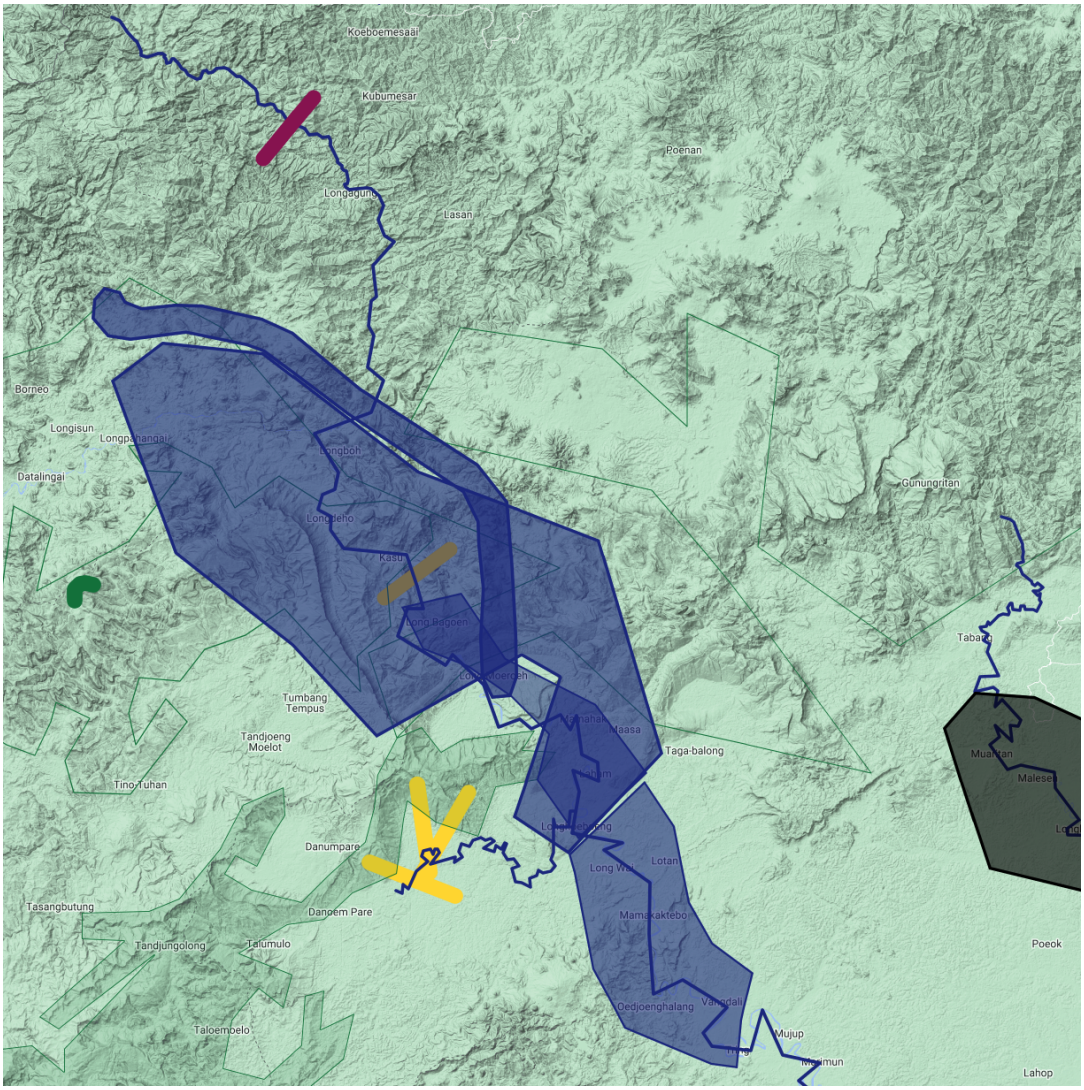


Figure 33. Map of Languages on the Mahakam River

The Mahakam river sees a great many overlaps among Kayanic-speaking groups. The Bahau and Busang are very close, but careful social distinctions are made along identity lines. The Mahakam Kayan are close neighbors with the Bahau and Busang. Despite this their language is very distinctive (45% for both respectively). Their language is probably so distinctive probably due to possible 'creolization' effects from having taken more slaves from very different language subfamilies, overwhelming the native Kayan speaking population (Guerreiro 1996).

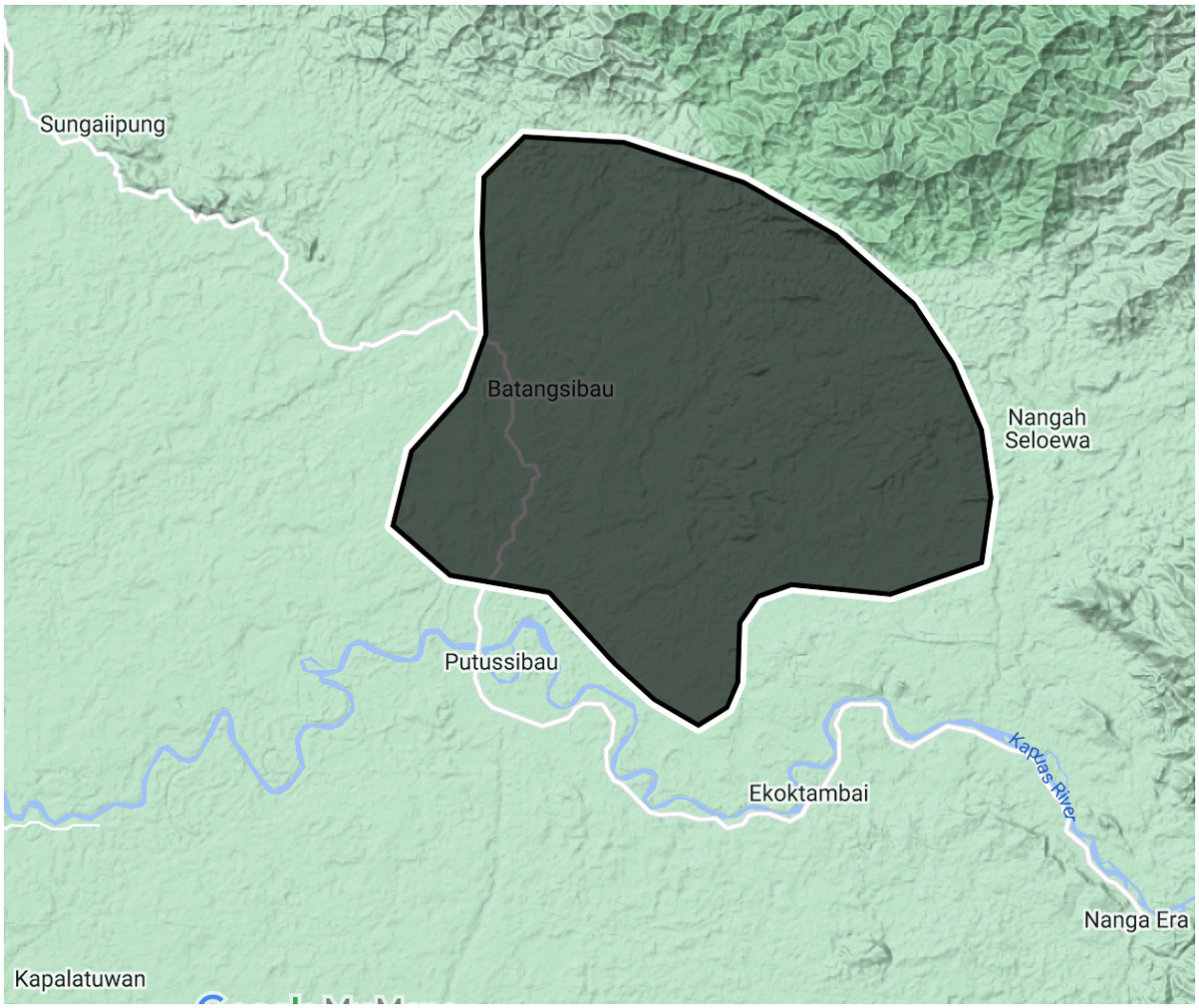


Figure 34. Map of Mendalam Kayan spoken on the Kapuas River (West Kalimantan)

There is no data for Mendalam Kayan, spoken on the Kapuas River on the western half of the island, near Putussibau (Figure 34).

The Kayanic languages have been color-coded below (Figure 36). Figures 37 and 38 provide closer views of the map in Figure 36. The Central Kayan Dialect Chain is in green. Bahau is in purple and Mahakam Kayan is in light purple. Ngorek ‘Proper’ or Murik is in dark purple. Hueng Bau-Pua’ is red. Merap is pink and Nyibun is light pink. Wehea is depicted in

dark orange. Kelinjau Modang-Long Gelat is in orange. Segai is light orange and bone-shaped. Punan Kelai is in light yellow-orange. The undescribed varieties are in black.

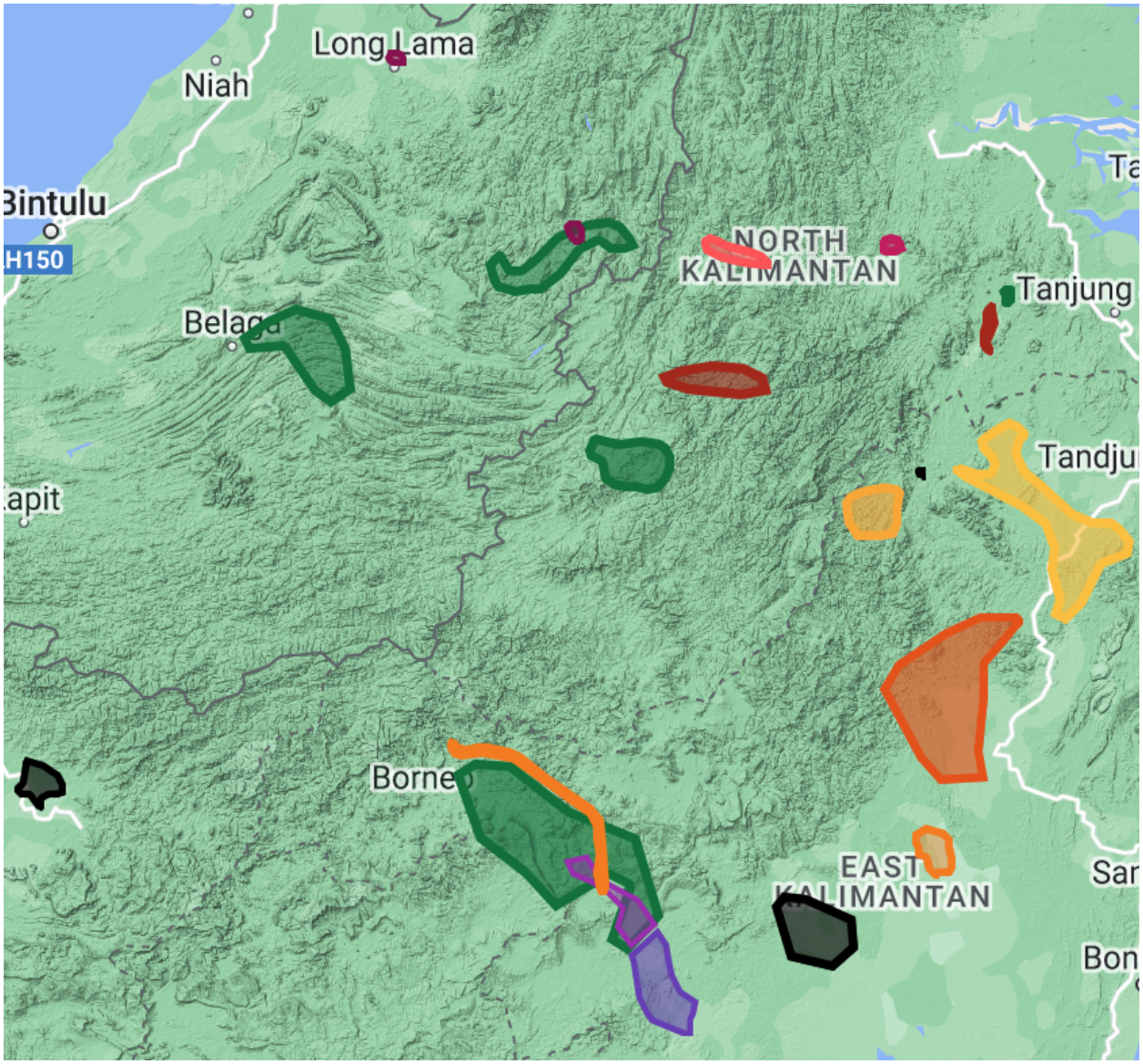


Figure 36. Overview Close-Up of Kayanic Dialectology

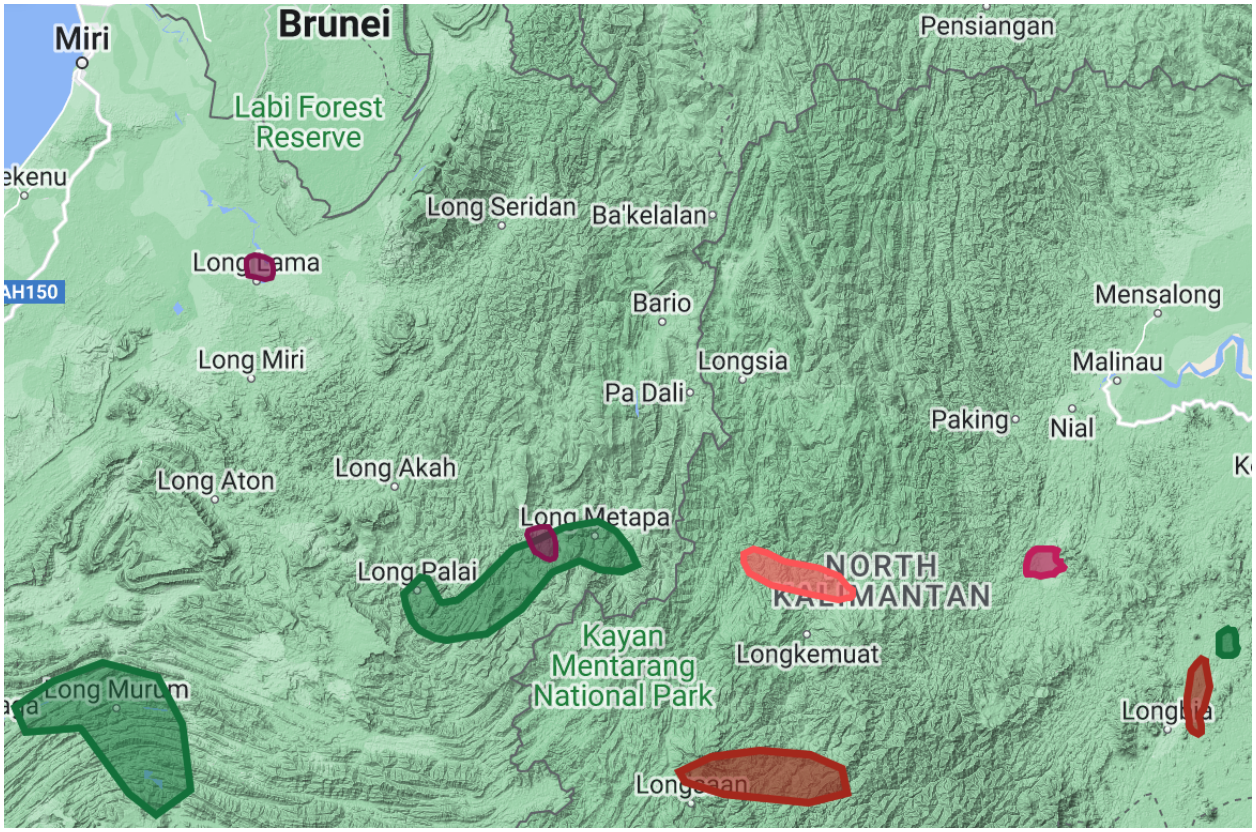


Figure 37. Overview Close-Up of Kayanic Languages of Malaysian Borneo and North Kalimantan

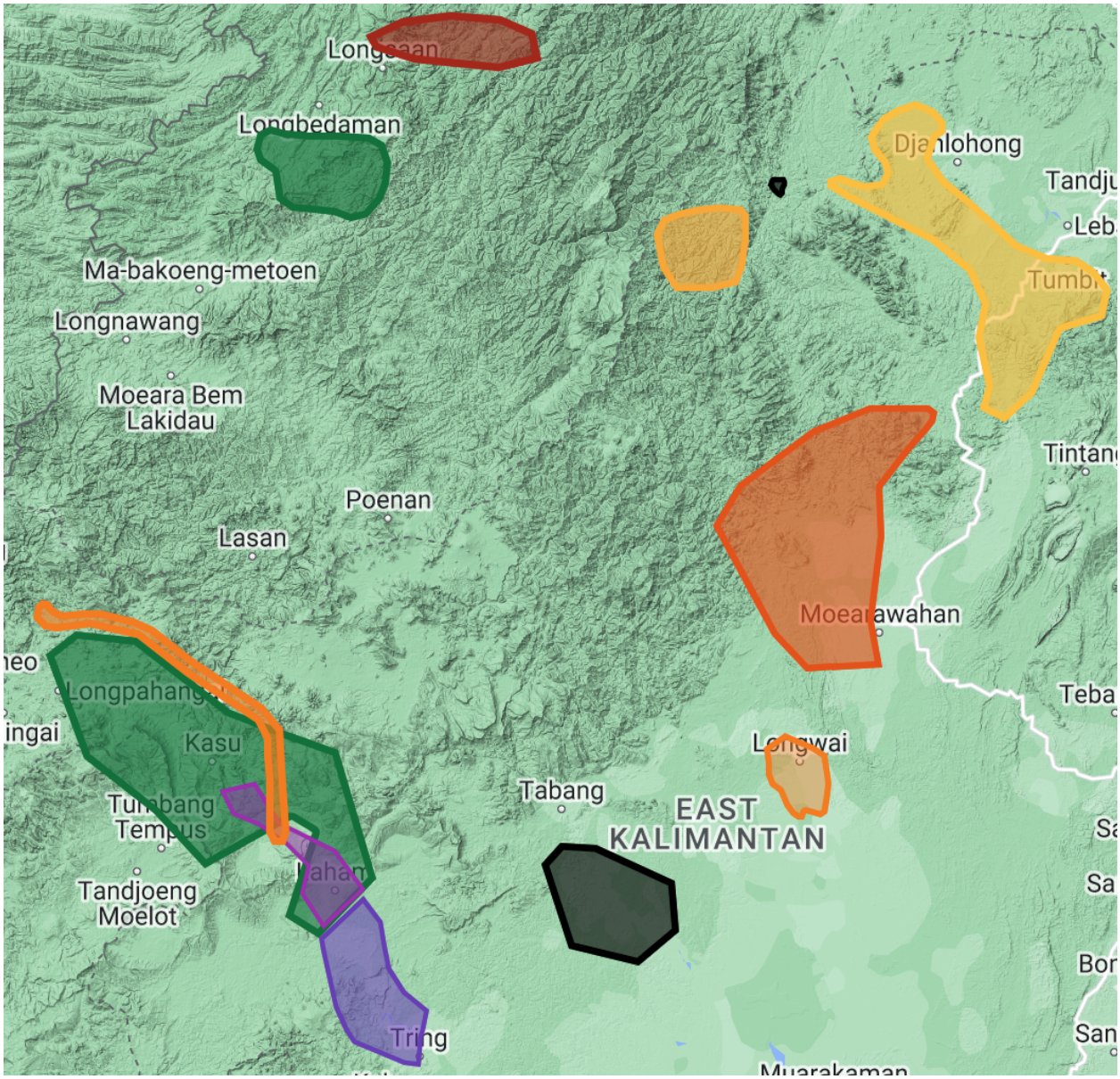


Figure 38. Overview Close-Up of Kayanic Languages of East Kalimantan

In addition to river, rapids, and mountains, sheer distance may play a role in language diversification. Physical niche factors appear to contribute to 'dialect-level' differentiation while social factors seem to take prominence over other factors in molding language diversification.

4.7 Summary: The organization of a language family

There are many different axes upon which language families can be organized. The most common is to speak of 'relatedness' in a phylogenetic sense. This is established by comparative

reconstruction. Since this study takes a *synchronic* orientation, here we speak of how the Kayanic language family is organized in terms of language varieties connected through calculated intelligibility. We have established that several dialect and language chains exist among the Kayanic language varieties. *Implicative coherence* was established among the different language varieties, forming a wave-like distribution of intelligibility within the family. In the second 'half' of our cartography of language diversification, we have established *geographical coherence* by plotting the distribution of languages and dialects spatially upon a physical map of Borneo.

What is clear is that social factors (strong in-group identity or ethnocentrism) counts more than physical niche factors, although they too play a role. In the allopatric speciation of these languages, the physical niche creates segregative conditions, powerful enough to yield dialect-level linguistic differentiation. These findings suggest that the power of the social niche plays an overwhelming role in linguistic differentiation, although, again, the physical niche can contribute too. The truly strong niche factor here is social—in-group identities and social hierarchies mold the praxis of language variation which in turn produces the differentiation patterns seen in our cartography of diversification.

CHAPTER 5

Conclusions: Progress and Prospects

> There are currently as many as 7,000 different languages spoken, or 7,000 mutually unintelligible systems of communication in one species, marking out at least 7,000 distinct societies. There are more different systems of communication in a single mammal species—for that is what we are—than there are mammal species. It is 7,000 different ways of saying, "Good morning," or, "Looks like rain today," and means that humans uniquely and strangely among animals often cannot communicate with other members of their own species.
>
> —Mark Pagel[22]

5.1 Introduction

Let us end where we began. *Why are there so many languages in the world? How did this diversity come about?* Part of the answer can be found in the *evolution of the language ecology*[23] in which speakers live and interact. This study on the Kayanic languages of Central Borneo offers evidence toward confirming the *Linguistic Niche Hypothesis*—asserting that languages adapt to their language ecologies. Languages, as complex adaptive systems, indeed, do evolve in ways that fit the specific environments they inhabit. Based on the data in this analysis, spatial factors such as geographic extensiveness, rivers, rapids, and mountains have an effect on predicted intelligibility. Even more so, social factors predominate as an organizing principle in the patterns of intelligibility. I conclude that the language ecology (a composite of physical, social, and linguistic niche factors, in this case) motivate the rise of linguistic complexity, not only grammatically as past studies have shown, but also lexically and phonologically. This lexical and phonological complexity inevitably has consequences for intelligibility—one of the principle measures by which linguists demarcate variants to be of

[22] This quote is from Mark Pagel's book *Wired for Culture: Origins of the Human Social Mind* (2012: viii).
[23] This is, of course, a play on the title of Mufwene's book *The ecology of language evolution*, which has greatly influenced the theory of this book in its population genetics approach to language diversification (2001).

'separate' languages, and therefore more linguistic diversity, and of a 'common' language, therefore less diversity. This means that linguistic diversification—the process of forming that diversity—is 'inextricably linked' (cf. Maffi 2005) to the language ecology.

5.2 Summary of findings: Evolutionary cartographies of language diversification

To remind us of our ultimate guiding inquiry, this study is concerned with whether the lexical composition of these Kayanic language varieties, as dictated by the phonological structure of their lexemes and whether they show adaptation to the language ecology. In other words, *does the phonological structure and lexical composition of the Kayanic languages, and, consequently, the patterns of dialectal intelligibility derived from it, support the Linguistic Niche Hypothesis—showing adaptation to ecological factors in this language subgroup's environment?* It appears to be so. Based on the lexicostatistical/phonostatistical analysis, three main Kayan 'languages' are predicted to exist: Bahau, Mahakam Kayan, and the Central Kayan Dialect Chain. The Central Kayan Dialect Chain is predicted to include the intelligible dialects of Baram Kayan, Rejang Kayan, Data Dian Kayan, Kayan River Kayan, and Busang. All of these are probably intelligible with each other. The Ngorek subgroup is composed of Ngorek proper, Merap, Nyibun, and the Hueng Bau-Pua' dialect chain. Modang is predicted to consist of Wehea, Segai, Punan Kelai, and the Kelinjau Modang-Long Gelat Modang dialect chain.

Once mapped, the *allopatric speciation* of these language varieties—created either by *segregative* or *integrative* conditions for their speakers—is based on physical niche factors of distance, rivers, rapids, and mountains, but more importantly of social niche factors. The social factors include in-group orientations such as ethnocentrism and social hierarchy which dictate the patterns of *diffusion* and *divergence* within these languages (especially the heterogeneity of the Modang group). Isolation orientation mixed with contact for other more powerful groups

likely plays an important role in the heterogeneity of the Ngorek subgroup. Their location in the more mountainous regions must play a role as well. In the linguistic niche of the Kayan, its role as a lingua franca and the openness of the Kayan people to language learning—despite their desire to preserve markers and distinctions of identity—likely play a role in their homogenization.

From these findings, it is proposed that language adaptation in this way, is a driving force in the creation of *linguistic diversity*, because *ecological niche factors*—in some cases slightly and in other cases more heavily—appear to bias patterns of intelligibility to diverge (or converge) with other language varieties—the main substance of *language diversification*.

5.3 Research contribution and recommendations

The contribution of this research centers around its theoretical and practical goals. This study provides new and up-to-date information on the linguistic situation of a severely under-described area of the Austronesian world. In addition to providing new lexical data and initial-stage documentation of several Kayanic languages—Mahakam Kayan (never before documented), Bahau, Wehea Modang, and Kelinjau Modang, this thesis has mapped the language varieties within the Kayanic subgroup and established language boundaries, based on predicted intelligibility.

By reaching these goals, certain practical objectives become possible, albeit preliminarily so. For starters, it is clear that the ISO 639-3 code situation among the Kayanic language varieties requires some revision. Bahau [bhv] and Busang [bfg] should probably maintain separate ISO codes, the methods here predict the merging of Baram Kayan [kys], Rejang [ree], and, perhaps, Kayan River Kayan [xkn]. Data Dian Kayan, labeled [ddk] here does not currently have an ISO code and perhaps belongs under this subsumed code. Mahakam Kayan [xay] is

almost certainly predicted to maintain its own code. In the Ngorek subgroup, only Ngorek or Murik [mxr] and Merap [puc] currently have ISO codes. Not only does this study predict that those ISO codes should be maintained, there should also probably be ISO codes for Nyibun (proposed code and abbreviation used here [nyi]), Hueng Bau (proposed code and abbreviation used here [hue]), and Pua' (proposed code and abbreviation used here [pua]). It is predicted that all Modang languages will require their own codes. There are currently only codes for Segai [sge] and general Modang [mxd]. Perhaps, there could be the following proposed codes: Wehea [weh], Kelinjau Modang [mxd], Long Gelat [lga], and Punan Kelai [puk]. Wehea should probably get a code closer to the preferred language name[24] while Kelinjau Modang could get the current code since it is closer to the ethnonym. Since there is no data for Mendalam Kayan [xkd], Punan Mahkam, or Long Belah Modang, no predictions can be made concerning their ISO code status.

This is the perfect time to emphasize that **no changes should occur to ISO codes based on these data alone**. This method only mades certain predictions to better inform hypotheses that require further testing with intelligibility/proficiency and perceptual dialectology methods. The sociolinguistic situation requires more depth and elucidation as well.

Despite this limitation, this study, furthermore, provides geo-spatial data on the maps as well as initial-level predictions for updates to language inventory maps (Ethnologue, Glottolog, Endangered Languages Project, UNESCO Atlas, etc.). Some suggestions concerning language and educational policy are provided below (Section 5.4).

The four major theoretical objectives, outlined in Chapter 1 were also addressed. The lexicostatistic/phonostatistic data presented here correlate with the genetic relationships,

[24] While the Wehea acknowledge that they are a 'Modang' group, they have a dispreference for the general name 'Modang' and prefer 'Wehea.'

established in the literature. This reifies the established integrity of the Kayanic family. This cartography of diversification also provides a description of the evolution of Kayanic, including potential roots of the theoretically problematic Modang subgroup. This study also demonstrates the potential for synchronic dialect mapping/atlases to empirically inform theories of language variation and language evolution by putting forth a case—that of Central Borneo—demonstrating the dynamics of language diversification.

Several recent studies, published during the writing of this thesis provides further support of the claims made here. Benz's et al. (2018) computational model testing neutral drift against environment factors, having some data from Borneo, found that the distribution and diversification of language families including Austronesian could not have occurred by neutral drift alone, suggesting an environmental impact on the diversification process. Smith & Rama (2022) uses Bayesian phylogenetic modeling to test for the temporal and cladistic structure of the language varieties of Borneo. Their findings support the a 'riverine homeland hypothesis' of language diversification on the island. These findings support the outcomes and accomplished objectives of this study.

Accomplishing these goals are significant not only for building a deeper understanding of the languages of Borneo (and, therefore, Austronesian languages and linguistics), but also for demonstrating the contribution of sociolinguistic survey work and field linguistics, more generally, to studies in the evolution of linguistic diversity.

5.4 Potential pathways: Linking research with education and development

While the contributions of this study to language development are modest—a first step on a much longer path to language testing and language program management—this study points to a few suggested future pathways for providing good mother tongue-based multilingual education

(MTB-MLE) in Indonesia. According to UNESCO, Indonesia scores the second worst in Southeast Asia—only slightly higher than Brunei, a country with much less linguistic diversity—for MTB-MLE programs.

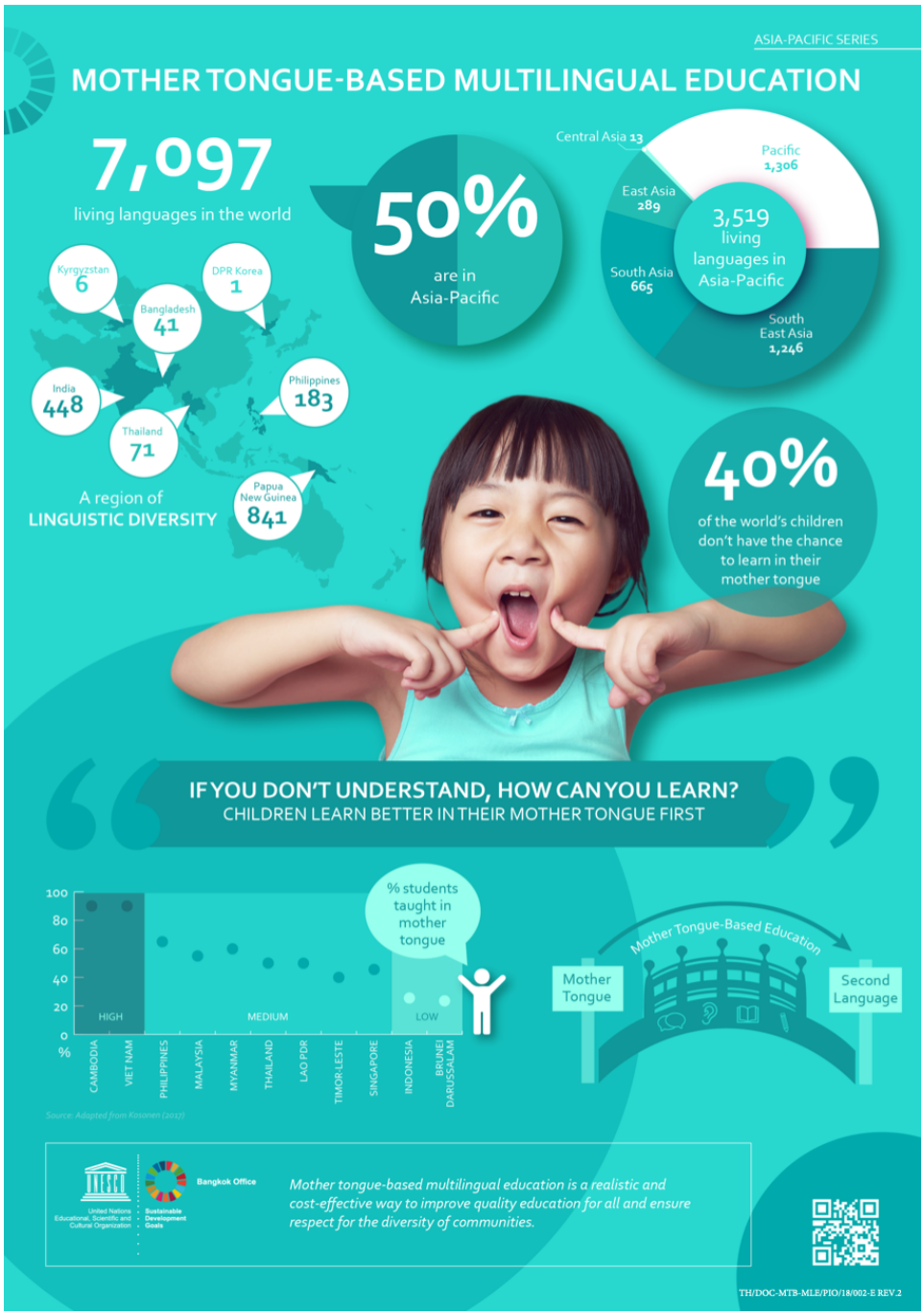


Figure 39. Unesco Mother Tongue Education Infographic[25]

[25] Used under Creative Commons license (CC-BY-SA 3.0 IGO): https://bangkok.unesco.org/content/mother-tongue-based-education-unesco-infographics-1.

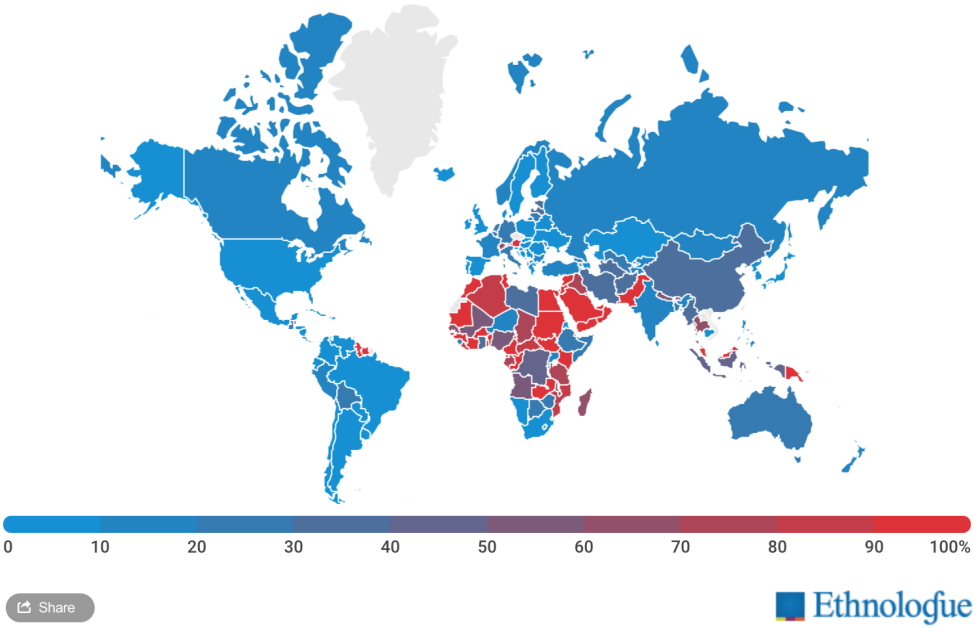


Figure 40. Ethnologue Literacy Map (https://bit.ly/3B18ShH)

According to the Ethnologue website (https://bit.ly/3B18ShH), "35% of the world's children begin their education in an unfamiliar language." In Indonesia, roughly 48% of the country's children attend school in a language other than their mother tongue (L1) which is less than the regional average (68% for East Asia, Southeast Asia, and the Pacific Islands).

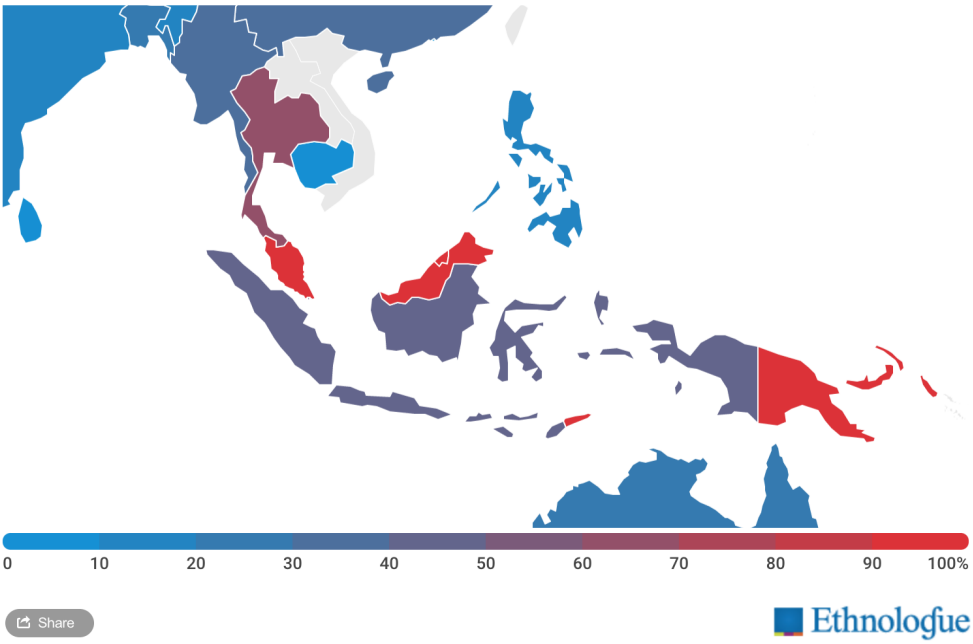


Figure 41. Ethnologue Literacy Map: Indonesia (https://bit.ly/3B18ShH)

In describing his time in Borneo, Gerry Abbott (2000: 216)[26], an international education development specialist, who has taught all around the developing world, describes the importance of literacy:

> After spending many years ... in some of the poorest countries ..., I became convinced that healthy development (as opposed to mal-development) was not possible in ethnic groups which no longer took pride in their own language; and that in the modern world the main prerequisite for such pride was literacy in those languages, so many of which were hitherto unwritten.

After all, as Nelson Mandela once said, "If you talk to a man in a language he understands, that goes to his head. If you talk to him in his language, that goes to his heart." Language and literacy is an important component of the UN Sustainability Goals (Sachs 2015).

The lexicostatistic/phonostatistic similarity percentages provided in table 5 point to clusters of language varieties to be tested further with intelligibility and/or proficiency testing and perceptual dialectology. Based on the data above all, Mahakam Kayan is predicted to require its own literacy materials as it bears extremely low similarity with other Kayan varieties. However, because it is purported that practically all Mahakam Kayan speakers are familiar with other more widespread varieties of Kayan, this fact will need to be considered in the language development decision making. Baram Kayan, Rejang Kayan, and Data Dian cluster together and should be tested for intelligibility/proficiency and perceptual dialectology across communities. The situation with the other varieties of Kayan (Bahau, Busang, Kayan River Kayan) is more inchoate with the very preliminary prediction that Bahau will require its own program.

For the Ngorek varieties (Ngorek proper, Hueng Bau-Pua', Merap, and Nyibun), it is preliminarily predicted that all varieties will require separate programs. With 83% between Hueng Bau and Pua', special attention should be given here to provide further intelligibility/ proficiency testing and perceptual dialectology research.

[26] In the quote above, the emphasis on 'literacy' is my own.

For the Modang varieties (Wehea, Kelinjau Modang-Long Gelat Modang, Segai, and Punan Kelai), it is preliminarily predicted that all varieties will require separate language and literacy programs. At 88% between Kelinjau Modang and Long Gelat, special attention should be given here to provide further intelligibility/proficiency testing and perceptual dialectology research. Again, due to a lack of data for Mendalam Kayan, Punan Mahkam, or Long Belah Modang, no predictions can be made here concerning language programs.

5.5 The limitations of this study and future research on the Kayanic languages of Central Borneo

Futures studies might investigate the effects of other niches (such as the genetic niche and physical (climatic) niche on the possibility of 'pre-nascent' or incipient tonogenesis, particularly among the Modang varieties and Mahakam Kayan (and, of course, other similar, phonologically aberrant languages in Borneo). No data here is provided for Mendalam Kayan (spoken in West Kalimantan off the Kapuas River), Long Belah Modang and Punan Mahkam which is a real limitation to a comprehensive treatment of Kayanic. There are sources on Mendalam Kayan but they are hard to gain access to and there is no data on Punan Mahkam or Long Belah Modang in the literature. They constitute major gaps in our knowledge of Kayanic languages.

Especially, given the aberrant sound changes in the Modang subgroup and the current inability to connect Kayanic to larger subgroups in Borneo and within Austronesian, having this data gap poses potential problems of the validity of our language classifications and could help solve these outstanding issues. The role of missing languages and dialects in linguistic classification, due to low documentation or endangerment is well documented (Evans 2010).

Some outstanding questions include whether Nyibun is genetically Kayanic or Kenyah. Evidence is provided here for its classification as Kayanic and therefore it was treated like so in

this study. However, its close proximity with Kenyah requires a set a criteria for making this demarcation.

The fact that lexicostatistics/phonostatistics was performed only within already established historical-comparative subgroups is another limitation of this study. The nature of the wave model suggests the possibility of linkages beyond the subgroup, but within the Kayanic family which simply were not addressed here. Future studies should investigate these linkages, perhaps through lexicostatistics/phonostatistics across subgroups or through glottometry across the entire family.

It is my hope the future of field-based evolutionary linguistics is bright and that, despite its limitations, the role of language survey as a main vehicle of investigating the evolution of linguistic diversity becomes a norm.

REFERENCES

Abbott, Gerry. 2000. Languages in lifelong education for third world development. *Development in Practice* 10(2). 216-222.

Adelaar, K. Alexander. 1995. Borneo as a cross-roads for comparative Austronesian linguistics. In Peter Bellwood, James J. Fox, & Darrell Tryon (eds.), *The Austronesians: Historical and comparative perspectives*, 75-95. Canberra: ANU E-Press.

Adelaar, K. Alexander & Nikolaus Himmelmann. 2005. *The Austronesian languages of Asia and Madagascar*. New York: Routledge.

Adelaar, Alexander K. 2010. Language documentation in the West Austronesian world and Vanuatu: An overview. In Margaret Florey (ed.), *Endangered Languages of Austronesia*, 12-44. Oxford: Oxford University Press.

Adelung, Johann Christoph & Johann Severin Vater. 1806. *Mithridates oder allgemeine Sprachenkunde – mit dem Vater Unser als Sprachprobe in bey nahe fünf hundert Sprachen und Mundarten*. Reprint: Hildesheim: Olms, 1970. Berlin: Voss.

Agheyisi, Rebecca & Joshua A. Fishman. 1970. Language attitude studies: A brief survey of methodological approaches. *Anthropological Linguistics* 12(5). 137-157.

Anderbeck, Karl. 2008. *Malay dialects of the Batanghari River Basin (Jambi, Sumatra)*. Dallas: SIL International.

Anderbeck, Karl Ronald. 2018. *Mapping the dialect network of western Bornean Malayic*. Bangi: Universiti Kebangsaan Malaysia dissertation.

Anderbeck, Karl & Herdian Aprilani. 2013. *The improbable language: Survey report on the Nasal language of Bengkulu, Sumatra*. Dallas: SIL Electronic Survey Report.

Anderbeck, Karl & Tessa Yuditha. 2020. *Mapping Bornean languages intelligibility* (Paper presented at the Indonesian Languages and Linguistics: State of the Field (InLaLi) Conference, Jakarta, 16-18 January 2020).

Anderson, Benedict. 1991. *Imagined communities: Reflections on the origin and spread of nationalism* (2nd edition). London: Verso.

Andersen, Henning. 1990. The structure of drift. In Henning Andersen & Konrad Koerner (Eds.), *Historical Linguistics* (Papers from the 8th International Conference on Historical Linguistics), 1-20. Amsterdam: Benjamins.

Anonby, Stan & David J. Holbrook. 2010. *A survey of the languages of the Javari River Valley, Brazil*. Dallas: SIL Electronic Survey Report.

Anonby, Stan & Dave Eberhard. 2016. A tale of two worlds: A comparative study of language ecologies in Asia and the Americas. *Language Documentation & Conservation* 10. 601–618.

Antunes, Nicolas, Wulf Schiefenhövel, Francesco d'Errico, William E. Banks & Marian Vanhaeren. Quantitative methods demonstrate that environment alone is an insufficient predictor of present-day language distributions in New Guinea. *PLoS ONE* 15(10). 1-27.

Aragon, Lorraine V. 2011. Living in Indonesia without a please or thanks: Cultural translations of reciprocity and respect. In Kathleen M. Adams and Kathleen A. Gillogly (eds.), *Everyday Life in Southeast Asia*, 16-26. Bloomington: Indiana University Press.

Aritonang, Buha, Non Martis, Hidayatul Astar, & Wati Kurniawati. 2002. *Kosakata dasar Swadesh di kabupaten Bulungan*. Jakarta: Pusat Bahasa.

Arizona, Yance, Muki Trenggono Wicaksono & Jacqueline Vel. 2019. The role of indigeneity NGOs in the legal recognition of Adat communities and customary forests in Indonesia. *The Asia Pacific Journal of Anthropology* 20(5). 487-506.

Aronoff, Mark, Irit Meir, Carol Padden & Wendy Sandler. 2008. The roots of linguistic organization in a new language. *Interaction Studies* (A Special Issue on Holophrasis Vs. Compositionality in the Emergence of Protolanguage) 9(1). 131-150.

Astar, Hidayatul, Buha Aritonang, Non Martis, & Wati Kurniawati. 2002. *Kosakata dasar Swadesh di kabupaten Kutai*. Jakarta: Pusat Bahasa.

Badia i Margarit, Antoni M. 1969. *La Llengua dels Barcelonins: Resultats d'una enquesta Sociològico-lingüística*. Barcelona: Edicions 62.

Bamba, John, Sujarni Alloy, Albertus, & Chatarina Pancer Istiyani (eds.). 2008. *Mozaik Dayak: Keberagaman subsuku dan bahasa Dayak*. Pontianak: Institut Dayakologi.

Barth, J. P. J. 1910. *Boesangsch-Nederlandsch Woordenboek*. Batavia: Government Printing Office.

Bastardas-Boada, Albert. 1996. *Ecologia de les llengües: Medi, contactes i dinàmica sociolingüística*. Barcelona: Proa.

Bastardas-Boada, Albert. 2002. *Biological and linguistic diversity: Transdisciplinary explorations for a socioecology of languages*. Diverscité langues. https://www.teluq.ca/diverscite/SecArtic/Arts/2002/bastarda/ftxt.htm

Bastardas-Boada, Albert. 2002. The ecological perspective: Benefits and risks for sociolinguistics and language policy and planning. In Alwin F. Fill, Hermine Penz & Wilhelm Trampe (eds.), *Colourful Green Ideas*, 77–88. Berna: Peter Lang.

Bastardas-Boada, Albert. 2014. Linguistic sustainability for a multilingual humanity. *Sustainable Multilingualism* 5. 134-163.

Bastardas-Boada, Albert. 2018. The ecology of language contact: Minority and majority languages. In Alwin F. Fill & Hermine Penz (eds.), *The Routledge Handbook of Ecolinguistics*, 26-39. New York: Routledge.

Baxter, G. J., R. A. Blythe, W. Croft & A. J. McKane. 2006. Utterance selection model of language change. *Physical Review* E73(4). 1-20.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59(1). 1-26.

Bedner, Adriaan & Yance Arizona. 2019. Adat in Indonesian land law: A promise for the future or a dead end? *The Asia Pacific Journal of Anthropology* 20(5). 416–434.

Bellwood, Peter. 1991. The Austronesian dispersal and the origin of languages. *Scientific American*. 265(1). 88-93.

Bender, Marvin L. 1971. The languages of Ethiopia: A new lexicostatistic classification and some problems of diffusion. *Anthropological Linguistics* 13(5). 165-288.

Benítez-Burraco, Antonio & Vera Kempe. 2018. The emergence of modern languages: Has human self-domestication optimized language transmission? *Frontiers in Psychology* 9(551). 1-8.

Bentley, R. Alexander & Stephen J. Shennan. 2003. Cultural transmission and stochastic network growth. *American Antiquity* 68(3). 459-485.

Bentz, Christian. 2018. *Adaptive languages: An information-theoretic account of linguistic diversity*. Berlin: De Gruyter Mouton.

Bentz, Christian, Dan Dediu, Annemarie Verkerk, & Gerhard Jäger. 2018. The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behavior* 2. 816–821.

Bentz, Christian & Bobo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics & Change* 3. 1–27.

Bergsland, Knut & Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3(2). 115-153.

Bernard, H. Russell. 2000. *Social research methods: Qualitative and quantitative approaches*. Thousand Oaks: Sage Publications, Inc.

Beyer, Robert, Joy S. Singarayer, Jay T. Stock & Andrea Manica. 2019. Environmental conditions do not predict diversification rates in the Bantu languages. *Heliyon* 5(10). 1-4.

Bickford, J. Albert. 1991. Lexical variation in Mexican Sign language. *Sign Language Studies* 72. 241-276.

Bickford, J. Albert. 2005. *The signed languages of Eastern Europe*. Dallas: SIL Electronic Survey Reports.

Bischoff, Shannon T. & Carmen Jany. 2006. *Insights from Practices in Community-Based Research: From theory to practice around the globe*. Boston: Walter de Gruyter.

Blair, Frank. 1990. *Survey on a shoestring: A manual for small-scale language surveys*. SIL/University of Texas at Arlington.

Blasi D. E., S. Moran, S. R. Moisik, P. Widmer, D. Dediu & B. Bickel. 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432). 1-10.

Blust, Robert. 1974a. A Murik vocabulary, with a note on the linguistic position of Murik. In Jérôme Rousseau (ed.), *The peoples of central Borneo* (Special issue of the Sarawak Museum Journal), 153-189. Kuching: Sarawak Museum.

Blust, Robert. 1974b. *The Proto-North Sarawak vowel deletion hypothesis*. Manoa, HI: University of Hawai‘i. dissertation.

Blust, Robert A. 1976. Austronesian culture history: Some linguistic inferences and their relations to the archaeological record. *World Archaeology* 8(1). 19-43.

Blust, Robert A. 1977. Sketches of the morphology and phonology of Bornean languages 1: Uma Juman (Kayan). In Hein Steinhauer (ed.), *Papers in Bornean and Western Austronesian languages*, no. 2, 7-122. Canberra: Pacific Linguistics.

Blust, Robert A. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36: 149-79.

Blust, Robert A. 1995. The prehistory of Austronesian-speaking peoples: A view from language. *Journal of World Prehistory* 9(4): 453-510.

Blust, Robert. 1998. The position of the languages of Sabah. In Maria Lourdes S. Bautista (ed.), *Pagtanaw: Essays on language in honor of Teodoro A. Llamzon*, 29–52. Manila: Linguistic Society of the Philippines.

Blust, Robert. 2000. Low vowel fronting in northern Sarawak. *Oceanic Linguistics* 39. 285-319.

Blust, Robert A. 2002. Formalism or phoneyism? The history of Kayan final glottal stop. In K. Alexander Adelaar and Robert Blust (eds.), *Between Worlds: Linguistic Papers in Memory of David John Prentice*, 29–37. Canberra: Pacific Linguistics.

Blust, Robert. 2007. Òma Lóngh historical phonology. *Oceanic Linguistics* 46(1). 1–53.

Blust, Robert. 2010. The Greater North Borneo hypothesis. *Oceanic Linguistics* 49(1). 44–118.

Blust, Robert. 2011. Austronesian: A Sleeping Giant? *Language and Linguistics Compass* 5(8). 538-550.

Blust, Robert. 2013. *The Austronesian languages* (2nd edition). Canberra: Asia-Pacific Linguistics.

Blust, Robert. 2014. Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics* 53(2). 300-391.

Blust, Robert A. 2014. *Blust Field Notebooks*. Archival Collection. Kaipuleohone: University of Hawai'i Digital Language Archive. https://scholarspace.manoa.hawaii.edu/handle/10125/33115

Blust, Robert. 2019. The Austronesian homeland and dispersal. *Annual Review of Linguistics* 5. 417-434.

Blust, Robert & Alexander D. Smith. 2014. *A bibliography of the languages of Borneo (and Madagascar)*. Phillips: Borneo Research Council, Inc.

Boberg, Charles, John Nerbonne, & Dominic Watt (eds.). 2018. *The handbook of dialectology*. Hoboken: John Wiley & Sons.

Bock, Carl. 1985 [1881]. *The headhunters of Borneo: A narrative of travel [up the Mahakkam and down the Barito]*. Oxford: Oxford University Press.

Boerger, Brenda H., Sarah Ruth Moeller, Will Reiman, & Stephen Self. 2018. *Language and culture documentation manual*. Leanpub.

Brooke Low, H. 1896. Kyan vocabulary. In H. Ling Roth (Ed.). *The Natives of Sarawak and British North Borneo*. London: Truslove & Hanson.

Brunn, Stanley D. & Roland Kehrein (eds.). 2020. *Handbook of the changing world language map*. Cham: Springer.

Bryant, David & Vincent Moulton. 2003. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*. 21(2). 255–265.

Bryman, Alan. 2008. *Social research methods* (3rd Edition). Oxford: Oxford University Press.

Burns, Robert. 1849. A vocabulary of the Kayan language of the Northwest of Borneo. *Journal of the Indian Archipelago and Eastern Asia* 3. 182-192.

Calvet, Louis-Jean. 1999. *Pour une écologie des langues du monde*. Paris: Plon.

Campbell, Lyle. 2013. *Historical linguistics: An introduction* (3rd Edition). Edinburgh: University of Edinburgh.

Casad, Eugene H. 1974. *Dialect intelligibility testing*. Norman: The Summer Institute of Linguistics and the University of Oklahoma.

Castro-Caldes, A., K. M. Petersson, A. Reis, S. Stone-Elander & M. Ingvar. 1998. The illiterate brain: Learning to read and write during childhood influences the functional organization of the adult brain. *Brain* 121. 1053–1063.

Catalogue of Endangered Languages. 2021. University of Hawaii at Manoa. http://www.endangeredlanguages.com

Cavalli-Sforza, Luigi Luca. 1997. Genes, peoples, and languages. *PNAS* 94. 7719-7724.

Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi & Joanna Mountain. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *PNAS* 85(16). 6002-6006.

Cense, A. A., and E. M. Uhlenbeck. 1958. *Critical survey of studies on the languages of Borneo* (Koninklijk Instituut Voor Taal-, Land- En Volkenkunde, Bibliographical Series). The Hague: Martinus Nijhoff.

Chamberlain, Alexander Francis. 1900. *The child: A study in the evolution of man*. London: Walter Scott, Ltd.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology* (2nd edition). Cambridge: Cambridge University Press.

Charpentier, Jean-Michel & Alexandre François. 2015. *Linguistic atlas of French Polynesia / Atlas linguistique de la Polynésie française*. Berlin: De Gruyter Mouton.

Chen, Sibo. 2016. Language and ecology: A content analysis of ecolinguistics as an emerging research field. *Ampersand* 3: 108-116.

Chomsky, Noam. 2005. Three factors in language design. *Linguistics Inquiry* 36(1). 1-22.

Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral & Brain Sciences* 31(5). 489-508.

Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences* 39. 1-17.

Clifton, John M. 2002. *Studies in languages of Azerbaijan*. Baku: Academy of Sciences of Azerbaijan/SIL International.

Clifton, John M. 2005. *Studies in languages of Tajikistan*. Dushanbe: National State University of Tajikistan/SIL International.

Clynes, Adrian. 2005. Belait. In Alexander Adelaar and Nikolaus P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*, 429-455. New York: Routledge.

Clayre, Beatrice & L. Cubit. 1974. An outline of Kayan grammar. *Sarawak Museum Journal* 43. 43-91.

Clayre, Iain F.C.S. 1972. Punan Ba': Melanau link with the Ulu Rejang. *Sarawak Gazette* XCVIII. 23-25.

Cleary, M.C. & F.J. Lian. 1991. On the geography of Borneo. *Progress in Human Geography* 15(2). 163-177.

Comrie, Bernard. (ed). 1987. *The world's major languages*. New York: Oxford University Press.

Connolly, Jennifer. 2003. *Becoming Christian and Dayak: A study of Christian conversion among Dayaks in East Kalimantan, Indonesia*. New York: New School University dissertation.

Cooke, Fadzilah Majid. (eds). 2006. *State, communities and forests in contemporary Borneo*. Canberra: ANU E Press.

Coomans, Mikail. 1987. *Manusia Daya: Dahulu, sekarang, masa depan*. Jakarta: PT Gramedia.

Connolly, Jennifer. 2003. *Becoming christian and Dayak: A study of christian conversion among Dayaks in East Kalimantan, Indonesia*. New York: New School University dissertation.

Cooper, Robert L. 1980. Sociolinguistic surveys: The state of the art. *Applied Linguistics* 1(2). 113-128.

Coseriu, Eugenio. 1981. "Los conceptos de 'dialecto', 'nivel' y 'estilo de lengua' y el sentido propio de la dialectología." *Lingüística española actual* 3. 1-32.

Crawfurd, John. 1853. *A grammar and dictionary of the Malay language with a preliminary dissertation (Volume I: Dissertation and grammar)*. London: Smith, Elder, and Co.

Crevello, Stacy. 2004. Dayak Land Use Systems and Indigenous Knowledge. *Journal of Human Ecology* 16. 69-73.

Croft, Willaim. A. 2000. *Explaining language change*. Harlow: Longman Linguistic Library.

Crowley, Terry & Claire Bowern. 2010. *An introduction to historical linguistics* (4th Edition). Oxford: Oxford University Press.

Cubit, L. E. 1964. Kayan phonemics. *Bijdragen tot de Taal-, Land- en Volkenkunde* 120(4). 409-423.

Dąbrowska, Ewa. 2015. Language in the mind and in the community. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms - New paradoxes: Recontextualizing language and linguistics*. Berlin: De Gruyter Mouton.

Dale, Rick & Gary Lupyan. 2012. Understanding the origins of morphological diversity: The Linguistics Niche Hypothesis. *Advances in Complex Systems* 15(03n04). 1-16.

Dalby, David. 2000. *The Linguasphere register of the World's languages and speech communities*. Hebron: Linguasphere Press. http://www.linguasphere.info/jr/index.php?l1=maps

Darwin, Charles R. 2003 [1859]. *The origin of species*. New York: Signet Classics.

Darwin, Charles R. & Wallace, Alfred. R. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London: Zoology*, 3(9). 45-62.

De Beer, J. H. & M. J. McDermott. 1996. *The economic value of non-timber forest products in Southeast Asia* (2nd edition). Amsterdam: IUCN.

Decker, Ken & John Grummitt. 2017. *Understanding language choices: A guide to sociolinguistic assessment*. Dallas: SIL International.

Dediu, Dan. 2011. Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? *Human Biology* 83(2). 279-296.

Dediu, Dan. 2015. *An introduction to genetics for language scientists*. Cambridge: Cambridge University Press.

Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *PNAS* 104(26). 10944-10949.

Dediu, Dan, Rick Janssen & Scott R. Moisik. 2017. Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Language & Communication* 54. 9-20.

Departemen Pendidikan Dan Kebudayaan. 1982. *Permainan rakyat daerah Kalimantan Timur* (Proyek inventarisasi dan dokumentasi kebudayaan daerah). Jakarta: Departemen Pendidikan Dan Kebudayaan.

Dixon, R. M. W. 2010. *The languages of Australia*. Cambridge: Cambridge University Press.

Douglas, R. S. 1911. A comparative vocabulary of the Kayan, Kenyah and Kelabit dialects. *Sarawak Museum Journal* 1(1). 75–119.

Douglas, R. S. 1912. Some Murik words. *Sarawak Museum Journal* 1. 86-87.

Drude, Sebastian. 2018. Reflections on diversity linguistics: Language inventories and atlases. In Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton (Eds.), *Reflections on Language Documentation 20 Years after Himmelmann 1998* (Language Documentation & Conservation Special Publication no. 15), 122-131. Honolulu: University of Hawai‘i Press.

Dyen, Isidore. 1965. A lexicostatistical classification of the Austronesian languages (Indiana University Publications in Anthropology and Linguistics, International Journal of American Linguistics). Baltimore: The Waverly Press.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (Eds.). 2022. *Ethnologue: Languages of the world* (25th edition). Dallas: SIL International. Online version: http://www.ethnologue.com

Effendy, Chairil. 1989. *Morfologi Bahasa Kayan*. Pontianak: Universitas Tanjungpura (Departemen Pendidikan dan Kebudayaan).

Effendy, Chairil, A. R. Muzammil, Firman Susilo, Dedy Ari Aspar, Agus Syahrani, Hangga Dwitika & Ediyanto. 2006. *Kamus Bahasa Kayan–Bahasa Indonesia*. Pondianak: STAIN Pontianak Press.

Eghenter, Cristina. 1999. Migrants' Practical Reasonings: The Social, Political, and Environmental Determinants of Long-Distance Migrations among the Kayan and Kenyah of the Interior of Borneo. *Journal of Social Issues in Southeast Asia* 14(1):1-33.

Eghenter, Cristina. 2000. What is Tana Ulen good for? Considerations on indigenous forest management, conservation, and research in the interior of Indonesian Borneo. *Human Ecology* 28(3). 331-357.

Eghenter, Cristina. 2001. Towards a causal history of a trade scenario in the interior of East Kalimantan, Indonesia, 1900-1999. *Bijdragen tot de Taal-, Land- en Volkenkunde* 157(4). 739-769.

Eghenter, Cristina & Bernard Sellato (eds.). 1999. *Kebudayaan dan pelestarian alam: Penelitian interdisipliner di pedalaman Kalimantan*. Jakarta: WWF.

Eghenter, Cristina, Bernard Sellato, & G. Simon Devung (eds.). 2003. *Social science research and conservation management in the interior of Borneo: Unravelling past and present interactions of people and forests*. Bogor: CIFOR.

Elliott, L. L. 1962. Backward and forward masking of probe tones of different frequencies. *Journal of the Acoustical Society of America* 34. 1116-1117.

Ellis, S. James. 2007. Language bending in Micronesia. *Working Papers in Linguistics: University of Hawai'i Mānoa* 38(4). 1-17.

Emerson, Robert M., Rachel I. Fretz & Linda L. Shaw. 2011. *Writing Ethnographic Fieldnotes* (Second Edition). Chicago: University of Chicago Press.

Espree-Conaway, DeAndré. 2012. *Language attitudes, acquisition, and usage of Osob Kiwalan Ngalam: An Indo-Javanese language of Malang.* Biehl International Research Scholarship. Sewanee: The University of the South, 1-22. https://dspace.sewanee.edu/bitstream/handle/11005/2127/EspreeConwayDeAndreLanguageCLD2012.pdf?sequence=1

Espree-Conaway, DeAndré. 2013. Bahasa walikan malangan and the building of Indo-Javanese Urban Spaces. *Planum: The Journal of Urbanism* 27(2), 1-6.

Espree-Conaway, DeAndré. 2014. *Fieldnotes from Watson fellowship Asia-Pacific research, 2013-2014*. Unpublished ms.

Espree-Conaway, DeAndré A. 2020. *Fieldnotes from linguistics research in Indonesia, 2019-2020*. Unpublished ms.

Ethnologue Website [*How many children lack access to education in their first language?*] https://www.ethnologue.com/guides/languages-of-instruction

Evans, Nicholas. 2010. *Dying words: Endangered languages and what they have to tell us*. Malden: Wiley-Blackwell.

Everett, Caleb, Damián E. Blasí & Seán G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *PNAS* 112(5). 1322-1327.

Everett, Caleb, Damián E. Blasí & Seán G. Roberts. 2016. Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution* 1(1). 33-46.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46(4). 621-646.

Everett, Daniel L. & Miguel Oliveira Jr. 2010. *The perspicuity of the syntax-prosody interface*. Keynote talk at the workshop on intonation at the meeting of the Deutsche Gesellschaft für Sprachwissenschaft; 2010. Available from: https://daneverettbooks.com/wp-content/uploads/2014/04/%20intonationworkshopberlin.pdf

Evers, Hans-Dieter & Solvay Gerke. 1992. The culture of planning: Transmigration area development in East Kalimantan, Indonesia. *International Sociology* 7(2). 141-151.

Fall, Moustapha. 2020. *The impact of mother tongue illiteracy on second language acquisition: The case of French and Wolof in Senegal*. New York: Routledge.

Favier, Saoradh, Antje S. Meyer & Falk Huettig. 2021. Literacy can enhance syntactic prediction in spoken language processing. *Journal of Experimental Psychology: General*. Advance online publication. 1–8.

Fill, Alwin F. 1998. Ecolinguistics — State of the Art 1998. *AAA: Arbeiten aus Anglistik und Amerikanistik* 23(1). 3-16.

Fill, Alwin F. & Hermine Penz (eds.). 2018. *The Routledge Handbook of Ecolinguistics*. New York: Routledge.

Fill, Alwin F. & Peter Mühlhäusler (eds.). 2001. *The ecolinguistics reader: Language, ecology and environment*. London: Continuum.

Fisher, Simon E. & Sonja C. Vernes. 2015. Genetics and the language sciences. *Annual Review of Linguistics* 1(1). 289-310.

Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The Evolution of the Language Faculty: Clarifications and Implications. *Cognition* 97(2). 179-210.

Floyd, Simeon, Lila San Roque & Asifa Majid. 2018. Smell is coded in grammar and frequent in discourse: Cha'palaa olfactory language in cross-linguistic perspective. *Journal of Linguistic Anthropology* 28(2). 175-196.

Flydal, Leiv. 1951. "Remarques sur certains rapports entre le style et l'état de langue." *Norsk Tidsskrift for Sprogvidenskap* 16. 240-257.

Frank, Michael C., Daniel L. Everett, Evelina Fedorenko & Edward Gibson. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108(3). 819-824.

Freeman, Scott & Jon C. Herron. 2007. *Evolutionary analysis* (4th edition). New York: Pearson.

Futrell, Richard, Laura Stearns, Daniel L. Everett, Steven T. Piantadosi & Edward Gibson. 2016. A corpus investigation of syntactic embedding in Pirahã. *PLoS ONE* 11(3): 1-20.

Gal, Susan & Judith T Irvine. 1995. The Boundaries of Languages and Disciplines: How Ideologies Difference. *Social Research* 62(4). 967–1001.

Gawne, Lauren & Hiram Ring. 2016. Mapmaking for Language Documentation and Description. *Language Documentation & Conservation* 10. 188-242.

Gerke, Solvay. 1997. Ethnic relations and cultural dynamics in East Kalimantan: The case of the Dayak lady. *Indonesia and the Malay World* 25(72). 176-187.

Gesner, Conrad. 1555. *Mithridates: De differentiis linguarum tum veterum tum quae hodie apud diversas nationes in toto orbe terrarum in usu sunt observationes*. Neudruck: Aalen: Scientia, 1974, hrsg. u. eingel. v. Manfred Peters. Zürich: Froschoverus.

Ginsburgh, Victor & Shlomo Weber. 2011. *How many languages do we need? The economics of linguistic diversity*. Princeton: Princeton University Press.

Gooskens, Charlotte & Cindy Schneider. 2016. Testing mutual intelligibility between closely related languages in an oral society. *Language Documentation & Conservation* 10. 278–30.

Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306(5695). 496–499.

Graham, Sarah A. & Simon E. Fisher. 2015. Understanding language from a genomic perspective. *Annual Review of Genetics* 49. 131-60.

Greenhill, Simon J. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37(4). 689-698.

Greenhill, S.J., Blust. R, & Gray, R.D. (2008). The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics. *Evolutionary Bioinformatics* 4. 271-283.

Grierson, George A. 2005 [1928]. *Linguistic survey of India*. Calcutta: Government Press.

Griffiths, Sascha Sebastian & Laura Robson. 2010. Cultural ecologies of endangered languages: The cases of Wawa and Njanga. *Anthropological Linguistics* 52(2). 217-238.

Grimes, Joseph E. 1995. *Language survey reference guide*. Dallas: Summer Institute of Linguistics.

Grimes, Joseph E. 2001. The value of comparative linguistics. *Notes on Linguistics* 4(1–4). 87–94.

Grimes, Joseph E. & Frederick B. Agard. 1959. Linguistic divergence in Romance. *Language* 35(4). 598-604.

Großmann, Kristina. 2017. The (ir)relevance of ethnicity among the Punan Murung and Bakumpai in Central Kalimantan. In Cathrin Arenz, Michaela Haug, Stefan Seitz & Oliver Venz (eds.), *Continuity under change in Dayak societies*, 141-162. Wiesbaden: Springer VS.

Guerreiro, Antonio J. 1983. A note on pronouns in the Long Gelat and Busang languages (Upper Mahakam, Kalimantan Timur). *Borneo Research Bulletin* 15(2). 98-105.

Guerreiro, Antonio. 1985. An ethnological survey of the Kelai river area, Kabupaten Berau, East Kalimantan. *Borneo Research Bulletin* 17. 106-120.

Guerreiro, Antonio J. 1991. Le mythe de Long Déang Yung: Récit modang wehès de l'origine du riz et des augures (Kalimantan, Indonésie). *La Mandragore: Revue des littératures orales* 4. 143-163.

Guerreiro, Antonio J. 1993. *Modang*. In Paul Hockings (Ed.), East and Southeast Asia, 185-187. Boston: G. K. Hall.

Guerreiro, Antonio. 1996. Homophony, sound changes and dialectal variation in some central Bornean languages. *Mon-Khmer Studies* 25. 205–226.

Guerreiro, Antonio J. 2002. A note on 'spirit helpers' in the Lung Lejie epic of the Wehèa Modang (East Kalimantan). In Alexander K. Adelaar & Robert Blust (eds.), *Between Worlds: Linguistic Papers in memory of David John Prentice*, 95-110. Canberra: ANU Press.

Haas, Walter. 2010. A study of areal diffusion. In Peter Auer, Alfred Lameli & Walter Haas (eds.), *Language and space: Theories and methods: An international handbook of linguistic variation*, vol. 1, 649–668. Berlin: De Gruyter Mouton.

Harrisson, Tom. 1962. Borneo Death. *Bijdragen tot de taal-, land- en volkenkunde* 118(1). 1-41.

Hartmann, Stefan & Michael Pleyer. 2020. A new hope (for common ground)? Convergences between biolinguistic and usage-based approaches to language and its evolution. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd & T. Verhoef (eds.), *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13)*, 141-143. Nijmegen: Max Planck Institute for Psycholinguistics.

Haugen, Einar. 2001 [1970]. The ecology of language. In Alwin Fill & Peter Mühlhäusler (eds.), *The Ecolinguistics Reader: Language, Ecology and Environment*, 57-66. New York: Continuum.

Heeringa, W. & J. Nerbonne. 2001. Dialect areas and dialect continua. *Language Variation and Change* 8. 375-400.

Hernandez, Arturo E., Maya R. Greene, Kelly A. Vaughn, David J. Francis, Elena L. Grigorenko. 2015. Beyond the bilingual advantage: The potential role of genes and environment on the development of cognitive control. *Journal of Neurolinguistics* 35. 109-119.

Hervás y Panduro, Lorenzo. 1800. *Catalogo de las lenguas de las naciones conocidas, y numeracion, division y clase de estas segun la diversitad de sus idiomas y dialectos*. re-published as facsilime by the Biblioteca Nacional (España), Madrid, 2008.6 vols. Madrid: [s.ed.].

Hills, Thomas T., James S. Adelman & Takao Noguchi. 2017. Attention economies, information crowding, and language change. In *Big data in cognitive science*, 270–293. New York: Routledge.

Holmsen, Katherine. 2006. *Out of the forest and into the market: Social and economic transformations in a Bornean foraging society*. Tucson: University of Arizona dissertation.

Honkola, Terhi, Kalle Ruokolainen, Kaj J. J. Syrjänen, Unni-Päivä Leino, Ilpo Tammi, Niklas Wahlberg & Outi Vesakoski. 2018. Evolution within a language: Environmental differences contribute to divergence of dialect groups. *BMC Evolutionary Biology* 18(132). 1-15.

Hornberger, Nancy H., & Francis M. Hult. 2008. Ecological language education policy. In Bernard Spolsky & Francis M. Hult (Eds.), *Handbook of Educational Linguistics*, 280-296. Malden: Blackwell.

Hruschka, Daniel J., Simon Branford, Eric D. Smith, Jon Wilkins, Andrew Meade, Mark Pagel, & Tanmoy Bhattacharya. 2015. Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology* 25(1). 1-9.

Hua, Xia, Simon J. Greenhill, Marcel Cardillo, Hilde Schneemann, and Lindell Bromham. 2019. The ecological drivers of variation in global language diversity. *Nature Communications* 10. 1-10.

Hudson, Alfred B. 1967. *The Barito isolects of Borneo: A classification based on comparative reconstruction and lexicostatistics* (Southeast Asia Programs, Data Paper no. 68.). Ithaca: Cornell University Press.

Hudson, Alfred B. 1978. Linguistic relations among Bornean peoples with special reference to Sarawak: An interim report. *Studies in Third World Societies* 3. 1–44.

Hughes-Hallett, H. 1938. An account of a Berhantu ceremony called "Perakong" by the Orang Belait of Brunei. *Journal of the Malayan Branch of the Royal Asiatic Society* 16(1). 102-108.

Hult, Francis M. 2009. Language ecology and linguistic landscape analysis. In E. Shohamy & D. Gorter (Eds.), *Linguistic Landscape: Expanding the Scenery*, 88-104. London: Routledge.

Hult, Francis M. 2010. Analysis of language policy discourses across the scales of space and time. *International Journal of the Sociology of Language* 202. 7-24.

Hult, Francis M. 2012 . Ecology and multilingual education. In Carol A. Chapelle (Ed.), *Encyclopedia of Applied Linguistics, Vol. 3*, 1835-1840. Malden, MA: Wiley-Blackwell.

Humboldt, Wilhelm von. 1999 [1836]. *'On Language' On the diversity of human language construction and its influence on the mental development of the human species*. Cambridge: Cambridge University Press.

Inoue, Makoto, Lugan, & Igin Bilung. 1991. Changes in economic life of the hunter and gatherers: The Kelay Punan in East Kalimantan. *Tropics*. 1(2/3). 143-153.

Jackson, Frederick Henry. 1983. *Internal and external relationships of the Trukic languages of Micronesia*. Manoa: University of Hawaii dissertation.

Jalong, Philip Ngau. 1989. The Ngurek. *Sarawak Museum Journal* 40. 157-168.

Journal of Language Survey Reports (formally SIL Electronic Survey Reports). https://www.sil.org/resources/search/series/journal-language-survey-reports

Kaboy, Tuton. 1974. The Penan Aput. *Sarawak Museum Journal* 22. 287-293.

Kalmár, Ivan. 1985. Are there really no primitive languages? In D. O. Olson, N. Torrance & A. Hildyard (eds.), *Literacy, Language, and Learning*, 148-166. New York: Cambridge University Press.

Kaltofen, R.E. 1917. *Borneo: Volkenkaart van Borneo*. Koninklijk Instituut voor de Tropen (KIT). http://hdl.handle.net/1887.1/item:2011882 (23 August 2020).

Karan, Mark E. 1996. *The dynamics of language spread: A study of the motivations and the social determinants of the spread of Sango in the Republic of Central Africa*. Philadelphia: University of Pennsylvania dissertation.

Karan, Mark E. 2001. *The dynamics of Sango language spread*. Dallas: SIL International.

Karjus, Andres, Richard A. Blythe, Simon Kirby & Kenny Smith. 2020. Quantifying the dynamics of topical fluctuations in language. *Language Dynamics & Change*. 1–40.

Kaskija, Lars. 2002. *Claiming the forest: Punan local histories and recent developments in Bulungan, East Kalimantan*. Bogor: CIFOR.

Keller, Rudi. 1994. *On language change: The invisible hand in language*. New York: Routledge.

Kempe, Vera & Patricia J. Brooks. 2018. Linking adult second language learning and diachronic change: A cautionary note. *Frontiers in Psychology* 9(480). 1-5.

Kennedy, Raymond. 1935. *The ethnology of the greater sunda islands*. New Haven, CT: Yale University dissertation.

Keppel, Henry. 1846. *The expedition to Borneo of H.M.S. Dido for the suppression of piracy* (Volume 1). London: Chapman and Hall.

Kimura, Motoo. 1985. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

King, Victor T. 1993. *The peoples of Borneo*. Oxford: Blackwell.

Koplenig, Alexander. 2019. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science* 6. 1-11.

Kroeger, Paul R. 1998. Language Classification in Sarawak: A Status Report. *Sarawak Museum Journal* LIII(74).137-173.

Kurniawati, Wati, Non Martis, Buha Aritonang & Hidayatul Astar. 2002. *Kosakata dasar Swadesh di kabupaten Berau, kotamadya Samarinda dan kotamadya Balikpapan*. Jakarta: Pusat Bahasa.

Jackendoff, R. 1978. Grammar as evidence for conceptual structure. In M. Halle, J. Bresnan & G.A. Miller (eds.). *Linguistic Theory and Psychological Reality*. Cambridge: MIT Press.

Jalong, Philip Ngau. 1989. The Ngurek. *Sarawak Museum Journal* 40. 157-168.

Jespersen, Otto. 1922. *Language, its nature, development, and origin*. H. Holt.

Labov, William. 1966. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Ladefoged, Peter, Ruth Glick & Clive Criper. 1971. *Language of Uganda*. Nairobi: Oxford University Press.

Lai, Cecilia S. L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem & Anthony P. Monaco. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413. 519-523.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Lameli, Alfred. 2010. Linguistic atlases - traditional and modern. In Peter Auer and Jürgen Erich Schmidt (eds.), *Language and space: An international handbook of linguistic variation. Theories and methods*, vol. 1, 567-592. Berlin: De Gruyter Mouton.

Lebar, Frank. 1972. *Ethnic Groups of Insular Southeast Asia* (Volume 1: Indonesia, Andaman Islands, and Madagascar). New Haven: Human Relations Area Files Press.

Lee, S. & T. Hasegawa. 2014. Oceanic barriers promote language diversification in the Japanese Islands. *Journal of Evolutionary Biology* 27(9). 1905-1912.

Leinonen, Therese, Çağrı Çöltekin & John Nerbonne. 2016. Using Gabmap. *Lingua* 178. 71-83.

Lewis, Molly & Michael C. Frank. 2016. Linguistic niches emerge from pressures at multiple timescales. In Anna Papafragou, Daniel Grodner, Daniel Mirman, & John C. Trueswell (Eds.), (*Proceedings of the 38th Annual Meeting of the Cognitive Science Society, CogSci 2016*, Philadelphia, 10-13 August 2016.)

Lewontin, Richard C. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1. 1-18.

Lieberson, Stanley. 1980. Procedures for improving sociolinguistic surveys of language maintenance and language shift. *International Journal of the Sociology of Language* 25. 11-27.

Ludwig, Ralph, Peter Mühlhäusler, & Steve Pagel (eds.). 2019. *Linguistic ecology and language contact.* Cambridge: Cambridge University press.

Lupyan, Gary and Rick Dale. 2010. Language Structure Is Partly Determined by Social Structure. *PLoS ONE* 5(1). 1-10.

Lupyan, Gary and Rick Dale. 2015. The role of adaptation in understanding linguistic diversity. In Rik De Busser & Randy J. LaPolla (eds.), *Language Structure and Environment: Social, Cultural, and Natural Factors*, 289–316. Amsterdam: John Benjamins Publishing.

Lupyan, Gary and Rick Dale. 2016. Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences* 20(9). P649-660.

MacKinnon, K., H. Gusti, H. Halim, & A. Mangalik. 1996. *The ecology of Kalimantan* (The Ecology of Indonesia Series, Volume 3). Singapore: Periplus Editions.

MacNamara, John. 1966. *Bilingualism and primary education*. Edinburgh: Edinburgh University Press.

Maddieson, Ian. 2018. Language adapts to environment: Sonority and temperature. *Frontiers in Communication* (3)28. 1-8.

Maddieson, Ian & Christophe Coupé. 2015. Human spoken language diversity and the acoustic adaptation hypothesis. *Journal of the Acoustical Society of America* 138(3). 1-14.

Maffi, Luisa. 2005. Linguistic, cultural, and biological diversity. *Annual Review of Anthropology* 29. 599–617.

Magenda, B. 2010. *East Kalimantan: The decline of a commercial aristocracy*. Singapore: Equinox Publishing.

Malyshev, Vladislav, Viktoria Malysheva, Angelina Gutz, Irina Novaya, Anastasia Panina, Alyona Yurkova, John Clifton & Calvin Tiessen. 2019. *The sociolinguistic situation of the Dargwa in Dagestan*. Dallas: SIL Electronic Survey Reports.

Martis, Non, Buha Aritonang, Hidayatul Astar & Wati Kurniawati. 2002. *Kosakata dasar Swadesh di kabupaten Pasir*. Jakarta: Pusat Bahasa.

Masfufah, Nurul. 2018. Hubungan kekerabatan bahasa Benuaq, Tonyooi, dan Bahau: Kajian linguistik historis komparatif. *LOA: Jurnal Ketatabahasaan dan Kesusastraan* 13(1): 27-37.

McFarland, David. 2006. *A Dictionary of Animal Behaviour*. Oxford: Oxford University Press.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746-748.

Mckaughan, Howard. 1964. A study of divergence in four New Guinea languages. *American Anthropologist* 66(4). 98-120.

McMahon, April M. S. 1994. *Understanding language change*. Cambridge: Cambridge University Press.

McMahon, April & Robert McMahon. 2013. *Evolutionary linguistics*. Cambridge: Cambridge University Press.

McWhorter, John H. 2003. *The power of Babel: A natural history of language*. New York: HarperCollins.

McWhorter, John H. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Boston: De Gruyter Mouton.

McWhorter, John H. 2014. *The language hoax*. Oxford: Oxford University Press.

McWhorter, John H. 2016. *Words on the move: Why English won't - and can't - sit still (like, literally)*. New York: Henry Holt and Company.

Metcalf, Peter. 2010. *The life of the longhouse: An archaeology of ethnicity*. New York: Cambridge University Press.

Meyer, Julien. 2021. Environmental and linguistic typology of whistled languages. *Annual Review of Linguistics* 7. 493-510.

Meyer, Ronny & Renate Richter. 2003. *Language use in Ethiopia from a network perspective: Results of a sociolinguistic survey conducted among high school students*. New York: Peter Lang Publishing.

Meyers, Jim, Heather Meyers, Ben Rice, Susan Rice & Pak Penyang. 2003. *Report on the Seruyan river surveys in Central Kalimantan*. Dallas: SIL Electronic Survey Reports.

Mithun, Marianne. 1984. How to avoid subordination. In Claudia Brugman & Monica Macaulay (eds.), *Proceedings of the tenth annual meeting of the Berkeley Linguistics Society*, 493-509. Berkeley: Berkeley Linguistics Society.

Montemagni, Simonetta, Martijn Wieling, Bob de Jonge & John Nerbonne. 2013. Synchronic patterns of Tuscan phonetic variation and diachronic change: Evidence from a dialectometric study. *Literary & Linguistic Computing* 28(1). 157-172.

Moseley, Christopher & R.E. Asher (Eds.). 1994. *Atlas of the world's languages*. London: Routledge.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2008. *Language evolution: Contact, competition and change*. New York: Continuum.

Mufwene, Salikoko S. & Cécile B. Vigouroux. 2012. Individuals, populations, and timespace: Perspectives on the ecology of language. *Cahiers de Linguistique* 32(2). 111-137.

Mühlhäusler, Peter. 1992. Preserving languages or language ecologies? A top-down approach to language survival. *Oceanic Linguistics* 31(2). 163-180.

Mühlhäusler, Peter. 2002. *Linguistic ecology: Language change and linguistic imperialism in the Pacific region*. London: Routledge.

Nagaraja, K. S., Paul Sidwell, & Simon Greenhill. 2013. A lexicostatistical study of the Khasian languages: Khasi, Pnar, Lyngngam, and War. *Mon-Khmer Studies* 42. 1-11.

Nahhas, Ramzi W. 2007. *The steps of language survey: An outline of practical methods*. Chiang Mai: SIL MSEAG Survey Team (Payap University Course Materials), Unpublished.

Nahhas, Ramzi W. & Noel W. Mann. 2006. *The steps of eliciting and analyzing word lists: A practical guide*. Chiang Mai: Payap University, Unpublished.

Needham, Rodney. 1954. Penan and Punan. *Journal of the Malayan Branch of the Royal Asiatic Society* 27(1). 73-83.

Needham, Rodney. 1955. Punan Ba. *Journal of the Malayan Branch of the Royal Asiatic Society* 28(1). 24-36.

Nettle, Daniel. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions: Biological Sciences* 367(1597). 1829-1836.

Newberry, Mitchell G., Christopher A. Ahern, Robin Clark & Joshua B. Plotkin. 2017. Detecting evolutionary forces in language change. *Nature* 551(7679). 223-226.

Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85(2). 355-404.

Ngo, Ding. 1978. *Dayung Kiaan*. Padua (Mendalam). Unpublished ms.

Ngo, Ding. 1987. *Sejarah Orang Kayan*. Padua (Mendalam).Unpublished ms.

Okushima, Mika. 1999. Wet rice cultivation and the Kayanic peoples of east Kalimantan: Some possible factors explaining their preference for dry rice cultivation. *Borneo Research Bulletin* 30. 74-104.

Okushima, Mika. 2006. Ethnohistory of the Kayanic peoples of northeast Borneo (Part 1): Evidence from their languages, old ethnonyms, and social organization. *Borneo Research Bulletin* 37. 86-126.

Okushima, Mika. 2008. Ethnohistory of the Kayanic peoples in northeast Borneo (Part 2): Expansion, regional alliance groups, and segai disturbances in the Colonial era. *Borneo Research Bulletin* 39. 184-238.

Omar, Asmah Haji. 1983. *The malay peoples of Malaysia and their languages*. Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia.

Omar, Asmah Haji & Norazuna Norahim. 2020. Lower and upper Baram sub-groups: A study of linguistic affiliation. Asmah Haji Omar, Susan & Nathan Hill (Eds.) *The GLOCAL Proceedings Asia 2020*, 92-100. Bintulu: GLOCAL Conference on Asian Linguistic Anthropology.

Oliver, Michelle. 2013. Language mapping for education policy realisation from below. *Language Matters* 44(3). 6-32.

Oliveira Jr., Miguel & Daniel L. Everett. 2010. Remarks on the Pirahã suffix –sai and complex syntax (Preliminary Report). Unpublished ms.

Osborne, Milton. 2016. *Southeast Asia: An introductory history* (12th edition). Sydney: Allen & Unwin.

Ouattara, Karim, Alban Lemasson & Klaus Zuberbühler. 2009. Campbell's monkeys use affixation to alter call meaning. *PLoS ONE* 4(11). 1-7.

Padoch, Christine & Nancy Lee Peluso. 2003. *Borneo in transition: People, forests, conservation, and development* (2nd Edition). Kuala Lumpur: Oxford University Press.

Pagel, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics* 10(6). 405–415.

Pagel, Mark. 2012. *Wired for culture: Origins of the human social mind*. New York: W.W. Norton & Company.

Pashler, H. 1988. Familiarity and visual change detection. *Perception & Psychophysics* 44. 369-378.

Pelkey, Jamin R. 2011. *Dialectology as dialectic: Interpreting Phula variation*. Berlin: De Gruyter Mouton.

Peluso, N. L. 1983. *Markets and merchants: The forest products trade of East Kalimantan in historical perspective*. Ithaca, NY: Cornell University masters thesis.

Pierce Colfer, Carol J. 1983. Change and indigenous agroforestry in East Kalimantan. *Borneo Research Bulletin* 15(1). 3–20.

Pinker, Steven. 2010. The cognitive niche: Coevolution of intelligence, sociality, and language. *PNAS* 107(2). 8993–8999.

Pinker, Steven & Paul Bloom. 1990. Natural language and natural selection. *Behavioral & Brain Sciences* 13(4). 707–784.

Pleyer, Michael. 2020. Pretend play as a scaffold for language evolution. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd & T. Verhoef (eds.), *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13)*, 345-355. Nijmegen: Max Planck Institute for Psycholinguistics.

Progovac, Ljiljana, Natalia Rakhlin, William Angell, Ryan Liddane, Lingfei Tang & Noa Ofen. 2018. Diversity of grammars and their diverging evolutionary and processing paths: Evidence from functional MRI study of Serbian. *Frontiers in Psychology* 9(278). 1-13.

Pullum, Geoffrey K. 1991. *The Great Eskimo Vocabulary Hoax and other irreverent essays on the study of language*. Chicago: University of Chicago Press.

Punan Mahkam. Kampung Punan Mahkam. http://punanmahkam.com/first/kategori/1 (15 March 2020).

Puri, Rajindra K. 2001. *Bulungan ethnobiology handbook*. Bogor: CIFOR.

Radloff, Carla F. 1991. *Sentence repetition testing for studies of community bilingualism*. Dallas: Summer Institute of Linguistics/University of Texas at Arlington.

Ray, Sidney H. 1913. *The languages of Borneo*. London: West, Newman & Co.

Reali, Florencia, Nick Chater & Morten H. Christiansen. 2018. Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society of London B: Biological Sciences* 285(1871). 1-6.

Remez, R. E., D. F. Ferro, K. R. Dubowski, J. Meer, R. S. Broder & M. L. Davids. 2010. Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, and Psychophysics* 72. 2054-2058.

Revel, Nicole. 1982. Synchronical description at the phonetic and syllabic level of Modang (Kalimantan Timur) in contrast to Kenyah-Kayan and Palawan (Philippines). In Paul Geraghty, Amran Halim, Lois Carrington & Stephen A. Wurm (Eds.). *FOCAL 1: Papers from the International Conference on Austronesian Linguistics*, 321-331. Canberra: Pacific Linguistics.

Revel-MacDonald, Nicole. 1978. La danse des hudoq (Kalimantan Timur). *Objets et Mondes* 18(1/2). 31–44.

Revel-Macdonald, Nicole & Maryke Leyenaar. 1982. Dayak and Kalimantan masks. *The Drama Review: TDR* 26(4). 70-72.

Riwut, Tjilik. 1979. *Kalimantan membangun*. Jakarta: Endang.

Romaine, Suzanne. 2000. *Language in society: An introduction to sociolinguistics*. Oxford: Oxford University Press.

Rooy, Raf Van. 2020. *Language or dialect? The history of a conceptual pair*. Oxford: Oxford University Press.

Roth, Henry Ling. 1896. *The natives of Sarawak and British North Borneo*. London: Truslove & Hansen.

Rousseau, Jérôme. 1974a. A vocabulary of Baluy Kayan. *Sarawak Museum Journal* 22. 93-152.

Rousseau, Jérôme. 1974b. The Baluy area. *Sarawak Museum Journal* 22. 17-27.

Rousseau, Jérôme. 1983. Kayan personal names. *Sarawak Museum Journal* 32(53). 251-72.

Rousseau, Jérôme. 1990. *Central Borneo: Ethnic identity and social life in a stratified society*. New York: Oxford University Press.

Rousseau, Jérôme. 2001. Hereditary stratification in middle-range societies. *The Journal of the Royal Anthropological Institute* 7(1). 117-131.

Rousseau, Jérôme. 2006. *Rethinking social evolution: The perspective from middle-range societies*. Montreal: McGill-Queen's University Press.

Russell, A. Sue. 1999. *Conversion, identity, and power: The impact of christianity on power relationships and social exchanges*. lanham: University Press of America.

Sachs, Jeffrey. 2015. *The age of sustainable development*. New York: Columbia University Press.

Sagan, Jacob Dungau. 1989. The Kenyah people of Sarawak. *Sarawak Museum Journal* 40. 119-141.

Sánchez Carrión, José María. 1985. *La nueva sociolingüistica y la ecología de las lenguas*. Donostia-San Sebastián: Eusko Ikaskuntza.

Sandin, Benedict. 1985. Notes on the Sian (Sihan) of Belaga. *Sarawak Museum Journal* 34. 67-76.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace & Company.

Sauerland, U. 2010. Experimental evidence for complex syntax in Pirahã. Ms, ZAS Berlin. Available from: http://www.zas.gwz-berlin.de/574.html

Schiller, Anne. 2007. Activism and identities in an East Kalimantan Dayak organization. *The Journal of Asian Studies* 66(1). 63-95.

Schmidt, Jurgen Erich. 2010. Language and space: The linguistic dynamics approach. In Peter Auer and Jürgen Erich Schmidt (eds.), *Language and Space: An International Handbook of Linguistic Variation (Volume 1)*, 201-225. New York: De Gruyter Mouton.

Schneeberger, W. F. 1979 *Contributions to the ethnology of Central Northeast Borneo: Parts of Kalimantan, Sarawak and Sabah*. Berne: University of Berne Institute of Ethnology.

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35. 335-357.

Sellato, Bernard. 1980. The Upper Mahakam area. *Borneo Research Bulletin* 12. 40-46.

Sellato, Bernard. 1981. La région de la haute Mahakam, Est Kalimantan. *Archipel* 22: 21-42.

Sellato, Bernard. 1981. Three-gender personal pronouns in some languages of central Borneo. *Borneo Research Bulletin* 13. 48-49.

Sellato, Bernard. 1992. *Hornbill and dragon: Arts and culture of Borneo* (2nd Edition). Singapore: Sun Tree.

Sellato, Bernard. 1993. Collective memory and nomadism: Ethno-historical investigations in Borneo. *Indonesia* 57. 155-174.

Sellato, Bernard. 1994. Forêts tropicales et sociétés traditionnelles à Bornéo : Vers une histoire régionale 'en continu' de l'environnement et des systèmes de subsistence. *Écologie Humaine* 12(2): 3-22.

Sellato, Bernard. 2001. *Forest, resources and people in Bulungan: Elements for a history of settlement, trade, and social dynamics in Borneo, 1880-2000*. Bogor: CIFOR.

Sellato, Bernard. 2002a. *Innermost Borneo: Studies in Dayak cultures*. Paris: SevenOrients.

Sellato, Bernard. 2002. Non-timber forest products and trade in eastern Borneo. *Bois & forêts des tropiques* 271: 37-50.

Sellato, Bernard. 2007. Resourceful children of the forest: the Kalimantan Punan through the twentieth century. In Peter Sercombe & Bernard Sellato (eds.), *Beyond the Green Myth: Borneo's Hunter-Gatherers in the 21st Century*, 61–90. Copenhagen: Nordic Institute of Asian Studies.

Sellato, Bernard. 2016. The Ngorek of the central highlands and "megalithic" activity in Borneo. In Christian Jeunesse, Pierre Le Roux & Bruno Boulestin (eds.), *Living and past megalithisms: Interwoven approaches (Mégalithismes vivants et passés: Approches croisées)*, 117–149. Oxford: Archaeopress Archaeology.

Sellato, Bernard & Antonia Soriente. 2015. The languages and peoples of the Müller Mountains: A contribution to the study of the origins of Borneo's nomads and their languages. *Wacana* 16(2). 339–354.

Seyfarth, Robert M., Dorothy L. Cheney & Peter Marler. 1980. Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour* 28(4). 1070–1094.

Shackelford, Sarah. 2019. *A sociolinguistic study of Dewas Rai and Danuwar*. Dallas: SIL Electronic Survey Reports.

Shiva, Vandana. 1993. *Monocultures of the mind: Perspectives on biodiversity and biotechnology*. London: Zed Books Ltd.

Simons, Gary F. 1979. *Language variation and limits to communication*. Dallas: Summer Institute of Linguistics.

Sinnemäki, Kaius. 2020. Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics* 6(2). 1-39.

Sinha, C. 2009. Language as a biocultural niche and social institution. In V. Evans & S. Pourcel (eds.), New Directions in Cognitive Linguistics, 289-310. Philadelphia: John Benjamins.

Sitorus, Soaduon, Patrice Levang, Edmond Dounias, Dollop Mamung, & Darif Abot. 2004. *Potret Punan Kalimantan Timur: Sensus Punan 2002-2003*. Bogor: CIFOR.

Skeat, W. W. & C. O. Blagden. 1906. *Pagan Races of the Malay Peninsula* (2 Volumes). London: MacMillan.

Skutnabb-Kangas, Tove & David Harmon. 2018. Biological diversity and language diversity. In Alwin F. Fill & Hermine Penz (eds.). *The Routledge Handbook of Ecolinguistics*. New York: Routledge.

Slater, Evan, Jacob Hanley, Thomas Mulja, Marcos Zentilli & Corwin Trottier. 2020. Epithermal mineralization in the Busang Southeast Zone, Indonesia: New insight into the Au prospect at the center of the Bre-X fraud. *Minerals* 10(698). 1–16.

Smalley, William. 1993. *Linguistic diversity and national unity: Language ecology in Thailand*. Chicago: University of Chicago.

Smith, Alexander D. 2015. On the classification of Kenyah and Kayanic languages. *Oceanic Linguistics* 54(2). 333–357.

Smith, Alexander D. 2017a. *The languages of Borneo: A comprehensive classification*. Manoa, HI: University of Hawai'i dissertation.

Smith, Alexander D. 2017b. Merap historical phonology in the context of a central Bornean linguistic area. *Oceanic Linguistics* 56(1). 143–180.

Smith, Alexander D. 2018. *Modang Wahau and Woq Helaq Wordlist*. Archival Collection, Over 800 items. Kaipuleohone: University of Hawai'i Digital Language Archive. https://scholarspace.manoa.hawaii.edu/handle/10125/70560

Smith, Alexander D. 2019. A reconstruction of Proto-Segai-Modang. *Oceanic Linguistics* 58(2). 353–385.

Smith, Alexander D. & Taraka Rama. 2022. Environmental factors affect the evolution of linguistic subgroups in Borneo. *Diachronica* 39(2). 193–225.

Société de Linguistique de Paris, Statuts de 1866 [Article 2] in 1866. http://www.slp-paris.com/statuts1866.html

Sombroek, H. 1986. *Kamus Bahasa Busang-Indonesia-Belanda*. Samarinda: MSF.

Soriente, Antonia. 1997. The Kenyah isolects of Long Pujungan district in North-East Kalimantan. In Cecilia Odé & W. A. L. Stokhof (eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 713–738. Amsterdam: Rodopi.

Soriente, Antonia. 2003. *A classification of Kenyah languages in Sarawak and East Kalimantan*. Bangi: Universiti Kebangsaan Malaysia dissertation.

Soriente, Antonia. 2008. The classification of Kenyah languages: A preliminary statement. *Journal of the SEAsian Linguistics Society*. 14(2). 49-62.

Soriente, Antonia. 2013. Undergoer voice in Borneo: Penan, Punan, Kenyah and Kayan languages. *NUSA: Linguistic studies of languages in and around Indonesia*. 175-203.

Soriente, Antonia. 2014. Studying linguistic and cultural contact in Borneo: Prospects and challenges. *Antropologia* 1(1). 59–81.

Soriente, Antonia. 2015. Mbraa: A Modang-Bahau language? (Paper presented at 13-ICAL, Academia Sinica, Taipei, Taiwan, July 18-31).

Soriente, Antonia. 2020. Hunter-gatherers of Borneo and their languages. In Tom Güldemann, Patrick McConvell & Richard A. Rhodes (Eds.), *The Language of Hunter-Gatherers*, 262-307. Cambridge: Cambridge University Press.

Soriente, Antonia & Kazuya Inagaki. 2012. *Kalimantan languages: An overview of current research and documentation*. Tokyo University of Foreign Studies.

Southwell, C. Hudson. 1990. *Kayan-English dictionary*. Kuching: Sarawak Literary Society.

Sprachatlas der deutschen Schweiz (SDS) [Language Atlas of German-speaking Switzerland] https://sprachatlas.ch

Stalder, Jurg. 1996. Rapid appraisal. *Notes on Literature in Use and Language Programs* 48(6). 5–23.

Stanlaw, James. 2021. Ecolinguistics. In James Stanlaw (Ed.). *The International Encyclopedia of Linguistic Anthropology*, 1-2, Hoboken: John Wiley & Sons.

Starks, Donna & Zita McRobbie-Utasi. 2001. Collecting sociolinguistic data: Some typical and some not so typical approaches. *New Zealand Journal of Sociology* 16. 79–92.

Steels, Luc & Eörs Szathmáry. 2018. The evolutionary dynamics of language. *Biosystems* 164. 128-137.

Steffensen, Sune Vork & Joshua Nash (eds.). 2007. *Jørgen Christian Bang & Jørgen Døør: Language, ecology and society: An introduction to dialectical linguistics*. London: Continuum.

Steffensen, Sune Vork & Alwin F. Fill. 2014. Ecolinguistics: The state of the art and future horizons. *Language Sciences* 41. 6–25.

Strouthes, Daniel. 1993. Kenyah-Kayan-Kajang. In Paul Hockings (ed.). *Encyclopaedia of World Cultures (Volume V - East and Southeast Asia)*, 133-134. New York: G.K. Hall & Co.

Sugono, Dendy, S.S.T. Wisnu Sasangka, & Ovi Soviaty Rivay (eds.). 2017. *Bahasa dan peta bahasa di Indonesia*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa (Kementerian Pendidikan dan Kebudayaan).

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–37.

Szczepaniak, Renata. 2015. Syntaktische Einheitenbildung - typologisch und diachron betrachtet. In Christa Dürscheid & Jan Georg Schneider (eds.), *Handbuch Satz, Äußerung, Schema*, 104-124. Berlin: Walter de Gruyter.

Tagliamonte, S. 2006. *Analyzing sociolinguistic variation*. Cambridge: Cambridge University Press.

Talmy, Leonard. 2000. *Force dynamics in language and cognition: Toward a cognitive semantics 1: Concept structuring systems*. Cambridge: MIT Press.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tim Pemetaan Bahasa. 2017. *Bahasa dan peta bahasa di Indonesia*. Jakarta: Badan Bahasa.

Tinits, Peeter, Jonas Nölle & Stefan Hartmann. 2017. Usage context influences the evolution of overspecification in iterated learning. *Journal of Language Evolution* 2(2). 148–159.

Truong, Christina Lai, and Lilian Garcez. 2012. Participatory methods for language documentation and conservation: Building community awareness and engagement. *Language Documentation & Conservation* 6. 22–37.

UN Sustainability Goals https://sdgs.un.org/ (21 August 2020).

Uyo, Lah Jau. 1989. Kayan people of Sarawak. *Sarawak Museum Journal* 40. 56–88.

Vico, Giambattista. 1988 [1711]. *On the most ancient wisdom of the italians unearthed from the origins of the Latin language: Including the disputation with the giornale de' letterati d'Italia*. Ithaka: Cornell University Press.

Voegelin, C. F. & F. M. Voegelin. 1977. *Classification and index of the world's languages*. New York: Elsevier.

Voegelin, C.F., F. M. Voegelin & Noel W. Schutz, Jr. 1967. The language situation in Arizona as part of the Southwest culture area. In Dell Hymes & William E. Bittle (eds.), *Studies in Southwestern Ethnolinguistics: Meaning and history in the languages of the American Southwest*, 403-51. The Hague: Mouton.

Wadley, Reed L. 2000. Reconsidering an ethnic label in Borneo: The "Maloh" of West Kalimantan, Indonesia. *Bijdragen tot de Taal-, Land- en Volkenkunde* 156(1). 83–101.

Wati, Kuriawati, Non Martis, Buha Aritonang, & Hidayatul Astar. 2002. *Kosakata dasar Swadesh di kabupaten Berau, Kotamadya Samarinda, dan kotamadya Balikpapan*. Jakarta: Pusat Bahasa.

Wegmann, Monika. 2016. *Language in space: The cartographic representation of dialects*. Strasbourg: Editions de Linguistique et de Philologie.

Wehea Dayak people of Nehas Liah Bing. http://weheapetkuq.com/ (18 April, 2019).

Wexler, Kenneth N. 2017. A program for the genetics of grammar. *Biolinguistics* 11. 295-323.

Whalen, D. H. and Gary F. Simons. 2012. Endangered language families. *Language* 88(1). 155–173.

Wieling, Martijn. & John Nerbonne. 2010. Hierarchical spectral partitioning of bipartite graphs to cluster dialects and identify distinguishing features. *Proceedings of the 2010 ACL Workshop on Graph-based Methods for Natural Language Processing*. 33–41.

Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE* 6(9). 1-14.

Winters, James, Simon Kirby & Kenny Smith. 2015. Languages adapt to their contextual niche. *Language & Cognition* 7(3). 415-449.

Winzeler, Robert L. 2011. *The peoples of Southeast Asia today: Ethnography, ethnology, and change in a complex region*. Lanham: AltaMira Press.

Wong, Patrick C. M., Xin Kang, Kay H. Y. Wong, Hon-Cheong So, Kwong Wai Choy & Xiujuan Geng. 2020. ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances* 6(22). 1-10.

Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3). 543–578.

Wurm, S. A. & S. Hattori (Eds.) 1983. *Language atlas of the Pacific area. Part II: Japan area, Taiwan (Formosa), Philippines, Mainland and Insular South-East Asia* (Series C, No. 67). Tokyo/Canberra: The Japan Academy/Pacific Linguistics.

Yeq, Yohana. 2013. *Pengulangan kata dalam bahasa Dayak Bahau, di Long Lunuk, Kabupaten Mahakam Ulu, Kalimantan Timur*. Depok, Indonesia: Universitas Sanata Dharma thesis (skripsi).

Yoder, Zachariah. 2017. The reliability of recorded text test scores: widespread inconsistent intelligibility testing in minority languages. *Journal of Multilingual and Multicultural Development* 38(9). 843–855.

Z'Graggen, Johannes A. 1975. The Madang-Adelbert Range Sub-Phylum. In Stephen A. Wurm (ed.), *New Guinea area languages and language study (Volume 1): Papuan languages and the New Guinea linguistic scene* (Pacific Linguistics: Series C 38), 569-612. Canberra: Australian National University Press.

Z'graggen, Johannes A. 1980. *A comparative word list of the Rai Coast languages, Madang Province, Papua New Guinea*. Canberra: Australian National University Press.

Zhang, Menghan & Tao Gong. 2016. How many is enough?—Statistical principles for lexicostatistics. *Frontiers in Psychology* 7(1916). 1-12.