What's in a Name?

A Typological and Phylogenetic Analysis of the Names of Pama-Nyungan Languages

Katherine Rosenberg
Advisor: Claire Bowern

Submitted to the faculty of the Department of Linguistics in partial fulfillment of the requirements for the degree of Bachelor of Arts

Yale University
May 2018

Abstract

The naming strategies used by Pama-Nyungan languages to refer to themselves show remarkably similar properties across the family. Names with similar meanings and constructions pop up across the family, even in languages that are not particularly closely related, such as *Pitta Pitta* and *Mathi Mathi*, which both feature reduplication, or *Guwa* and *Kalaw Kawaw Ya* which are both based on their respective words for 'west.' This variation within a closed set and similarity among related languages suggests the development of language names might be phylogenetic, as other aspects of historical linguistics have been shown to be; if this were the case, it would be possible to reconstruct the naming strategies used by the various ancestors of the Pama-Nyungan languages that are currently known. This is somewhat surprising, as names wouldn't necessarily operate or develop in the same way as other aspects of language; this thesis seeks to determine whether it is indeed possible to analyze the names of Pama-Nyungan languages phylogenetically. In order to attempt such an analysis, however, it is necessary to have a principled classification system capable of capturing both the similarities and differences among various names. While people have noted some similarities and tendencies in Pama-Nyungan names before (McConvell 2006; Sutton 1979), no one has addressed this comprehensively. In the first section of this thesis, I therefore propose a classificatory system for Pama-Nyungan languages that can capture the similarity between names like *Pitta Pitta* and *Mathi Mathi*, while also recognizing that *Mathi Mathi* and *Nyawaygi*, which are both based on the word for 'no,' also have a feature in common. The result is a comprehensive hierarchical classification system that captures both the semantic and structural aspects of language names. I then use this classification system as the basis for my exploration of whether language names can in fact be considered to be phylogenetic; ultimately, I show that there is very little phylogenetic signal, which suggests that there must be other factors at play.

Acknowledgments

Words cannot convey how grateful I am to so many people who have helped me in the process of writing my senior essay. First and foremost, I would like to thank Claire Bowern, who was not only an incredibly patient, supportive, and insightful advisor this past year, but also first introduced me to the Pama-Nyungan language family and theories of phylogenetic approaches to language change. Without her, this thesis would truly not exist. I would also like to thank Raffaella Zanuttini and my fellow senior linguistics majors for their unfailing support and encouragement as we all navigated the thesis-writing process together. Thank you as well to my family, who have done their best to seem interested when I drone on about shibbolethnonyms; to my first years for being an unfailing source of support, love, and even chocolate; and to the entire Yale Linguistics Department and Ezra Stiles College for providing two spaces that have truly felt like home for the past four years.

Contents

List of Figures

1 Introduction

Linguists tend to gloss over the names of the languages we research, diving headfirst into the language without thinking about the name. This ignores a potentially useful source of information on linguistic identity, naming conventions, interactions between languages, and language change more generally. In this thesis, I look address the names of the Pama-Nyungan languages of Australia, focusing on how they are composed both semantically and structurally, using these patterns to address issues of language contact and change.

The naming strategies used by Pama-Nyungan languages to refer to themselves show remarkably similar properties across the family. Names with similar meanings and constructions pop up across the family, even in languages that are not particularly closely related, such as *Pitta Pitta* and *Mathi Mathi*, which both feature reduplication, or *Guwa* and *Kalaw Kawaw Ya* which are both based on their respective words for 'west.' This combination of variation within a closed set and similarity among related languages suggests the development of language names might have evolved within the family; if this is the case, it should be possible to reconstruct the naming strategies used by various ancestors of the Pama-Nyungan languages that are currently known.

Many aspects of language have been shown to be analyzable using computational biological methods; people have used phylogenetic methods to construct trees of language families (Gray et al. 2009; Bowern & Atkinson 2012), as well as to reconstruct past vocabulary (Haynie & Bowern 2016; Jordan 2013). Language naming strategies appear to have a distribution that might be the result of phylogenetic processes, in which case they too should be reconstructable. However, language names might not be conducive to this sort of analysis, as names wouldn't necessarily operate or develop in the same way as other aspects of language; this thesis seeks to determine whether it is indeed possible to analyze the names of Pama-Nyungan languages phylogenetically. In order to attempt such an analysis, however, it is necessary to have a principled classification system capable of capturing both the similarities and differences among various names. While people have noted some similarities and tendencies in Pama-Nyungan names before (McConvell 2006; Sutton 1979), no one has addressed this comprehensively.

In Section 2.1 of this thesis, I address the previous treatments of Pama-Nyungan names, and their shortcomings. In Section 2.1.3, I propose a classificatory system for Pama-Nyungan languages that can capture the similarity between names like *Pitta Pitta* and *Mathi Mathi*, while also recognizing that *Mathi Mathi* and *Nyawaygi*, which are both based on the word for 'no,' also have a feature in common. The result is a comprehensive hierarchical classification system that captures both the semantic and structural aspects of language names. I then use this system in Section 2.2 to present a typological survey of the naming strategies present within Pama-Nyungan. Finally, in Section 3, I use this classification system as the basis for my exploration of whether language names can in fact be considered to be phylogenetic; ultimately, I show that while there is some phylogenetic signal, much of the distribution of naming strategies cannot be explained as purely the result of evolution.

1.1 Language Names

Very few linguists study a language without learning its name, and yet we tend to ignore these names as sources of valuable information. As Isabelle Léglise and Bettina Migge explain in their 2006 paper on language names,

> "Although it is well accepted that a name for a language (or people) is never neutral...the potential of names for shedding light on the social and linguistic reality of a particular linguistic situation has not yet been fully explored. Naming conventions are rarely investigated in much detail. They are generally discussed only briefly in the introductory sections of studies dealing with specific (socio)linguistic topics about the language so named."

Even where the names of languages are explicitly discussed, Léglise and Migge point out that this discussion usually takes the form of the politically-influenced debates of what the "right" name for a language is, or whether two speaker communities should be considered to speak the same language. Léglise & Migge call attention to the famous debates of whether to use the exonym "Eskimo" or the endomym "Inuit," or whether "Serbo-Croat" should be considered two distinct languages (Léglise & Migge 2006: 314). While these questions are certainly interesting, the names themselves are not investigated in great detail; rather the true questions are whether to use an endonym or an exonym, and whether two names refer to the same language. It is the political connotations of the names that come into question, rather than the names themselves. Even in in-depth studies of a language, the etymology of its name is often glossed-over or ignored as unimportant.

But the names of languages reflect the identity of their speakers. As Jan Wohlgemuth 2015 discusses, many language names are built on an us-vs-them contrast, and are meant to differentiate the group of people who speak that language from other groups. This can take many different forms: some language names are truly glossonyms that describe the language itself, but many language names are also derived from the ethnonym for their speaker community, or a toponym referring to the place they are spoken.

Which naming strategy a language name uses therefore reflects an important aspect of that language's identity. Investigating naming strategies overall is therefore interesting, because it shows what aspects of identity are salient to language names; likewise the analysis of an individual language name might provide valuable insight into the self-image of that language's speakers. What is also interesting, however, is looking at the pattern of these semantically-laden strategies: how are they distributed across the family, and do they pattern like other aspects of language?

1.2 The Pama-Nyungan Languages of Australia

Pama-Nyungan is a family of about 300 languages that are distributed across 90% of mainland Australia; these languages make up roughly two thirds of the languages spoken in Australia pre-colonization (Bowern & Atkinson 2012). The family was first proposed in the 1960s by Kenneth Hale (1966), and various classifications of Pama-Nyungan have been attempted since. While there are many agreed-upon subgroups of the family, however, there has been disagreement over the higher level groupings within Pama-Nyungan (Koch 2014). For this thesis, I use the genetic grouping developed by Bouckaert, Bowern, and Atkinson in their 2018 paper "The origin and expansion of Pama-Nyungan languages across Australia."

1.3 Data Collection

The data that I use for this thesis was mostly gathered from reference grammars and dictionaries of the Pama-Nyungan languages, as well as some more general Australian language handbooks (Dixon & Blake 1979–2000; Wafer & Lissarague 2008) and input from researchers who study these languages.

Unfortunately, the nature of this type of research guarantees that data collection will be frustrating and incomplete, because I am bound by the work of others. In some cases, this means that there is simply nothing to be found: no one has documented the language in question, or recorded if its name meant something. Other Pama-Nyungan languages, however, have very thorough reference grammars written about them; in these grammars, there is often a short discussion about the name of the language, including the various alternate names and spellings that have been recorded for it. While this discussion sometimes includes the etymology of the name, it often does not. Even where the name is discussed, it is not always broken down completely, because only part of it seems to be interesting. Often, the author will explain that the name "is based on the word for *x*" and leave the discussion there, despite it being clear that there is another component to the name. Additionally, because there is no universal standard for discussing and analyzing language names, some amount of interpretation is required to determine whether two linguists are using different terms to identify similar phenomena. It is also possible that a name is only partially analyzable, like *Mayi-Yapi*, in which *Mayi* means 'language' but *yapi* is unknown (Breen 1981: 2). All of this combines to mean that not only is the data sometimes missing, but even where there is a description, it is sometimes incomplete.

Despite this, it is possible to find etymologies for many Pama-Nyungan language names. Of the 309 languages in Atkinson and Bowern's Pama-Nyungan tree, I have collected data on the meanings of 154 names, or roughly 50%. I did my best to only include *endonyms*, the names that languages bestow upon themselves, rather than *exonyms*, which are imposed by outsiders. Because of this, 5 of the names have been discounted because they are names like "Flinders Island" that were clearly bestowed by English-speaking colonialists. Because this thesis is focused on how speakers name their own languages, this is not relevant data. There also exist languages without names, such as the languages of Tasmania (Claire Bowern p.c.); while the lack of a name is an interesting naming strategy in its own right, if there are any of these languages within the Pama-Nyungan family, it would be impossible to distinguish them from the languages that had names that have been lost to history, and so I do not include any in my data.

2 A Typology of Australian Language Names

In order to discuss the evolution of various types of language names, it is necessary to first discuss what type of names exist. This requires having a principled system of referring to the meanings of names, as well as their similarities and differences. It is not possible to classify languages directly by their exact meaning and be at all descriptive: even counting all unanalyzable names as instances of "the same meaning", all names based on animal species as the same, and all place-based names as the same, there are about 30 different categories of naming types, many of which only have one token. Beyond involving numerically too many categories to be useful, this also misses the generalization that there is something in common between the languages whose names mean *no-language*, *no-no*, *no*, *no-having*, and *no-X*, or between names meaning *no-having*, *go-having*, and *this-having*.

2.1 Classification System

Ideally, a classificatory system will be able to capture both similarities and differences among language names. There has been little effort in the past to create a detailed description of the types of names that are used by Pama-Nyungan languages. The two descriptions that do exist, by McConvell (2006) and Sutton (1979), have fatal flaws that prevent them from applying to Pama-Nyungan names in general: they focus only on semantic features, ignoring important structural distinctions, and do not have categories that can apply to all Pama-Nyungan names. In the sections to follow, I will first look at Jan Wohlgemuth's general principles about the types of language names that exist; I will then look at these prior explorations of Australian language names and explain why they are inadequate for a large scale, detailed description of Pama-Nyungan names, before proposing my own system.

2.1.1 General Principles Concerning Language Names

Jan Wohlgemuth (2015) notes that many, if not most, language names in the world can be traced back to the name for their speaker community (*ethnonyms*), or for the region in which they were spoken (*toponyms*); some feature an ethnonym/toponym combined with a head noun that means 'language, speech'. He categorizes such names as being based on the word for 'people,' describing 'others,' marking 'esteem' (i.e. 'true people', 'upright people'), or describing 'physical features' (Wohlgemuth 2015: 8). Other possible bases for names include 'phenomenological' names which mean things like 'speech,' and 'mythological' names like *Hawai'i* that apparently refer to important cultural stories (Wohlgemuth 2015: 9).

He also discusses the shibboleth principle, in which languages are named after a particular word that distinguishes them from other languages. The choice of shibboleth varies by language; while many Australian languages are based on the word for 'no,' the French dialects called *langues d'oil* and *langues d'oc* are based on the shibboleths *oil* and *oc*, which mean 'yes' in their respective dialects (Wohlgemuth 2015: 10). Other possible shibboleths are interrogatives, numerals, demonstratives, prominent verbs, or distinctive pronunciations.

Wohlgemuth points out the necessity of distinguishing endonyms from exonyms, and notes that language names are often contrastive, meant to separate one language from another in an us v. them fashion. Typically, he suggests, endonyms have positive connotations while exonyms have negative ones. In my analysis, I have tried to limit my discussion to endonyms, which is why English names that are clearly exonyms have not been included. In some cases, it is possible that a name was originally an exonym that was adopted into the language it was describing, but while these are historically exonyms, they are contemporarily endonyms and therefore count for the purpose of analyzing how speakers refer to their own languages.

Wohlgemuth's analysis provides a good starting point for thinking about Pama-Nyungan languages, but it does not cover all the possibilities found in Pama-Nyungan. I turn now to two previous treatments of language names in Australia.

2.1.2 Previous Approaches to Australian Names

McConvell In his 2006 article "Shibbolethnonyms, ex-exonyms and eco-ethnonyms in Aboriginal Australia: the pragmatics of onymization and archaism" Patrick McConvell seeks to categorize Australian aboriginal language names into three groups. Generally, he claims that Australian language names are built on salient contrasts, either among the languages themselves or among the locations in which the languages are spoken. He also examines how historical processes have obscured some of the meanings of language names: some are completely opaque to etymological analysis, but others can be interpreted given enough linguistic and historical knowledge. By analyzing the names that can be interpreted, McConvell hopes to be able to make claims about population contact and self-identity.

McConvell divides aboriginal ethnonyms into the categories of *shibbolethnonyms*, which are built from the distinctive shibboleths that Wohlgemuth discusses; *directional exonyms*, built from cardinal direction terms; and *environmental ethnonyms*, based on names for places or dominant features of a specific environmental zone, like a type of plant. To come to this conclusion, he analyzes the processes by which he assumes various names were developed for languages in the Victoria River District of the Northern Territory. He looks at data both from languages with transparent meaning and others with what he calls "semi-transparent" meaning—names whose meanings can be interpreted using historical and linguistic knowledge, but are no longer perfectly transparent. After introducing these definitions, he also introduces the concept of an ethnonymic paradigm zone, in which the name for one language is derived to have a meaning that opposes the name of another. An example of this is the contrasting shibbolethnonyms Pitjantjatjarra and Yankunytjatjara, which are based on their respective words for "to go"; this is a salient contrast between two closely related languages, and so the language with *pitjan* is distinguished from the language with *yankuny.* These names are clearly not inherited from their common ancestor, and yet might appear to be the result of evolution, because the end result is two closely-related languages that use the same strategy.

McConvell argues that ethnonyms based on directional terms necessarily develop from exonyms that are borrowed from other languages; in the case of Karranga, for example, one population referred to another as "southerners," and that name was borrowed into the language, according to the phonological rules of the new language. These names are often hard to analyze, because not only are they based on terms from another language, but groups have also moved, so that the directional relationships are no longer the same. According to McConvell, the existence of such "ex-exonyms" emphasizes the interconnectedness of speaker populations, as members of one language community interacted enough with the other to adopt an outside name for their language.

Environmental ethnonyms, on the other hand, are based on distinctive environmental features of the area in which a population lives, such as Pirli-ngarna, which McConvell says means "hill-dwellers." This forms a paradigm with Pinka-ngarna, meaing "river dwellers." In this case, it is not the words of the languages that are being contrasted, as with a paradigm based on shibboleths, but the places where the speakers live.

Any of these etymologies can become opaque as the result of changes in language, which can prevent the meaning of the names from being analyzable, or cause a former shibboleth to become meaningless. Other changes like population movement can cause language names to be inaccurate if they refer to relative directions, or to salient environmental features. Any of these etymological changes can also cause the names to no longer be transparently contrastive. McConvell claims that the presence of such "ethnonymic paradigms" of language names that contrast with each other, shows that there was sustained contact among speakers of different languages, enough that they could borrow their names from each other, or refer to themselves by mutually contrastive names. This knowledge can be used to determine which populations had contact with each other, and when. If, for example, one language's name comes from another language's word for 'south' but the first population is not located to the south of the second, this name presumably originates from an earlier time when one of the two groups lived elsewhere.

McConvell's analysis is mainly focused on languages of the Victoria River District of the Northern Territory of Australia, but he suggests that it can be expanded to apply to the rest of Australia, and perhaps the world's languages more generally. While the concepts of shibbolethnonyms and ethnonymic paradigms are very interesting, and potentially very useful, this three-way classification of language names does not clearly generalize to the entire Pama-Nyungan family. It cannot be used to describe all the different naming strategies that occur in Pama-Nyungan languages, such as the many language names that are simply monomorphemic ethnonyms. Moreover, it is not clear that every case of a compass-based name is an instance of a former exonym, rather than a self-identification as a group that lives in a specific place in comparison to others. The three categories that he proposes, additionally, are very broad, and therefore do not provide much insight into the wide degree of variation found among Pama-Nyungan language names. Unfortunately, therefore, McConvell's system cannot adequately describe the range of naming strategies used by Pama-Nyungan languages, although the concept of names set up in opposition to each other is potentially very valuable.

Sutton Peter Sutton's 1979 paper "Australian Language Names" is meant to be a practical introduction to the issues facing linguists who study Australian languages. He begins with a discussion of how language names have traditionally been spelled, before beginning the discussion of what the names themselves mean, and the difficulty of gathering such data. Unfortunately, he gives more of an impressionistic description than a detailed typological survey, and so, despite making some interesting points, he does not fully describe the variation present in Pama-Nyungan names.

Sutton makes distinctions among names (1) referring to dialect characteristics, (2) referring to locations (both place names and compass terms), (3) referring to mythological stories, and (4) referring to human characteristics, although he notes that there are other names that do not fit these categories. This seems promising as a way to create meaningful categories without creating too many, but because Sutton is focused more on over-arching semantic differences than low-level or structural ones, it will not address the difference in naming strategies among four language names meaning *no-no, no-having, no-language*, and *this-having*; instead it will group them all as falling into category 1, referring to dialect characteristics. Moreover, the ideal descriptive system will be able to classify all (or nearly all) of the language names in the family, including the large number of languages whose name is simply a monomorphemic clan name, which are not covered by Sutton's description.

2.1.3 A Hierarchical Approach

As neither McConvell nor Sutton's approach to classifying languages is either accurate or descriptive enough to apply to all the languages in the Pama-Nyungan family, it was necessary to come up with an alternate method of classifying language names.

Trying to classify the names of languages by their specific meanings clearly does not have much descriptive power. There are 34 different categories, 14 of which have a single token each, and there is no way to show that some names, while different overall, clearly have something in common, like *no-no* and *no-having*. These strategies, with examples, are listed in Table 1. For the sake of clarity, I have also mapped the 10 most common strategies by the locations in which they are spoken, with bullseyes representing languages that use less common strategies, in the map in Figure 1.

Table 1: An Overly Specific Classification System

Number	Naming Strategy	Example	Map Label
26	unanalyzable clan name	*Warrwa*	A
12	place	*Paakantyi*	B
10	good language	*Gugu-Mini*	C
8	no language	*Djadjawurrung*	D
8	compass	*Gunggari*	E
7	X language	*Mayi-Thakurti*	F
6	this	*Dhangu*	G
6	no reduplication	*Yorta-Yorta*	H
5	man/people	*Kaurna*	I
5	language	*Dharuk*	J
4	no-having	*Wiradjuri*	
4	no X	*Wardandi*	
4	unanalyzable redup	*Nyulnyul*	
4	this X	*Ngunawal*	
4	species	*Nimanburru*	
3	this language	*Mayi-Kulan*	
3	no	*Janday*	
3	go having	*Kartujarra*	
2	this having	*Ngaanyatjarra*	
2	person's language	*Yinhawangka*	
1	X having	*Ngaliwurru*	
1	unanalyzable stranger	*Pintupi Lurritja*	
1	this mouth language	*Kuuk Thayorre*	
1	this having language	*Guugu-Yimidhirr*	
1	strong customs	*Lardil*	
1	place men	*Gunditjmara*	
1	place language	*Uw Oykangand*	
1	no reduplication language	*Dhudhuroa*	
1	more X	*Nyangumarta*	
1	meat having	*Kukatja*	
1	like X	*Bigambal*	
1	go language	*Dalabon*	
1	fire language	*Kalaamaya*	
1	compass place language	*Kala Lagaw Ya*	

Figure 1: Map of Naming Strategies

Having 34 different categories, especially with so many singletons, means that this approach is therefore too specific to be particularly useful for phylogenetic analysis. This challenge is also reflected in the map, where in the southeast, for example, there is a cluster of two different strategies, labelled D and H. It is not immediately obvious under this system that the name of each language in the cluster in fact has something in common: the use of *no*. Clearly, there should be some way of capturing this similarity.

On the other hand, there are perfectly accurate ways of describing language names that are so general as to be equally as uninformative. Many such possibilities exist, such as the binary approach of dividing the languages into those whose names describe the language itself and those whose names describe the speakers. This results in the data in Table 2, which, while accurate, doesn't tell us very much.

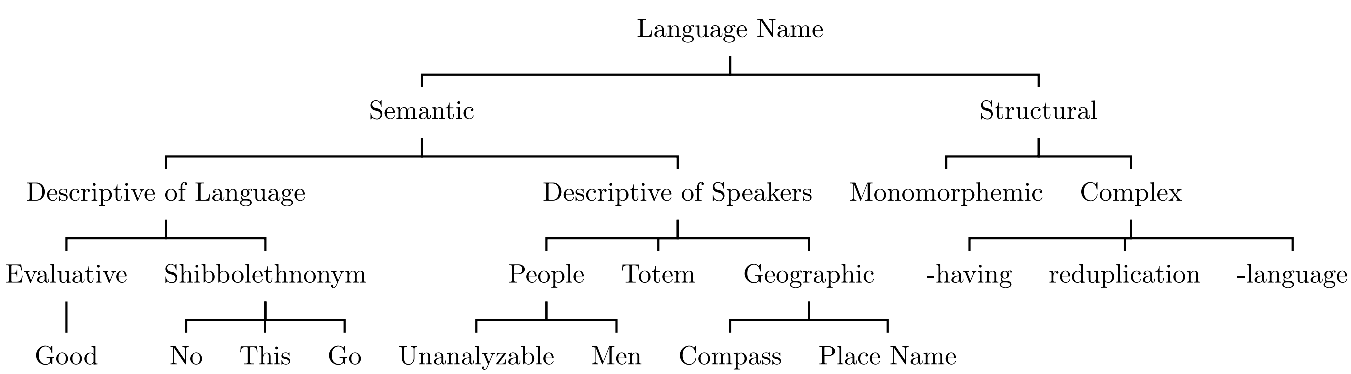
Table 2: An Overly General Classification System

Number	Naming Strategy
74	language
56	speakers

Because of the complicated patterns in the data, therefore, the biggest issue in categorizing the various names of Pama-Nyungan languages is finding a system that is neither too specific nor too general, but rather one that is historically meaningful.

In order to be able to capture both high-level similarities and low-level differences, I propose the hierarchal classification system in Figure 2 to describe Pama-Nyungan language names. Thus a language name like *this-having* and one like *no-having* will have in common the fact that they are shibbolethnonyms, and that they feature *-having*; however, they will also be classified as different types of shibbolethnonyms because the relevant shibboleth is different. At different levels of specificity, therefore, these employ either the same type of naming strategy or different ones. While this seems very complicated, it allows for both generalizations and fine-grained distinctions to be drawn within the same system.

Figure 2: Hierarchical Classification of Naming Strategies



2.1.4 Explaining the Hierarchy

Under this hierarchy, every language name can be analyzed in terms of both its semantic and its structural content. The difference between a language whose name means 'no-no' and one that means 'no-having' is now very simply the difference of one feature: the two languages have the same semantic features (*Descriptive of Language, Shibbolethnonym, No*) and both are structurally complex, but one is built by reduplication, the other by adding a comitative or proprietive morpheme, which I have grouped together under the label "-having."

While neither McConvell nor Sutton's descriptions of language names were usable wholesale, my hierarchy is based in large part on their analyses, as well as on Wohlgemuth's principles. None of them focused much on the structural differences in language names, so that part of the classification is entirely original; within the semantic domain, however, the division between *Descriptive of Language* and *Descriptive of Speakers* is based on Sutton's approach to differentiating languages.[1] I have collapsed Sutton's distinction between "referring to mythical stories," "referring to human characteristics," and "referring to locations" into subcategories of the single category *Descriptive of Speakers.* This is motivated by the fact that even the names that seem to be describing a place are describing the location of the people who live there, and are therefore describing the people, while names "referring to dialect characteristics" are purely descriptive of the language itself.

Within the category *Descriptive of Speakers* I have established three semantic categories: *People* (which includes both unanalyzable clan names and names based on the word for 'men'), *Totem* (names that refer to items in the natural world that are associated with the speakers) and *Geographic*, which is further divided into names built on directional terms and those built on place names.

Within the category *Descriptive of Language* I have created a two-way distinction. *Evaluative* names like "good language" describe the quality of the language and the speaker's attitudes toward it.[2] The other subcategory is *Shibbolethnonym*, a concept that I have borrowed from McConvell. These are names that are descriptive of the content of the language itself: they proclaim what lexeme that language uses to express a specific meaning, presumably to contrast their own language with a similar one that does not use the same word. The specific shibboleths in this hierarchy are *This*, *No,* and *Go*.

On the structural side, languages can be either monomorphemic or complex. If they are complex, there are three possible categories that they can fall under:

[1]While it is tempting to refer to this distinction as being between glossonyms and ethnonyms, I have refrained from doing so for two reasons. The first is that whatever their etymology, these names are almost all used as the name of both the language and the group of people who speak it, and are therefore both glossonyms and ethnonyms. Secondly, the category *Descriptive of Speakers* includes names that describe the location where the speakers live, and are therefore originally toponyms, not ethnonyms. I therefore prefer to use these less technical terms to avoid any confusion.

[2]Wohlgemuth would classify these as "esteem" names that refer to the people, not the language. Because the adjective always co-occurs with *language* in my data, I follow Sutton in considering them to be value judgments about the languages themselves.

Reduplication, meaning a morpheme is repeated, *Language*, meaning that the language name contains a component meaning 'language' or 'speech,' and *Having*, which I use as an umbrella term covering names with either a comitative or proprietive affix.

2.1.5 Important Notes for Understanding the Hierarchy

Feature Exclusivity Every language name has both a structural and a semantic component; it is not possible for a word to have semantic content but no structural content, or structural content without even the semantic "meaning" of 'unanalyzable.' Within either of these two main branches, however, the presence of one possibility seems to prevent any other possibilities on the same tier. Thus 'this language' and 'good language' are both possible, but 'this good language' is not, because it contains two different semantic features. There are two exceptions to this feature exclusivity on the structural side: *Dhudhuroa* means 'no-no language,' and therefore falls under both *-language* and *reduplication*, and *Guugu-Yimidhirr* means 'this-having language'. The only potential exception on the semantic side is *Kala Lagaw Ya*, which means 'west island language,' and might fall under both *compass* and *place name*; however, it seems likely that *Kala Lagaw* is a complex place name meaning 'west island' and therefore only one semantic strategy is being used.

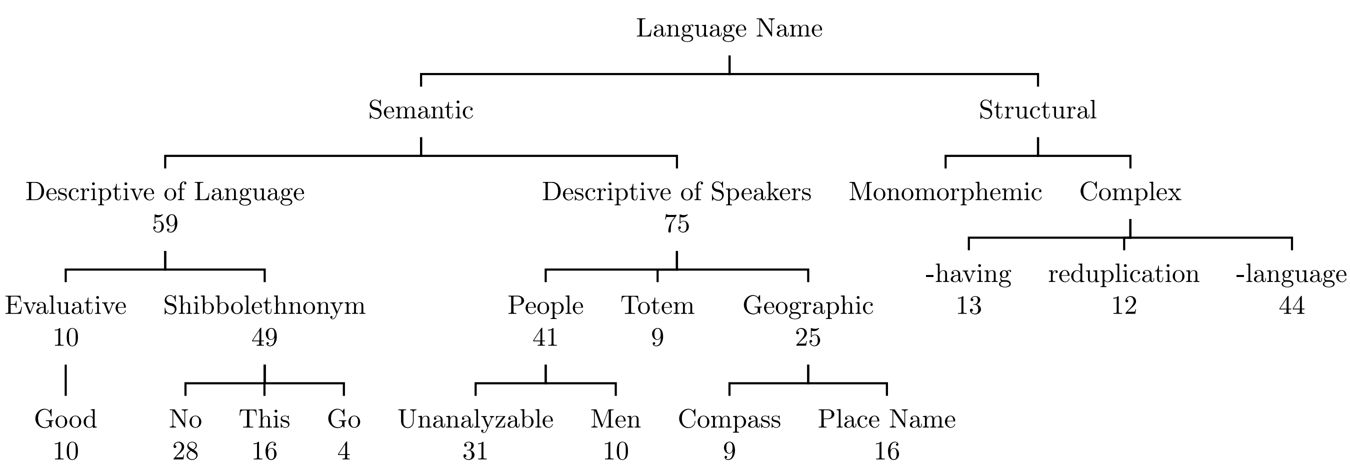
Non-Exhaustivity Because this is a hierarchy, the presence of a semantic feature on a lower tier necessitates that any feature that is a direct ancestor of that feature must also be present. Thus, because *this* is a subcategory of *shibbolethnonym*, a language name containing *this* must necessarily be a shibbolethnonym, and therefore also fall under *Descriptive of the Language.* This relationship does not, however, go in both directions. It is possible for a language to be a shibbolethnonym, but not have *no*, *go*, or *this.* This means that it is possible for a language name not to have a value on the very lowest tier, because, while the three types of shibbolethnonyms listed in the above hierarchy are the most common, there is no reason to assume that it is not possible for another word to become a shibboleth if that is the only salient word that is not shared between two closely related languages. In my dataset, there is exactly one other shibbolethnonym: *Kukatja* means 'meat-having' and is understood to highlight a contrast in vocabulary, rather than a cultural description about meat-eating practices (Claire Bowern p.c.).

Theoretically there are infinitely many possible projections on this hierarchical structure; I included only the ones that occur frequently in my data, but if many language names were found to be built from the word for 'meat,' that would motivate adding such a possibility to the subset of shibbolethnonyms. This helps solve the problem of McConvell and Sutton's analyses not applying to languages that they hadn't seen yet; any new language can be placed somewhere on the hierachy. It will be either monomorphemic or complex, and describe either the language or the speakers; even if it is not possible to classify a new language farther than this point, that will be an accurate way of describing that this language name has little in common with others.

2.2 Using the New Classification

Now that we have a classification system, it is possible to begin looking at how the naming patterns are represented across Pama-Nyungan languages. It is possible to come up with tallies for each level of the hierarchy, which I have included in Figure 3 below.[3] It should be noted that given the nature of the data at hand, these numbers are necessarily minima: a language like Mayi-Yapi, for example, whose name is a compound of the word for 'language' and another unanalyzable component, can only reliably be counted in the *language* category. It is possible that knowing the meaning of *Yapi* would allow Mayi-Yapi to be classified as a shibbolethnonym, or a totem-based language, etc.[4]

Figure 3: Totals for Individual Naming Strategies



The most basic semantic difference among language names according to the hierarchy is whether they describe a fact about the language itself (that it is a good language, or has a specific word) or describe the people who speak it, including the location in which those people live. Language names that describe the speakers of the language are somewhat more common in this set of data than those that describe languages, with 90 names that describe speakers and 59 names that describe languages. While speaker-descriptive names are more common, it is still interesting that so many names describe languages, given

[3]The one exception is the category *monomorphemic*, which, for reasons that I will discuss on page 15, cannot be tallied.

[4]It is possible, for example, that *yapi* is related to the widespread word *yapa*, which has a variety of meanings in different languages, including 'person' and 'sister' (Bowern p.c) so perhaps it falls under the 'People' category.

that most Pama-Nyungan language names are also ethnonyms; clearly language is a significant part of group identity.

Of the 59 language-descriptive names, 49 are shibbolethnonyms, based on a word that is representative of the language. The majority of these (28) contain the word for 'no,' 16 have the word for 'this,' and 4 are built on 'go.' There is one shibbolethnonym which is not built on any of these three words; this remaining shibbolethnonym is Kukatja, which means 'meat-having.'

The other 10 names that describe languages are all evaluative terms that describe themselves as 'good language.'

Of the names that are in the broad category of "descriptive of speakers," 57 describe the people themselves, of which 9 refer to 'men' or 'people' and the remaining 48 are (partially or fully) unanalyzable. 9 more languages have names that refer to a totem relating to the group of people who speak that language; this can be a species, like Nimanburru, which appears to be based on the word for 'flying fox,'[5] or another salient feature, like Adnyamathana, whose name means 'stone group' (Sutton 1979: 92). The remaining 24 languages that are descriptive of speakers refer to the area where the language is spoken; 16 are based on names for locations, while the other 9 feature a compass-point term like 'north' or 'west.'

On the structural side, it is unfortunately not possible to reach a definitive conclusion about how many names are monomorphemic or morphologically complex, because it is not always clear from the descriptions in grammars whether a name is a bare noun, a stem with an inflectional morpheme, or a larger noun-phrase. In a world with perfect data, it would in fact be possible to count the number of monomorphemic names. Even with imperfect data, however, it is possible to analyze the 72 languages that clearly contain more than one morpheme. More than half of these names (44) are noun phrases consisting of 'language' and one other morpheme. 13 fit the pattern 'X-having,' while 12 consist of a single morpheme reduplicated.

2.2.1 Overlapping Categories

It is also important to look at which semantic features co-occur with which structural features. Of the 12 language names that feature reduplication, for example, all but 4 are the result of reduplicating a negative morpheme; in the remaining 4 cases, the reduplicated morpheme is unanalyzable. It is tempting to draw the conclusion that the only type of reduplication in Pama-Nyungan language names is negative reduplication, and that the unanalyzable reduplicated names are based on archaic negative forms; this is not provable at the moment, but would be an interesting topic for further investigation.

Turning to the 'having' names, 12 of the 13 are clear shibbolethnonyms: 5 mean 'no-having,' 3 'this-having,' 3 'go-having,' and 1 'meat-having.' While the shibboleths can all co-occur with 'having,' there are no ethnonyms with a name

[5]N.B. Claire Bowern (p.c.) notes that while the name *Nimanburru* is phonologically identical to that language's term for 'flying fox,' it seems unexpected for it to be a totemic name, given that none of the other languages in the region have totemic names.

meaning 'good-having.' This is part of the reason that I treat ethnonyms with 'good' as evaluative terms rather than as shibbolethnonyms based on contrasting terms for 'good'; every instance of 'good' co-occurs with 'language,' and none are reduplicated or appear with 'having.' There are also no names that clearly describe speakers and fall into the 'having' category: the remaining name has an unanalyzable morpheme as its other component. That the vast majority of 'having' names are shibbolethnonyms is unsurprising given the semantic content of this structure, and suggests that *Ngaliwurru*, the language whose name means '[unanalyzable]-having,' might also be a shibbolethnonym.

The 'language' category is, perhaps unsurprisingly, more varied. Of the 44 names in this category, 5 simply mean 'language' and 8 consist of a morpheme meaning 'language' and another unanalyzable morpheme. The remaining 31 languages consist of 23 names that describe the languages and 8 that describe the speakers. There are 2 that consist of 'language' and a compass point term, 2 that consist of 'language' and a place name, 2 that feature a totem, and 2 that mean 'people's language.' Of the 23 that describe languages, 10 mean 'good language' and the other 13 are shibbolethnonyms: 4 have 'this,' 8 have 'no,' and 1 has 'go.'

The overlap between each semantic and structural feature is shown below in Figure 4.

2.2.2 Classifying by Subfamily

If we believe that ethnonyms have evolved through the family, there should be some evidence of patterns within subgroups. For many of the subgroups of Pama-Nyungan, I do not have enough data to come to any conclusions, but in other cases, there are very interesting patterns to note.

For the following subgroups, I only have data on one language, and therefore can clearly not make any assumptions about the subgroup as a whole: Arandic, Bunuban, Dhuduroa, Dyirbalic, Gippsland, Kalkatungic, Kartu, Macro NSW, Marrngu, Mirniny, Nyawaygic, Pilbara, Western Torres, Yotayotic.

There are also subgroups with only two members represented here. In Durubulic, both language names are built on the word for 'no.' Similarly, both Kanyara-Mantharta languages contain 'language,' but *Thalanyji* just means 'language' while *Warriyangga* is a shibbolethnonym meaning 'no-language.' These similarities are perhaps suggestive, but clearly do not provide enough data to generalize.

The Bandjalangic subgroup has five members in this dataset, but one is Upper Clarence River, whose name comes from English and is therefore discounted. The remaining four language names are evenly split between two that are descriptive of the language and two that are descriptive of the speakers.

In the Central NSW subgroup, one name is unanalyzable, while the other four are descriptive of the language. Of those that describe the language, two names mean 'no-having,' one simply means 'language' and the other means 'like [unanalyzable],' which seems to be describing the language as similar to something.

Figure 4: Tables of Semantic Features by Structure

(1)

-having	
this-having	3
no-having	5
go-having	3
people-having	1
Total shibboleth-having	12
Total speaker-having	1
Total	13

(2)

-language	
good-language	10
this-language	4
no-language	8
go-language	1
Total shibboleth-language	13
Total language-descriptive-language	23
men-language	2
totem-language	2
compass-language	2
place-language	2
Total geographic-language	4
Total speaker-language	8
[unanalyzable]-language	8
language	5
Total	44

(3)

Reduplication	
no	8
unanalyzable	4
Total	12

Karnic has 12 representatives in this sample. Two of the Karnic names are based on compass terms, four mean 'good language,' and the remaining six are unanalyzable, including two unanalyzable reduplicative names.

Of the 9 members of Kulin, all but one are *no*-based shibbolethnonyms: six mean 'no-language,' while two consist of 'no' reduplicated. The final language name is Boonwurrung, meaning '[unanalyzable] language'; given the rest of the family, it seems likely that this too was originally a shibbolethnonym.

Lower Murray has five members, one of which is an English-based name; one name means 'human,' one is based on a place name, and two are cases of 'no'-reduplication.

Maric has five geographic names: three compass-based and two place-based. There are also three language-descriptive names, of which one means 'good language,' and two mean 'no [unanalyzable].'

The Mayi languages all have the word for 'language' in their name, followed by another word; in two cases this second component is unanalyzable, while in the other two it is 'this.'

There is no clear pattern in Ngayarta: one is an unanalyzable clan name, one simply means 'language,' and one is based on a place name.

Similarly, in the Ngumpin-Yapa subgroup, three names are unanalyzable, one means 'language,' and one describes a place.

Nyungar has more speaker-descriptive names than language-descriptive ones. There are two totem-based names, one that means 'man,' and one place name. The other two languages are both shibbolethnonyms composed of 'no' and an unanalyzable morpheme.

Of the 16 Paman languages for which I have data, two have English names and are therefore irrelevant to this analysis. Three names are based on place names, one is unanalyzable, and three mean '[unanalyzable] language.' The remaining seven are descriptive of the language itself: there are two 'good languages,' and five *this*-based shibbolethnonyms. While there is a lot of variance in the semantic features, there is some structural similarity in that seven of the fourteen languages have 'language' in their name.

The names of the three Thura-Yura languages are all descriptive of their speakers. Two are totemic names, while one means 'people.'

Waka-Kabi is split between language-descriptive names and speaker-descriptive ones. Two are *no*-based shibbolethnonyms (one featuring 'language,' the other reduplication) while two refer to locations.

The Wati language names are predominantly descriptive of the language. Of the eight languages, five are shibbolethnonyms, all with 'having' in the name: three are based on 'go,' one on 'meat,' and one on 'this.' There is also one name that means 'good language' and one that means 'people's language,' which is the only speaker-descriptive name in the group. Finally, there is one unanalyzable Wati name.

Four of the five Western Nyulnyulan languages are unanalyzable, with two of those four being reduplicative. The final name is based on a totem, and so all the Western Nyulnyulan languages seem to be descriptive of speakers rather than the language.

No clear pattern can be seen in Yardli. One name means 'good language,' one is based on a totem and the third means '[unanalyzable]-language.'

Yolŋu has a very clear pattern of names: six are unanalyzable clan names, six mean 'this,' and one means 'this language.'

Of the six Yuin-Kuri languages, each has a name that means something different; more of these names describe the speakers than the language, but very little connection can be seen among the names.

3 Phylogenetic Analysis of the Evolution of Language Names

With this classification system, it is now possible to explore the distribution and development of various types of language names. The levels of both similarity and variance among related languages suggests that this might be able to be addressed as the result of phylogenetic evolution.

3.1 Cultural Phylogenetics

The question at the heart of cultural phylogenetics is whether cultural evolution is fundamentally the same as biological evolution. There are clearly similarities: languages, like species, are forever changing and developing different varieties, the relationships among which can be mapped with trees just like biological species. Even Charles Darwin himself noted the similarities between linguistic and biological evolution stating that there was a "curious parallel" (1871, quoted in Gray et al. 2007: 361) between the formation of new languages and of new species. The similarities are striking enough that many people began to question whether these processes were in fact the same. While the idea of cultural evolution is nothing new, this question has gained a particular strength in recent years as new computationally powerful methods have been created to analyze biological evolution; if linguistic change is evolution, could these tools also be applied to aspects of language (and other parts of culture)?

3.1.1 Darwinian Evolution

In order to answer this question, it is necessary to determine whether linguistic change meets the standard of evolution. In his 2011 book *Cultural Evolution: How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*, Alex Mesoudi addresses each of the tenets of Darwinian evolution, and shows how those can apply to culture.

According to Mesoudi, the three preconditions for Darwinian evolution are *variation*, *competition*, and *inheritance* (2011: 26-34). Without *variation* within a species, every individual would be identical, leaving nothing for natural selection to act on. This selection happens via *competition*, both between members of a single species and across different species: as individuals struggle for existence, whether against each other or simply against the environment, the only variants of a trait that will endure are the ones manifested by the individuals that survive. These variants are then *inherited*, which is key to evolution; if traits are not heritable, then beneficial traits will not be passed down more than others, and evolution will not occur because the same cycle will be repeated in each generation.

Mesoudi shows that each of these preconditions are met by change in language and other aspects of culture. There are, for example, thousands of languages, each of which has variation in the form of dialects, sociolects, idiolects, etc. That these variants compete on a high level is evidenced by the rate at which languages go extinct: when two languages are spoken in the same region, one often survives at the expense of the other. Within languages, too, there is competition: Mesoudi's example is the process by which irregular verb forms are replaced by regular ones. Finally, language and culture are clearly inherited from one generation to the next, largely the same, but with small variations. This cultural transmission is also evident in the progression of innovations, such as the long progress by which the invention of a written numeral system led to a place value system, which led to the written symbol for zero, and so forth, until finally calculus was invented. Likewise, languages change over time, but only gradually, so that each successive generation understands the last, but the language eventually becomes completely different from its distant ancestor.

Based on these similarities, Mesoudi argues, cultural evolution (including linguistic evolution) is indeed Darwinian, and it should therefore be possible to study it like biological evolution, using the same formal models (2011: 54).

3.1.2 Formal Phylogenetic Analyses

Just as with trees of languages, people have been constructing trees of biological species since the 19th century. These trees are known as *phylogenies*, and so the construction and analysis of such trees is called *phylogenetics* (Mesoudi 2011: 87). The essential concept of phylogenetics is very similar to the comparative method in historical linguistics: by comparing the similarities and differences among current species/languages, it is possible to both construct a tree that shows the relationships and history of the family, and to attempt to reconstruct how the trait (or phoneme, etc.) manifested in an ancestral language or species.

The difference between this and the traditional comparative method is that, rather than an individual painstakingly comparing the manifestation of each trait across species, phylogeneticists use formal computer models. This allows them to analyze a much larger amount of data in a much shorter time period, to test many more possibilities, and to quantify the (un)certainty of their results. Given that language evolves in a Darwinian manner, and these analyses are already predicated on such similar ideas, researchers have begun to use these powerful phylogenetic methods to analyze the evolution of language (Nunn 2011: 39-41).

Basic Phylogenetic Terminology and their Linguistic Correlates The basic concept of a phylogeny is identical to that of a linguistic tree, although some of the terminology is different. The living species at the *tips* (also called *taxa*) of the tree are all descended from the *root*, the now-extinct ancestor of all the species in the tree. Each internal *node* within the tree is also considered to be the extinct ancestor of any node or tip that it dominates, and descended from every node that dominates it. (Nunn 2011: 21-22). For languages, this means that the root represents the proto-language from which the daughter languages at the tips developed, with the interim stages reflected in the internal nodes.

The data points to be compared, such as the cognates or phonemes typically used in the comparative method, or the DNA used in biological analyses, are called *traits* or *characters*. Just as in the comparative method, it is necessary to construct a tree before analyzing how different traits developed. Just as, once the relationship among languages has been established, it is possible to reconstruct a proto-language, so too once a phylogeny is constructed by comparing the states of various characters across each species/language, can a researcher then perform *ancestral state reconstruction*. By comparing the current distribution of a given trait to the historical tree, it is possible to reconstruct the value of that trait at the root and any interior nodes, to determine how that trait developed.

While there are various methods for performing such phylogenetic analyses, Bayesian methods are currently very popular, in which, given a current trait distribution, the analysis measures the likelihood that a specific phylogeny (when constructing a tree) or trait history (for ancestral state reconstruction) is correct. This approach is often implemented using a Markov Chain Monte Carlo; in essence, these programs work by constructing trees of relationships among the relevant languages and determining how likely the distribution of cognates is to have evolved based on the relationships in a given tree. They then compare this to the likelihood of the distribution with a slightly altered tree and accept whichever tree is more likely. After doing this millions of times, this results in a sample distribution of trees that can be used to capture uncertainty: a subgroup that occurs in every sample tree has a higher probability of being accurate than a grouping that only occurs in 50%. The process for ancestral state reconstruction is very similar, but rather than selecting possible trees, it runs through possible trait histories and transition rates (Nunn 2011: 35-36, 71-75), resulting in a set of probabilities for each possible ancestral state of each character.

Linguistics researchers have used these Bayesian phylogenetic methods to, among other things, construct dated phylogenies of language families (Gray et al. 2009; Bouckaert et al. 2012) and trace the development of various constructions in language families (Haynie & Bowern 2016).

3.1.3 Phylogenetic Signal

Culture is clearly not entirely the result of evolution. Languages can, for example, borrow words from other languages, and frequently do so. It is not always easy to tell whether what appears to be the result of inheritance might actually be the result of languages borrowing from each other. One way of dealing with this is by measuring the amount of "phylogenetic signal" there is in a set of data; this term was originally used by Blomberg and Garland to mean the "tendency for related species to resemble each other more than the resemble species drawn at random from the tree" (2002: 905). In other words, the amount of signal reflects the degree to which the distribution of traits is consistent with evolution along a tree. There are many ways to test for phylogenetic signal, some of which will be discussed at greater detail later, but in general, high degree of signal will make it possible to reconstruct past states of a trait, while low levels of signal make it much less likely (Nunn 2011: 109). Thus a trait with a high level of borrowing will exhibit less phylogenetic signal than a more stable trait like core vocabulary terms.

3.2 Are Language Names Phylogenetic?

Language names seem to fit Mesoudi's three criteria for Darwinian evolution. There is *variation* in the strategies used to refer to languages, both within and across languages; these varieties necessarily *compete* to be the dominant way of referring to the language; and the names are *inherited* from one generation to the next.

Despite this, however, it is not clear that names evolve in the same way that other aspects of language do. Language names are much more on the level of consciousness than aspects like phonemic change, and they make much more of a political statement. Recently, for example, an 88 person council voted on whether to replace the name *Navajo* with *Diné*, meaning 'people' (Moore 2017: 11), in order to assert their identity.

Such changes are deliberate and sudden, designed to make a statement in a way that typically does not happen in evolution. Naturally, naming decisions that occurred in 20th century North America are different from those that occurred in precolonial Australia, but the fact remains that the decision of what to name a language is inherently political and meaningful, and may potentially be deliberately thought out in a way that phonemic changes tend not to be.

Thus, there are certainly factors influencing the development of language names that do not occur with all aspects of language. Perhaps name changes do not happen purely organically, as they are an attempt by a group of people not only to assert their own linguistic identity, but to very deliberately separate their language from other languages. But just because there is some intentionality behind naming does not necessarily mean that such change is not phylogenetic. This level of intention is not automatically exclusive of evolution; this sort of development seems to be somewhat akin to selective breeding, by which the randomness of natural evolution is manipulated by intentional processes. If phylogenetic analyses can be done on other aspects of culture like weaving (Buckley 2012) that are both consciously constructed and highly diffusible, then conscious impositions should not necessarily invalidate phylogenetic processes.

Language names certainly contain both variety and similarity. Languages within the same subgroup often have names with similar meanings, with patterns occurring that suggest some sort of family inheritance. Textile weaving patterns, like language names, are certainly at the level of consciousness, and while some of the change in weaving patterns is unintentional, some of it is definitely consciously implemented. What is striking about the variance in names is that, even though the names might have been consciously chosen, they seem to be chosen from a limited set of possibilities; almost every language in my data set has a name that fits neatly into my small hierarchy of names. Instead of languages all having vastly unrelated names, as might be expected, these names all seem to have developed from the same small set of traits, for which there are clear familial patterns. It is for this reason that it is worth investigating just how much of the distribution of naming strategies is due to evolution within the family.

3.3 Choosing the Analyses

Because the hierarchical structure of strategies so complicated, it is not feasible to attempt to reconstruct past names wholesale; instead, I focused on multiple smaller categories representing divisions within the hierarchy. I chose the levels to analyze based on both theoretical considerations and the distribution of the traits across the family. The levels I chose to analyze, and the relevant trees of trait distribution, are listed in Figures 5-10[6]:

[6] In the trees where there is a black category coded as '-,' this is an "other" category. This allows us to look at not only how a trait varies, but how common that trait is across the family. In the language-descriptive tree in Figure 8, for example, roughly half the dots are black, showing that they are not language-descriptive at all

Figure 5: Structural

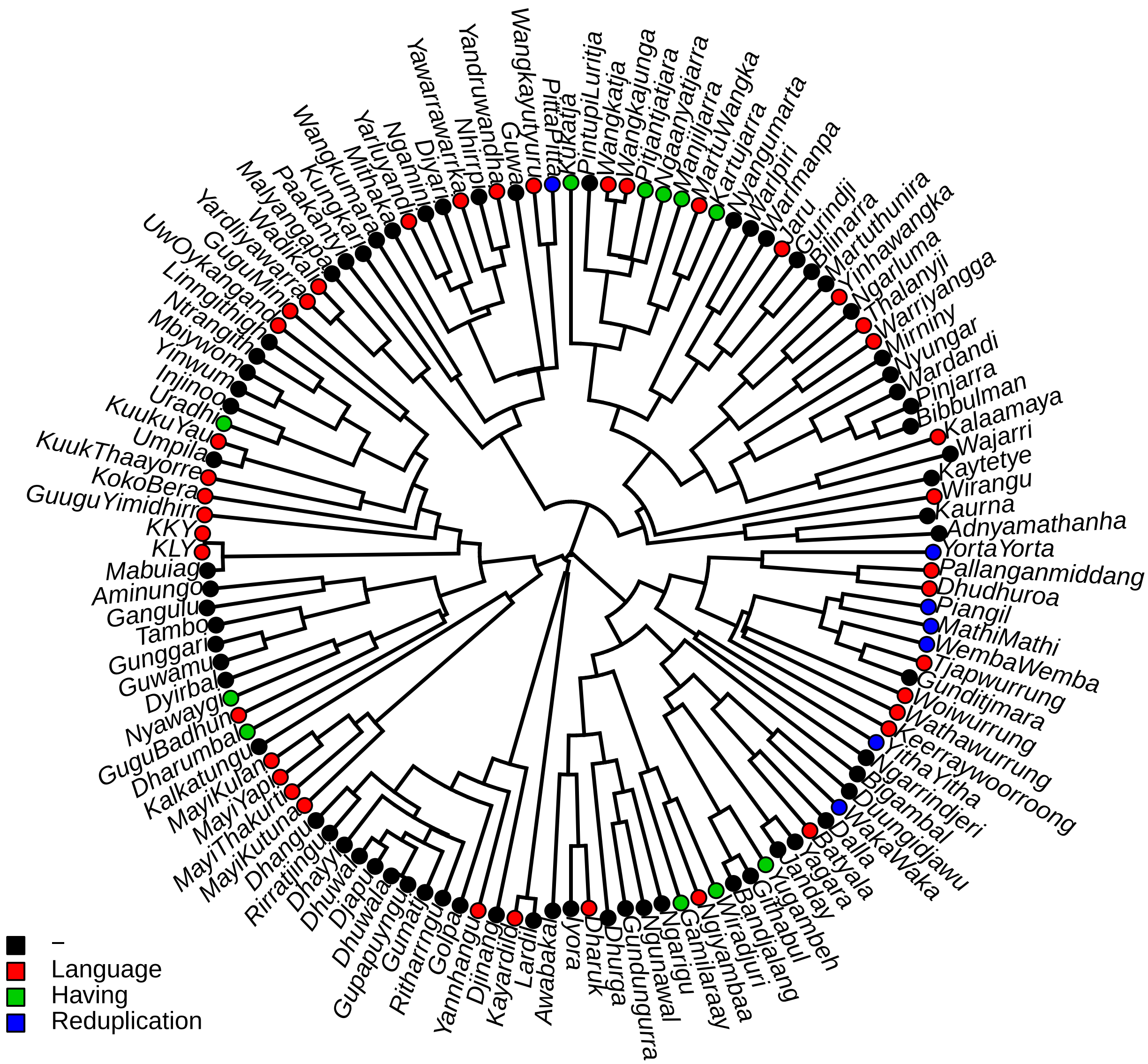


Figure 6: Describing Language or Speakers

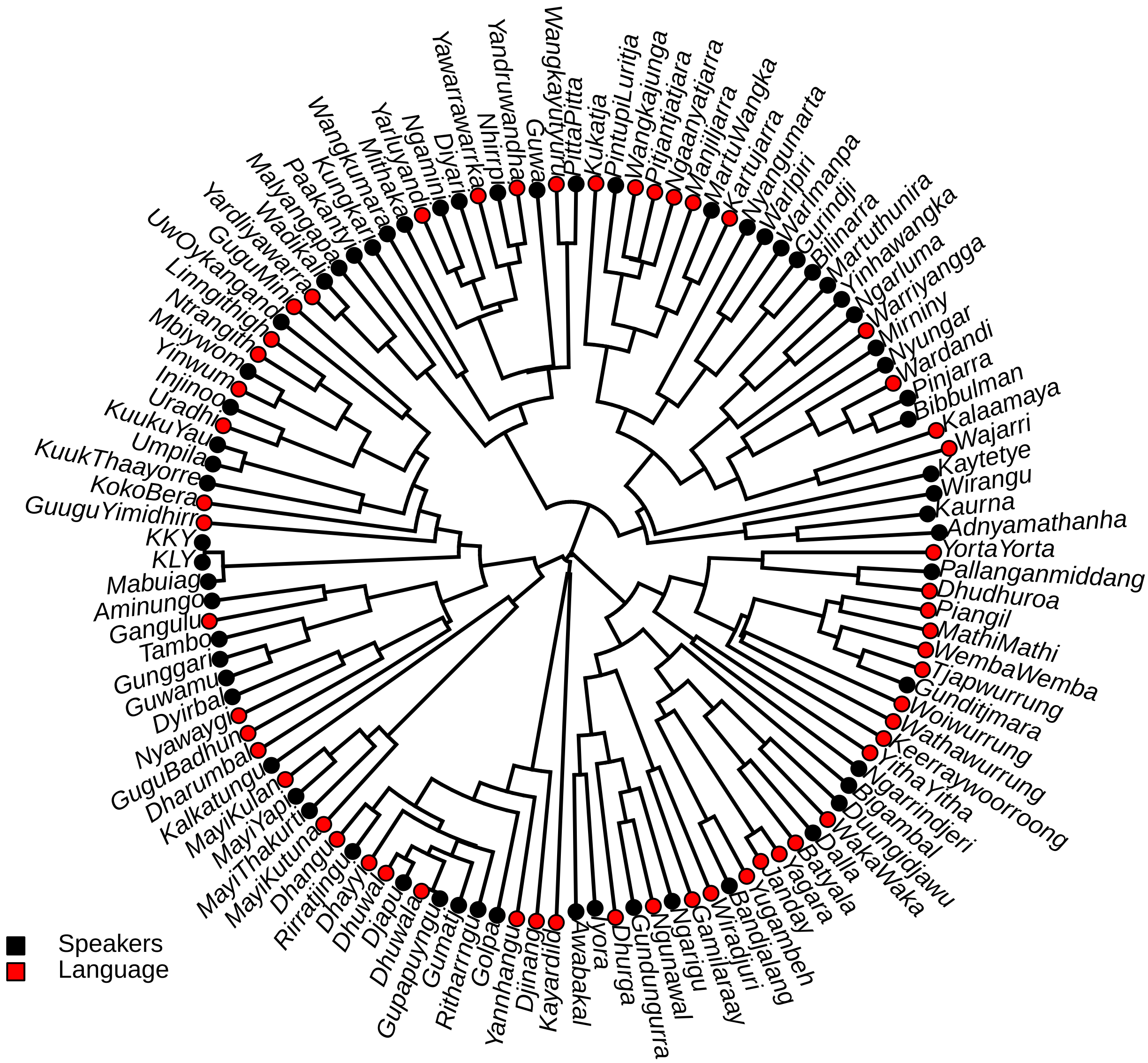


Figure 7: Shibbolethnonym or Not

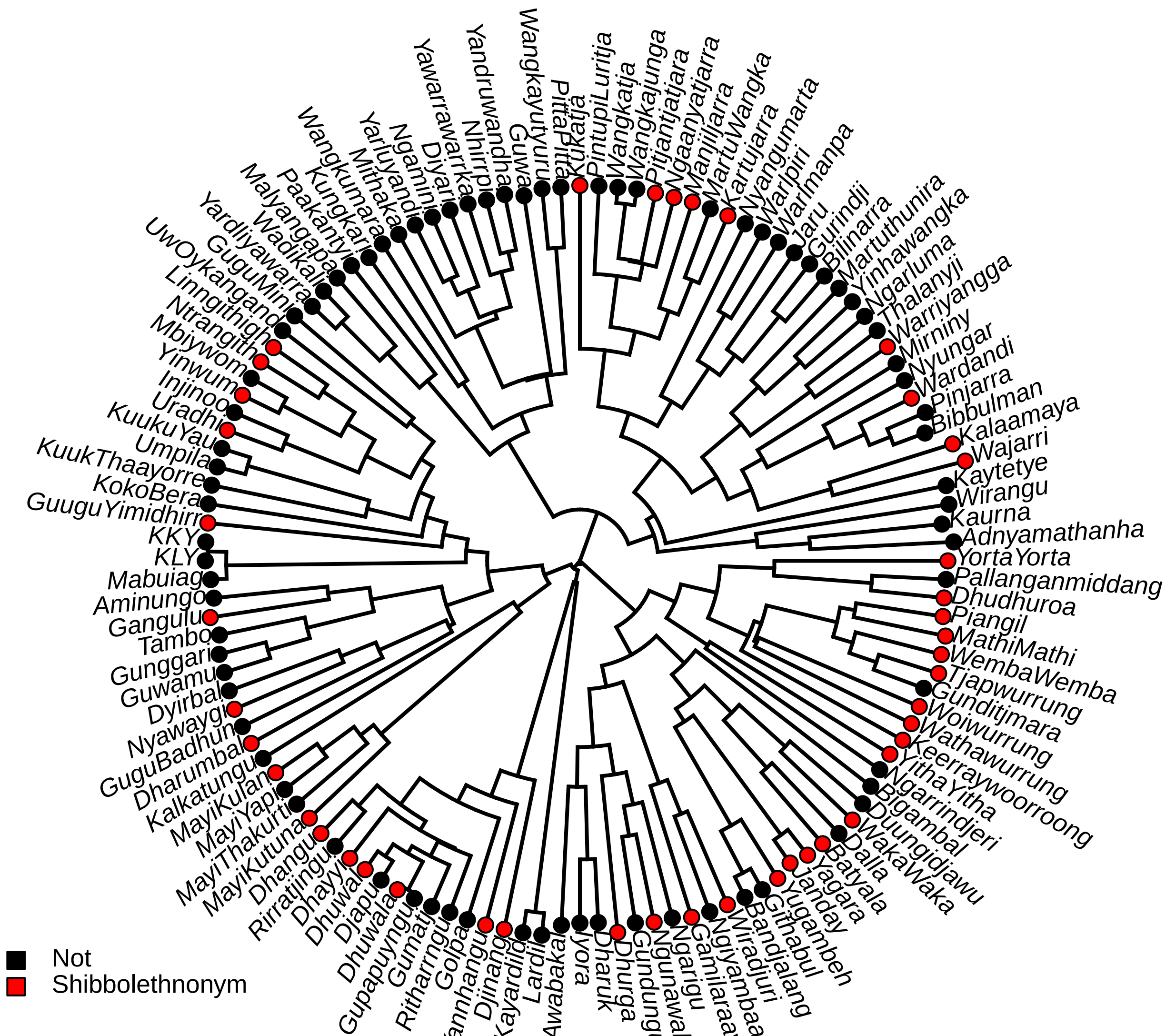


Figure 8: Language Descriptive: Evaluative or Shibbolethnonym

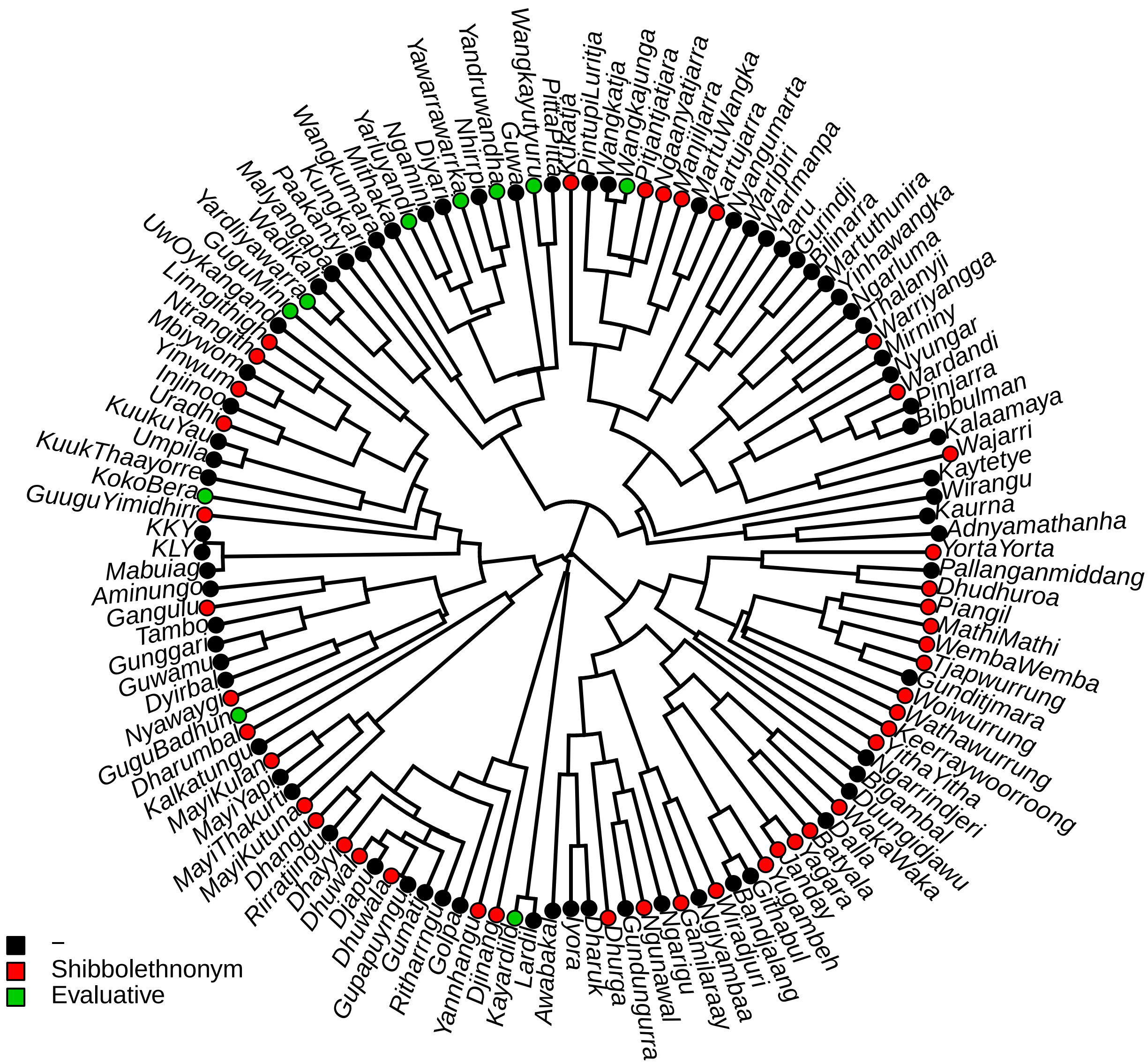


Figure 9: Type of Shibbolethnonym

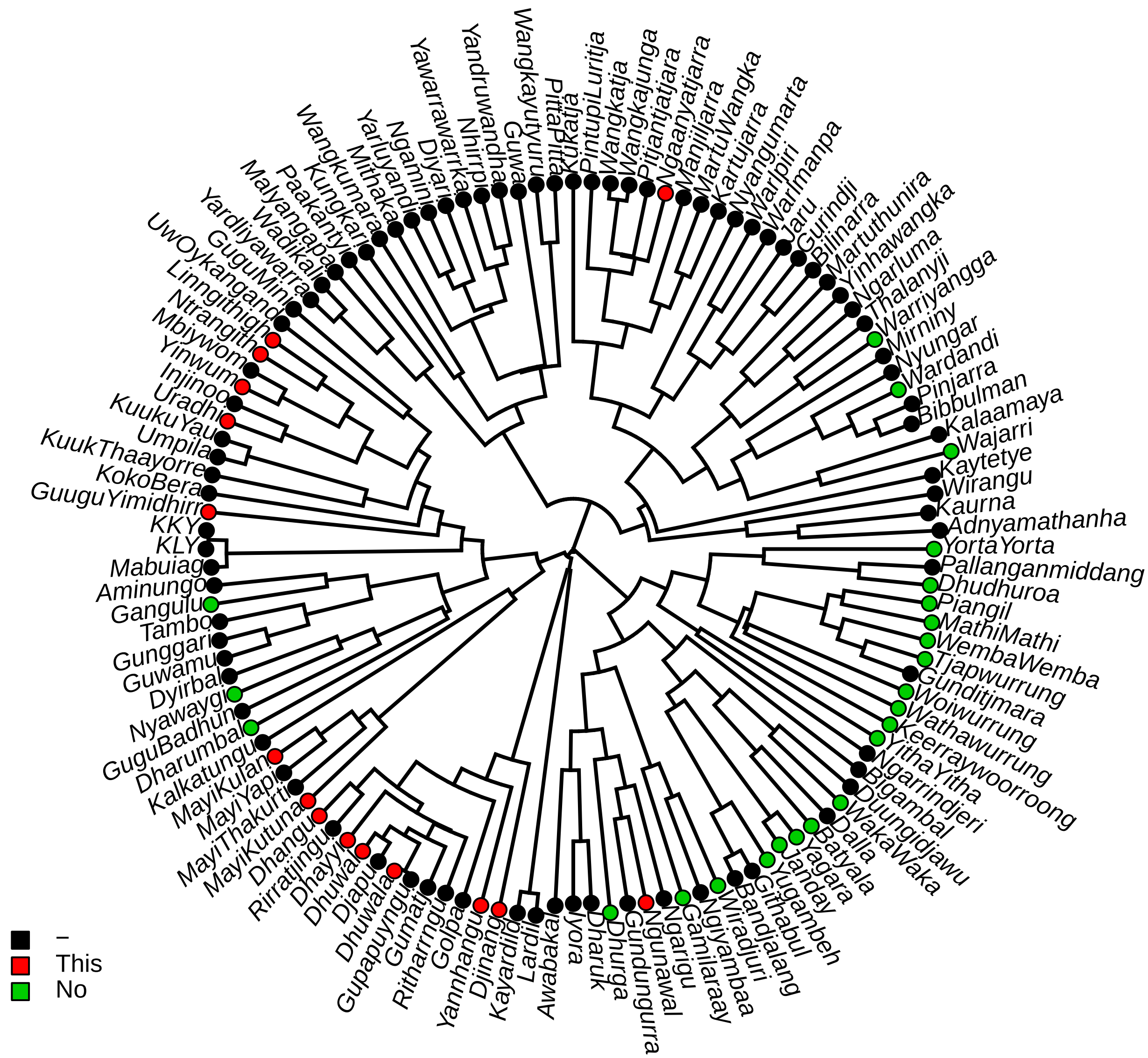
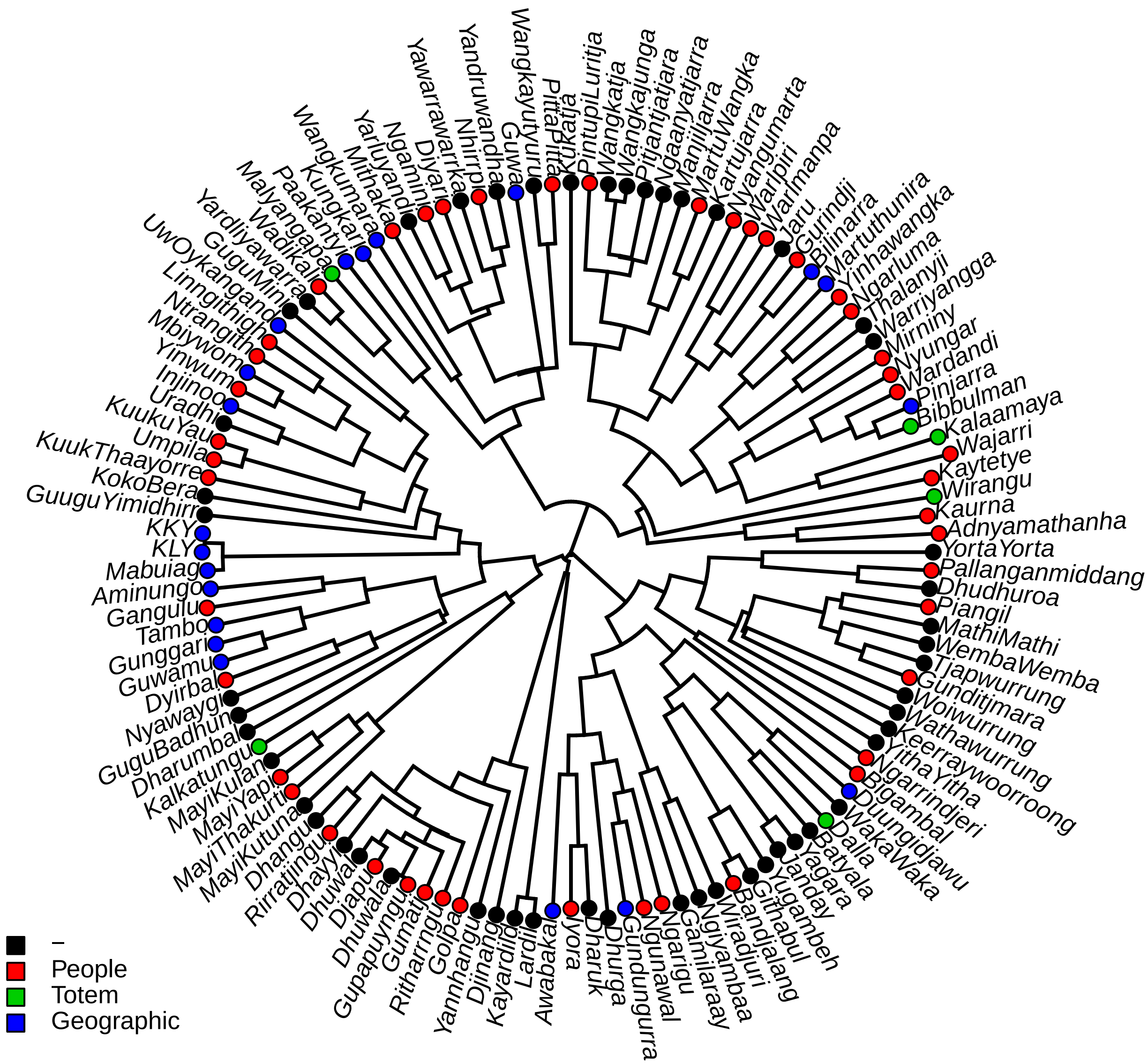


Figure 10: Speaker Descriptive Methods



I chose these analyses for a number of reasons. I intentionally chose both the highest level of semantic analysis (describing language vor speakers) and the lowest level (*this* vs. *no*)[7] in order to investigate whether there was a difference in the level of phylogenetic signal at different levels of the hierarchy. I predicted that the lowest level would have the least amount of signal, not only because it is the most specific, but also because the decision of what shibboleth to use as the basis of a name is highly dependent on the vocabulary of both the language being named and the languages it is being distinguished from. A language can only use 'no' as a shibboleth if the nearby languages have a substantially different word for 'no'; otherwise it will not provide the desired contrast. Thus, I expected that, while the use of shibbolethnonyms in general might develop phylogenetically, the specific shibboleths in use would show much less signal. The tree in Figure 9 supports the idea that this development is highly dependent on context; it is split almost down the middle of the family, and appears to be distributed even less randomly than would be expected due to evolution.

While I did not expect to find much phylogenetic signal for specific shibboleths, I did expect to find some for the higher-level semantic groupings. I predicted that there would be even more phylogenetic signal on the structural side of the hierarchy, because while ideological concerns might influence the semantic concepts used for naming, they seemed less likely to have an impact on the specific structure used. I did not expect to find that any of these analyses were purely phylogenetic, because no cultural evolution is, and there were clearly other factors that could influence the development of these names.

3.4 Testing for Phylogenetic Signal

3.4.1 Methods

In order to test the amount of phylogenetic signal in my data, I coded the values for each of my six traits into numerical data. In the Language-Descriptive v. Speaker-Descriptive category, for example, I coded a language as 0 if the name described the speakers and 1 if it described the language. For most of the traits, I had to include an "other" category, because the absence of a particular trait is as important phylogenetically as the state of that trait: in the analysis of the type of Language-Descriptive strategy used, for example, data was coded as 1 for shibbolethnonyms, 2 for evaluative names, and 0 for 'other,' including all non-language-descriptive names. I created a matrix of all of the trait values for every language; using this matrix, I ran three different sets of test of phylogenetic signal.

I used three different sets of tests of phylogenetic signal. First, I found the delta (Holland et al. 2002) and Q-residual (Gray et al. 2010) scores, which measure how tree-like the data is. To measure these scores, I converted analyzed my data using SplitsTree (Huson & Bryant 2006), which, rather than comparing

[7] While there are a handful of other shibbolethnonyms, the vast majority are built on either *this* or *no*, and so to simplify the analysis, I have only included these two. There is also a third category, which includes any language with a name that does not include either *this* or *no*.

the distribution of the true tree, constructs a Neighbor-Net (Bryant & Moulton 2004) network of relationships based on the phylogenetic trait information it is given. Neighbor-Nets are designed to account for the possibility that the relationships among languages are not completely due to evolution. In these networks, rather than a given node having at most one ancestor and two descendants, the end result is a weblike structure, in which each node can have many different connections; these reflect the similarities between multiple taxa that occur as the result of not only evolution but also horizontal influences like borrowing. If the data fits the result of evolution perfectly, the network will indeed take the form of a tree; however, since almost nothing is purely the result of evolution, the result is usually a true network, with multiple connections among different languages. The delta and Q-residual scores quantify how close that network is to a tree; they work by comparing the phylogenetic distances between every two species in a set of four taxa. If these distances exactly fit a tree, the delta and Q-residual scores for that quartet equal 0; otherwise, they use slightly different formulas to result in a score between 0 and 1. After repeating this for every set of four taxa, each taxon receives a delta and Q-residual score that is the average of all the quartets it is a member of; to determine the overall amount of tree-like signal, the scores for all the taxa are averaged.

After finding the delta- and Q-residual scores used the statistical software R to compare my data to Bouckaert, Bowern, and Atkinson's (Bouckaert et al. 2018) phylogeny of Pama-Nyungan languages in order to find the consistency and retention indices (henceforth CI and RI) (Maddison & Maddison 2000). These are actually measures of how much *homoplasy* is in the data; homoplasy is the phenomenon by which one trait has multiple origins throughout the tree. Evolutionary analyses assume that the same trait is unlikely to develop independently multiple times (although it is possible), and so higher levels of homoplasy are associated with lower phylogenetic signal. The CI works by dividing the minimum number of evolutionary steps along the tree by the actual number of steps required to result in the distribution of a given trait; trees with a high number of changes exhibit more homoplasy, and are therefore assumed to have less signal. RI also compares the actual number of changes to the minimum number of changes required, but also takes into consideration the maximum possible number of changes, and is considered to be a little more reliable (Nunn 2011: 31). In both cases, a value of 0 correlates to a high amount of homoplasy and low amount of signal, while a value of 1 correlates to a high amount of signal. Unlike the delta score, this involves comparing the trait distribution to the actual tree, and measures the level of signal for traits rather than taxa. Like with the delta and Q-residual scores, it is also possible to find the overall CI and RI for the entire tree by averaging the scores for each character.

Finally, I again used R to find Fritz & Purvis's (2010) D statistic, which also compares the distribution of data to the actual tree; like the CI and RI, it is used to measure the level of signal for each character. The D statistic results in a value of 0 if the data appears to be the result of evolution, and of 1 if it is perfectly random. It is possible to have a value lower than 0 if the trait is even more clumped among close relatives than expected by evolution, or greater than 1 if traits are even more dispersed than expected from truly random distribution. This test works by counting the number of sister taxa that have the same state; if the data is extremely overly clumped, almost every set of sisters will match, while if it is underclumped, they will never match, etc. For each trait, the analysis returns a value of D, as well as two p-values; a p-value of <0.0.5 for 0 tells you whether you can reject the hypothesis of Brownian evolution, while the p-value of 1 tells you whether you can reject the possibility of a random distribution. The D statistic is designed for binary data, and many of my traits were multistate (such as the Structural category, which was split into "language," "having," "reduplication," and "other"), so I had to convert each of my traits into a set of binary characters. Rather than having one value for Structural that reflected one of the four options, each language would now be coded individually for the presence or absence of "language," the presence or absence of "having," and the presence or absence of "structural." This meant that every column in my matrix now had only values of 1 (for presence) or 0 (for absence). Breaking down my six traits like that resulted in 17 binary characters, one for every measurable category of my classification.

3.4.2 Results

Delta and Q-Residual The mean delta score was 0.1426, which is remarkably small and implies that the signal is extremely tree-like. The mean Q-residual score was 0.0594, also very small. While this seems to suggest a significant amount of phylogenetic signal, however, the network constructed from this data (shown in Figure 11) failed to reconstruct most of the groupings of Pama-Nyungan. The Mayi languages, for example, are not identified as being closely related at all. While "Mayi" in each case means language, the second morpheme in Mayi Yapi and Mayi Thakurti is unanalyzable; these two are grouped with other partially unanalyzable names like Kuuk Thayorre and Kuuku Ya'u, which also mean '[unidentifiable] language', while Mayi Kutuna and Mayi Kulan are on the other side of the network with Yan'nhangu and Guugu-Yimidhirr. As all four of these languages mean something along the lines of 'this language,' it is clear why they were grouped together; it also makes sense that they are close to the group of languages meaning 'this-having' and 'this,' but the network failed to capture the similarities between 'this language' and 'X language' more generally.

The fact that the network looks so different from the actual PamaNyungan tree is interesting. However, given that it was built from only six characters, it is not surprising that it is not particularly accurate: this is not a lot of data to construct relationships from. Unfortunately, this also means that the delta and Q-residual scores are not informative, as the tree-like signal it is identifying is not related to the real Pama-Nyungan tree.

Consistency and Retention Indices The consistency and retention indices did not reveal any signal. The overall consistency index for my data was 0.05371901, while the retention index was 0.2683706. The individual con-

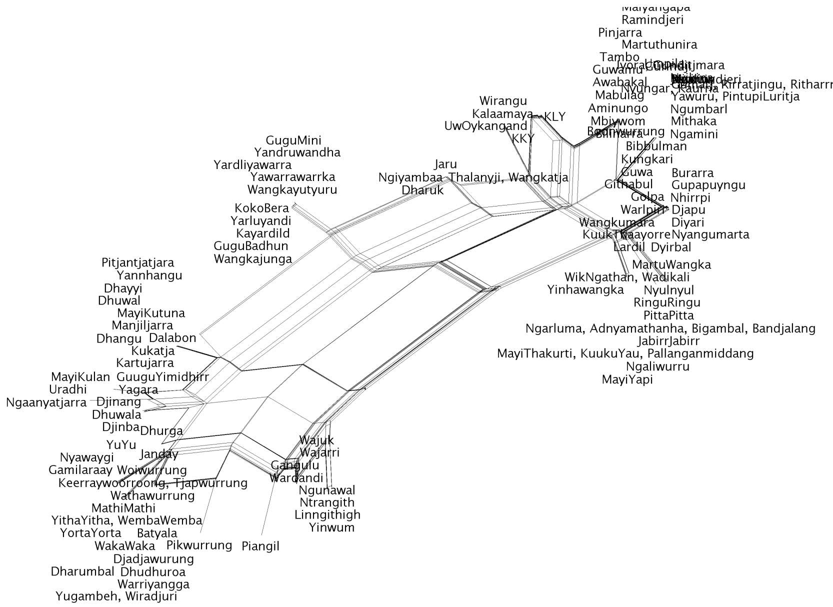


Figure 11: NeighborNet

sistency and retention indices for each character varied slightly, but were all very low and did not differ significantly from the mean. These values are much lower than expected, and suggest that there is no phylogenetic signal. However, CI and RI do not detect signal directly; rather, they identify homoplasy. It is possible for true evolutionary phylogenies to exhibit higher than normal levels of homoplasy, and so the CI and RI are informative, but not definitive.

D Statistic The D statistic results were varied. For some characters, they rejected the possibility of evolution, but in others, they supported it. The results of this analysis are listed in Table 3 below.

Of the 17 levels of the hierarchy that I tested, 8 allowed me to reject the hypothesis of random distribution, while 10 rejected the hypothesis of evolution. The meaningful p-values are marked with stars.[8] For most traits, one of the two hypotheses was rejected; the exceptions are for *Totem*, in which neither was rejectable, and *Shibbolethnonym* and *Language*, which each rejected both hypotheses.

Interestingly, while *Shibbolethnonym* rejected both hypotheses and had a relatively high D statistic, the individual shibboleths had some of the lowest D statistics and each rejected the random hypothesis, supporting the possibility of

[8]The p-value of 1 for *Go* is just barely over the acceptable value of 0.05, and so I have marked it with a star in parentheses; because the standard 0.05 threshold is arbitrary, and this is so close to that threshold, I consider it to be a marginal ninth case of non-randomness.

Table 3: D Statistic Results

	D Estimate	Pval(1)	Pval(0)
Having	0.4018	0.0087*	0.1914
Language	0.6355	0.0126*	0.005*
Reduplication	-0.0946	0.0005*	0.5946
Good	0.8941	0.3	0.0191*
No	0.1191	0 *	0.3769
This	0.4458	0.0077*	0.1293
Go	0.1298	0.0545(*)	0.5439
Unanalyzable	0.9514	0.3555	0*
Men	0.9161	0.3352	0.0238*
Totem	0.7152	0.157	0.1099
Compass	-0.4316	0.0002*	0.7902
Place	0.9657	0.4038	0.002*
Shibbolethnonym	0.7317	0.0395*	0.001*
Describing Language	1.0371	0.5906	0*
Describing Speakers	0.7792	0.0652	0.0008*
People	1.0929	0.7332	0*
Geographic	0.4607	0.0033*	0.0754

evolution. This was contrary to my hypothesis that these would have the least level of signal.

Also contrary to my hypothesis is the fact that the higher-level distinctions of *Describing Language* and *Describing Speakers* had high D statistics and p-values that strongly rejected the hypothesis of evolution. On the other hand, the structural traits all rejected the randomness hypothesis; both *Having* and *Reduplication* had lower D statistics, while *Language*'s D statistic was higher and rejected the evolution hypothesis as well as the hypothesis of randomness. The fact that *Language* seems less likely to be the result of evolution is perhaps unsurprising, given that the use of the word *language* in a language name is naturally extremely widespread.

The other two traits that rejected the hypothesis of randomness were *Compass*—which, along with *Reduplication* had a D statistic that suggests even more clumping than expected with evolution—and *Geographic*, which was probably very influenced by the high level of clumping in the subcategory *Compass*.

3.5 Ancestral State Reconstruction

Despite the low levels of phylogenetic signal recovered for the tree overall, the varied D statistics suggested that some reconstructions of past states should be possible. I therefore attempted to reconstruct the ancestral state of each of my six traits at the root and at each of 19 interior nodes representing both high and low-level subgroups. These subgroups are Karnic, Paman, Yuinkuri, Yolngu, Kaltu, Central, Thura Yura, Pilbara, Mayi, Wati, Wakakabi, Ngumpi, North, East, Western, Kulin, Maric, Central NSW, and Durubal. Just as with the levels of the hierarchy, I chose these groups to get a mix of high- and low-level groupings, and representation from across the entire tree. I expected that there would be clearer phylogenetic signal for more recent ancestors than for earlier ones, and so they would have clearer results.

3.5.1 Methods

To perform these ancestral state reconstructions, I again used the coded data for each trait (including the category "other" where needed) and Bouckaert, Bowern, and Atkinson's 2018 tree. As I did not have data for every language in the tree, I used R to prune the original tree so that it included all and only the languages for which I had data; the intersection of languages in the tree and my data set meant that my analyses had 124 languages each.

Once my data files and trees were matched up, I used BayesTraits (Pagel et al. 2004) to reconstruct the ancestral states of various nodes using a Markov Chain Monte Carlo. BayesTraits models the evolution of traits by using a Markov model that tests possible transition rates within a given prior distribution, and records the probabilities of every possible ancestral state given these rates; at each iteration, the chain proposes a new combination of rates, and measures the likelihood of the corresponding evolutionary history, then it modifies the rates slightly and compares the likelihood of the new trait history to the old one. If the new likelihood is higher, it too is recorded, otherwise it is either accepted or rejected according to a complex algorithm. In theory, if the chain runs infinitely long, it will settle on the most likely evolutionary history; practically, a few million runs is usually enough. The beginning of a chain will probably start at a very low likelihood and work its way up toward the true maximum likelihood; for this reason, the initial data points are discarded as "burn-in" values that are not reflective of reality. Because millions of runs is a vast amount of data, and analyses that are close together in the chain will be autocorrelated, only a sample of the results are recorded.

For each of my analyses, I ran five different chains of 22,000,000 iterations each, with a burn-in of 2,000,000 and a sampling interval of 1000, leaving me with 20,000 data points in each chain. By running multiple chains for each analysis, I decreased the possibility that my results were not overly influenced by outliers.

Each analysis required a different set of priors to ensure meaningful analyses. For the analysis of what type of language-descriptive strategy was used (i.e. evaluative or shibbolethnonym), I used an exponential distribution with a mean of 0.1. The remaining analyses all had uniform distributions of potential rates. The minimum and maximum for these distributions are listed in Table 4.

Once I had my results, I used Tracer (Rambaut et al. 2018) to combine the five chains for each analysis and find the average estimated rates of change between states and the probability of each state for the 20 nodes. I also used Tracer to visually represent the data, using trace plots and density plots of the transition rates and reconstructed probabilities. The trace plots allowed me to

Table 4: Prior Distributions

Character	Minimum	Maximum
Structural	0	0.3
Describing Language or Speakers	0	0.1
Shibbolethnonym or Not	0	0.4
Type of Shibbolethnonym	0	0.2

determine whether the program was testing a range of rates and probabilities, while the density plots allowed me to determine whether it ultimately settled on a single analysis.

3.5.2 Results

- Discuss rates

Perhaps unsurprisingly, given the levels of phylogenetic signal, not all of my results were meaningful. In each case, there was no prior distribution that resulted in each node converging on an analysis. Additionally, the results are very sensitive to changes in priors, which does not help with the validity of the results, as a small change in parameters can have a drastic change in posterior probabilities.

For each analysis, and each node that I attempted to reconstruct, I used Tracer to average my chains, get mean transition rates and probabilities, and create trace and density plots of the probabilities for each possible state of each character. Most of the transition rates between states for all of the traits were very small, suggesting that very little evolution occurs, and that naming strategies are fairly stable through generations.

As examples of the trace and density plots of probabilities, I have included in Figures 12-23 the plots of the probabilities that the root, North, and Durubal were either language-descriptive or speaker-descriptive.

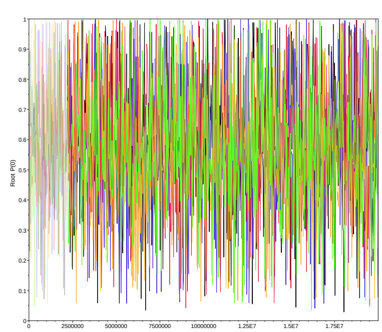


Figure 12: Trace Plot of Probability Root was Language-Descriptive

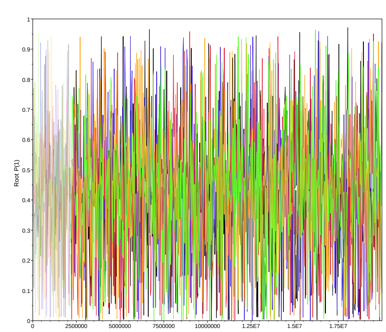


Figure 13: Trace Plot of Probability Root was Speaker-Descriptive

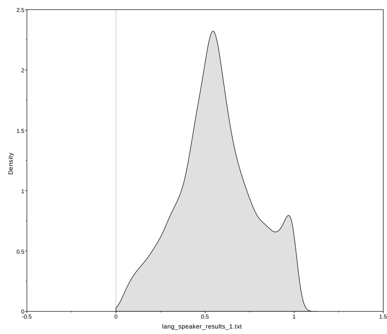


Figure 14: Trace Plot of Probability Root was Language-Descriptive

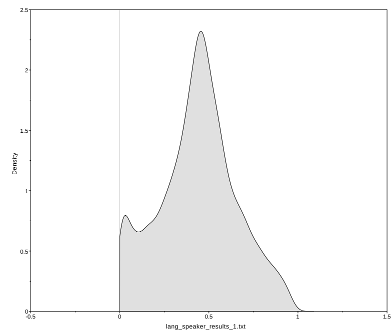


Figure 15: Density Plot of Probability Root was Speaker-Descriptive

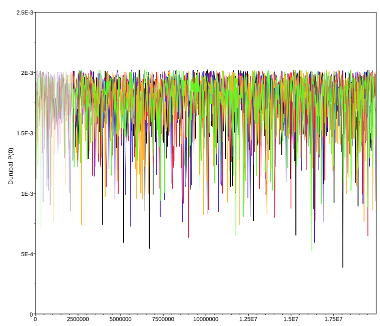


Figure 16: Trace Plot of Probability Durubal was Language-Descriptive

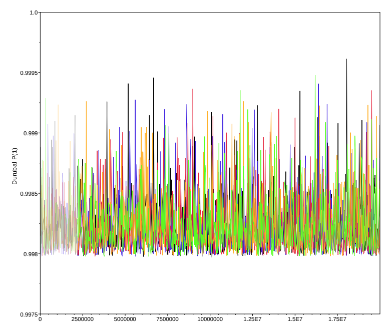


Figure 17: Trace Plot of Probability Durubal was Speaker-Descriptive

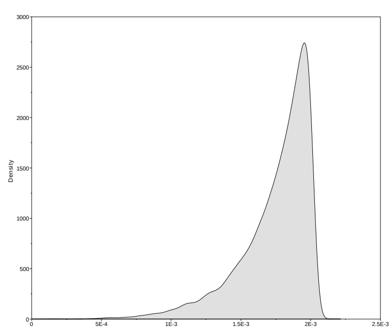


Figure 18: Density Plot of Probability Durubal Language-Descriptive

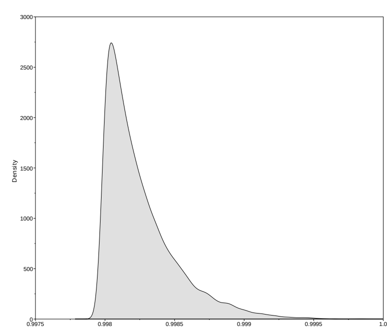


Figure 19: Density Plot of Probability Durubal was Speaker-Descriptive

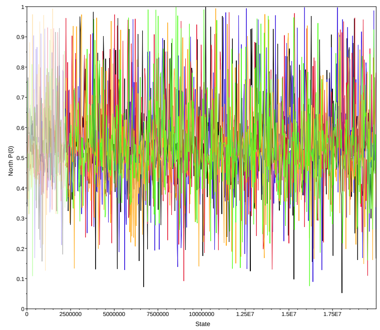


Figure 20: Trace Plot of Probability North was Language-Descriptive

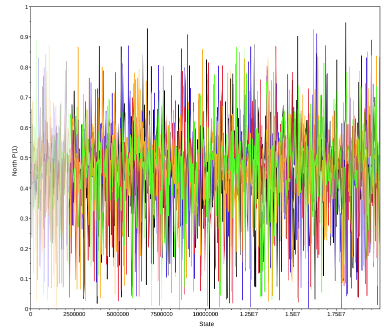


Figure 21: Trace Plot of Probability North was Speaker-Descriptive

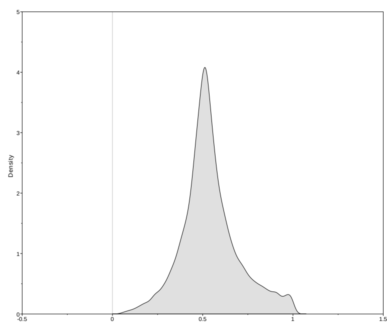


Figure 22: Density Plot of Probability North was Language-Descriptive

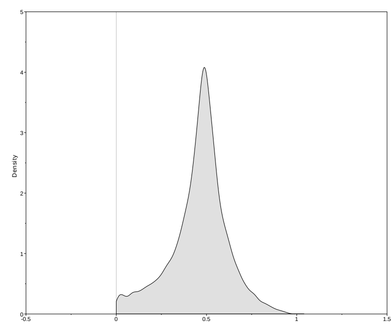


Figure 23: Density Plot Probability North was Speaker-Descriptive

Each node has two trace plots and two density plots because there were two possible states; if the trait had three states, there would be three of each type of plot for each node. Each color in the trace plots represents a different chain; the fact that they overlap very well suggests that they all converged on basically the same answers. In the case of the root and North, the trace plots show that the probabilities varied within a wide range, suggesting that by exploring different rates, it explored the whole range of possibilities for these nodes before settling on an analysis. The trace plots for Durubal look very different: each is grouped at either 0 or 1, with only miniscule differences. In a different scenario, this might be very worrying; however, there are only two Durubalic languages in my data, each of which is language-descriptive. Because the rate of change from speaker-descriptive to language-descriptive is so small, it is almost impossible for Durubal to have been anything but language-descriptive, and so it is unsurprising that the probabilities never strayed far from 1, for language-descriptive, or 0, for speaker-descriptive.

These three nodes also show the three different types of density plots that occurred. Of the three nodes I have shown here, only Durubal resulted in a meaningful reconstruction. Durubal has a 0.998 probability of being language descriptive, and both density plots feature a smooth curve up to a single peak. The density plots for North also have the appropriate shape, more or less reflecting a normal distribution; however, the peak of each of these plots is almost exactly at 0.5. This means that the analysis could not pick up any signal to reconstruct. The root case is slighly more complicated. Rather than having a single peak, both of its density plots have two: a high peak at 0.5, and a smaller peak close to 1 for the language-descriptive probability, and close to 0 for speaker-descriptive. This suggests that it was caught between two possible analyses: one in which it could pick up no signal, and one in which there was weak signal suggesting that the root had a language-descriptive name. It is not, however, possible for one thing to have a probability of both 0.5 and 1, and so this type of result is not meaningful and was excluded from the analysis. I attempted to find parameters that would allow the analysis to hone in on this signal, but was unsuccessful.

Despite the caveats about rates and that some analyses did not settle and others settled on 50-50 probabilities, many of the analyses did provide results; oespecially for many of the lower subgroups, some conclusions can in fact be drawn. The trees of each trait are included below, in Figures 24-29 on pages 41-46. The states of the descendant languages are represented at the tips, while the probabilistic reconstructions for each of the 19 subgroups and the root that I analyzed are represented by pie charts on the nodes; failed analyses are represented with a plain white circle. For ease of reference, each node is labeled with a letter corresponding to the name of a subgroup; the relationships between letters and subgroup names are listed in Table 5.

Table 5: Node Labels

Label	Subgroup
A	Root (Proto-Pama-Nyungan)
B	Karnic
C	Wati
D	Ngumpi
E	Pilbara
F	Kaltu
G	Central
H	Western
I	Thura Yura
J	Paman
K	North
L	East
M	Kulin
N	Maric
O	Mayi
P	Yolngu
Q	Yuinkuri
R	Central NSW
S	Wakakabi
T	Durubal

Figure 24: Reconstruction of Structural Strategies

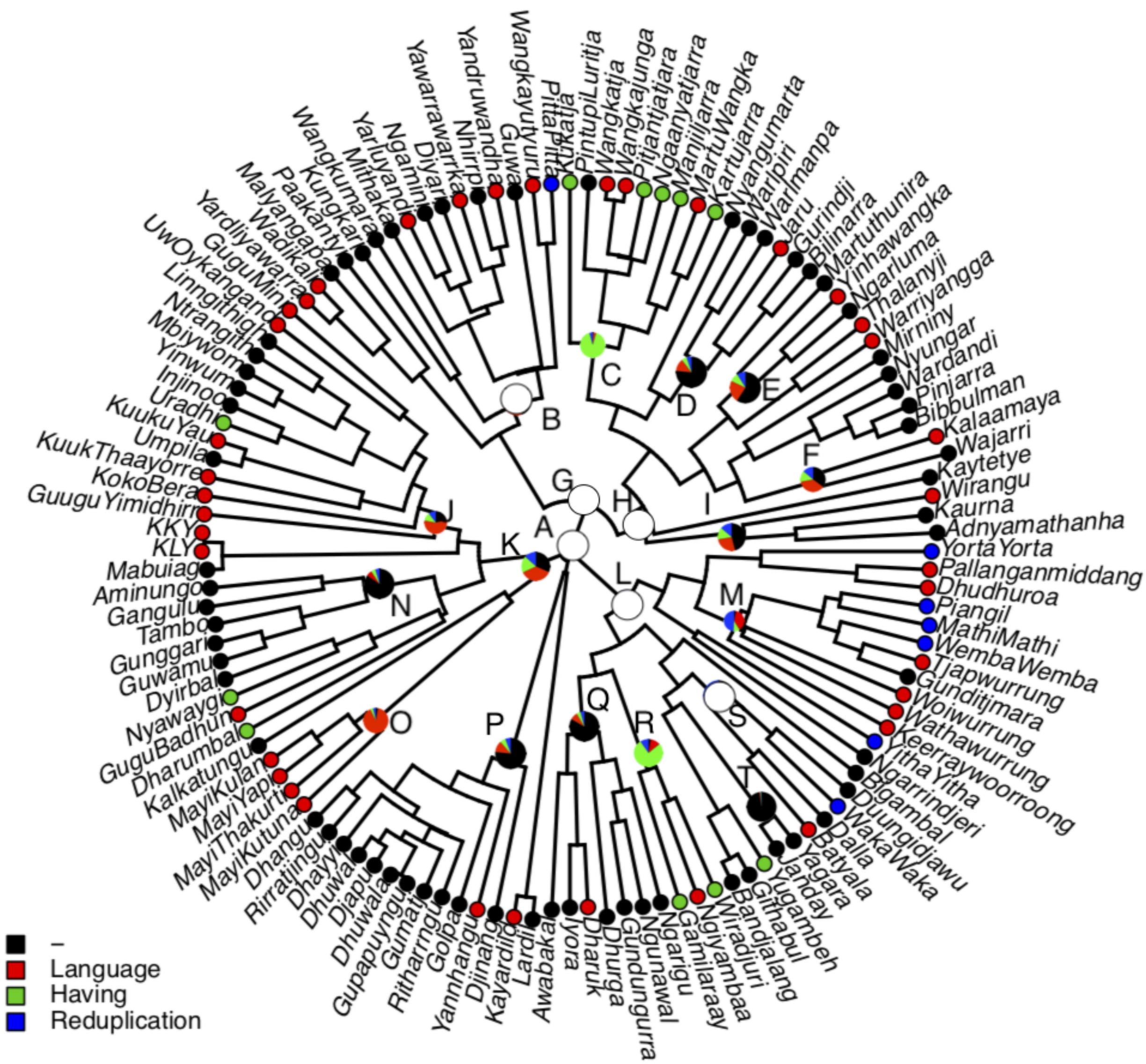


Figure 25: Reconstruction of Strategies Describing Language or Speakers

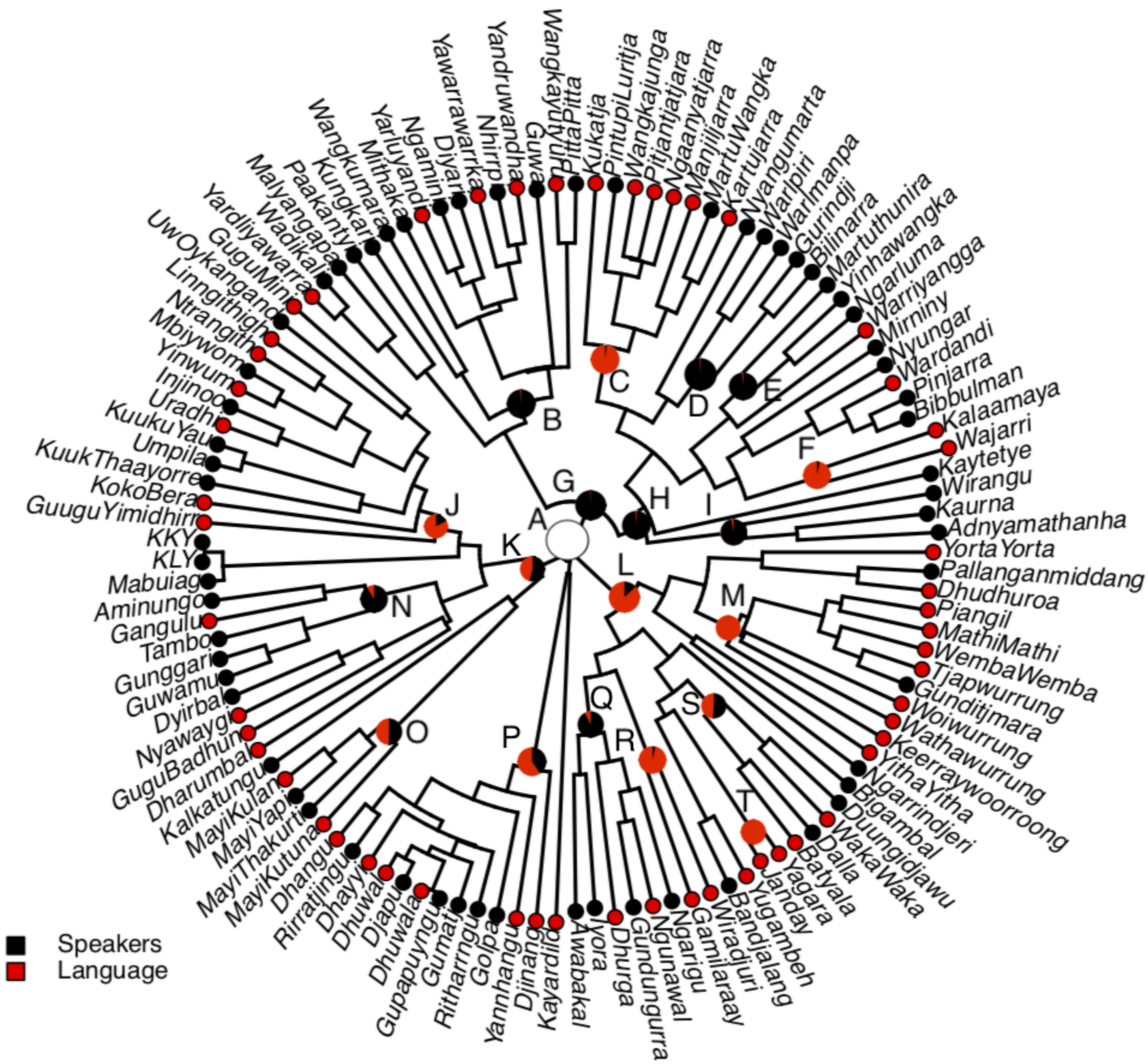


Figure 26: Reconstruction of Shibbolethnonym Presence

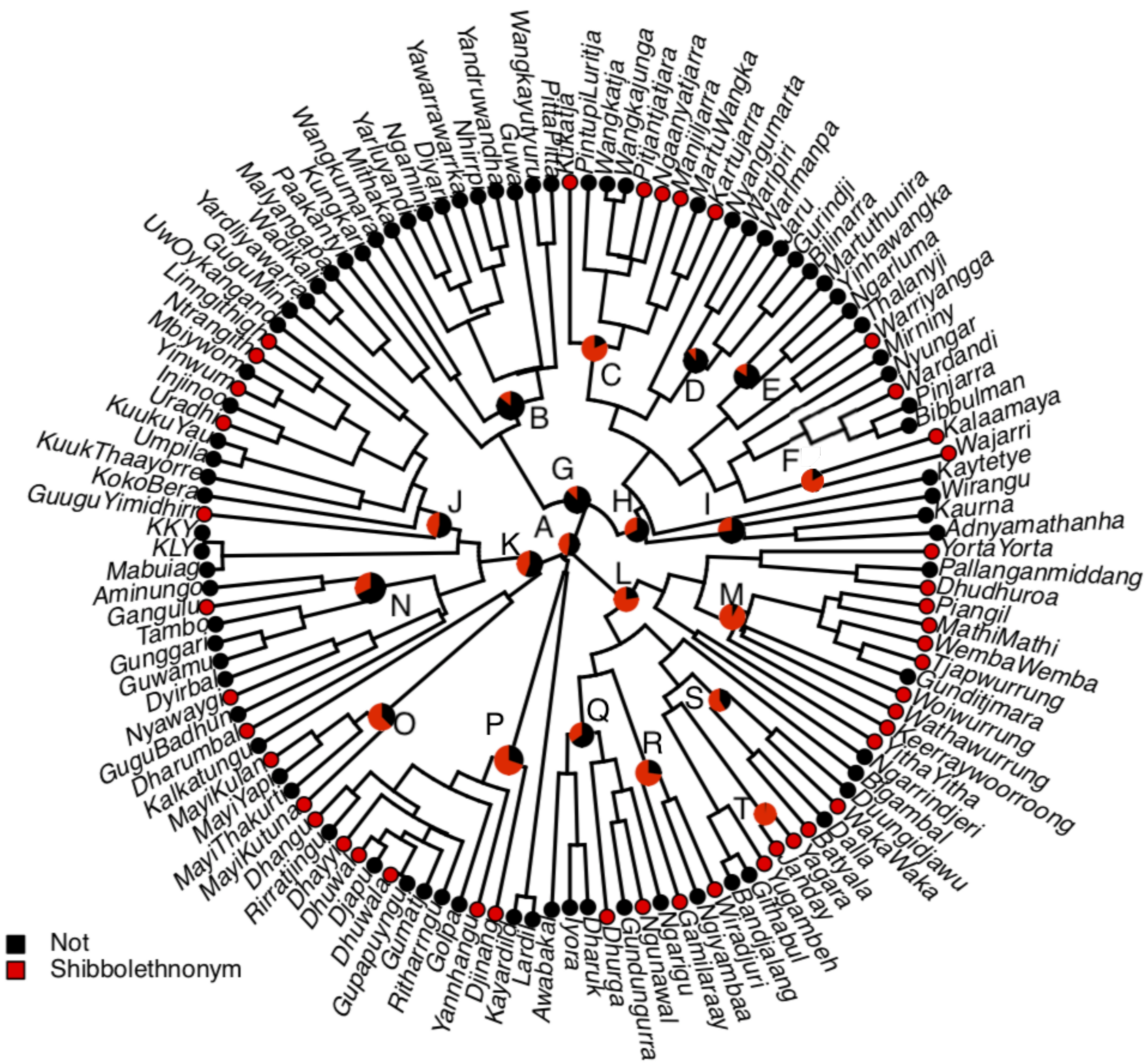


Figure 27: Reconstruction of Language Descriptive Strategies

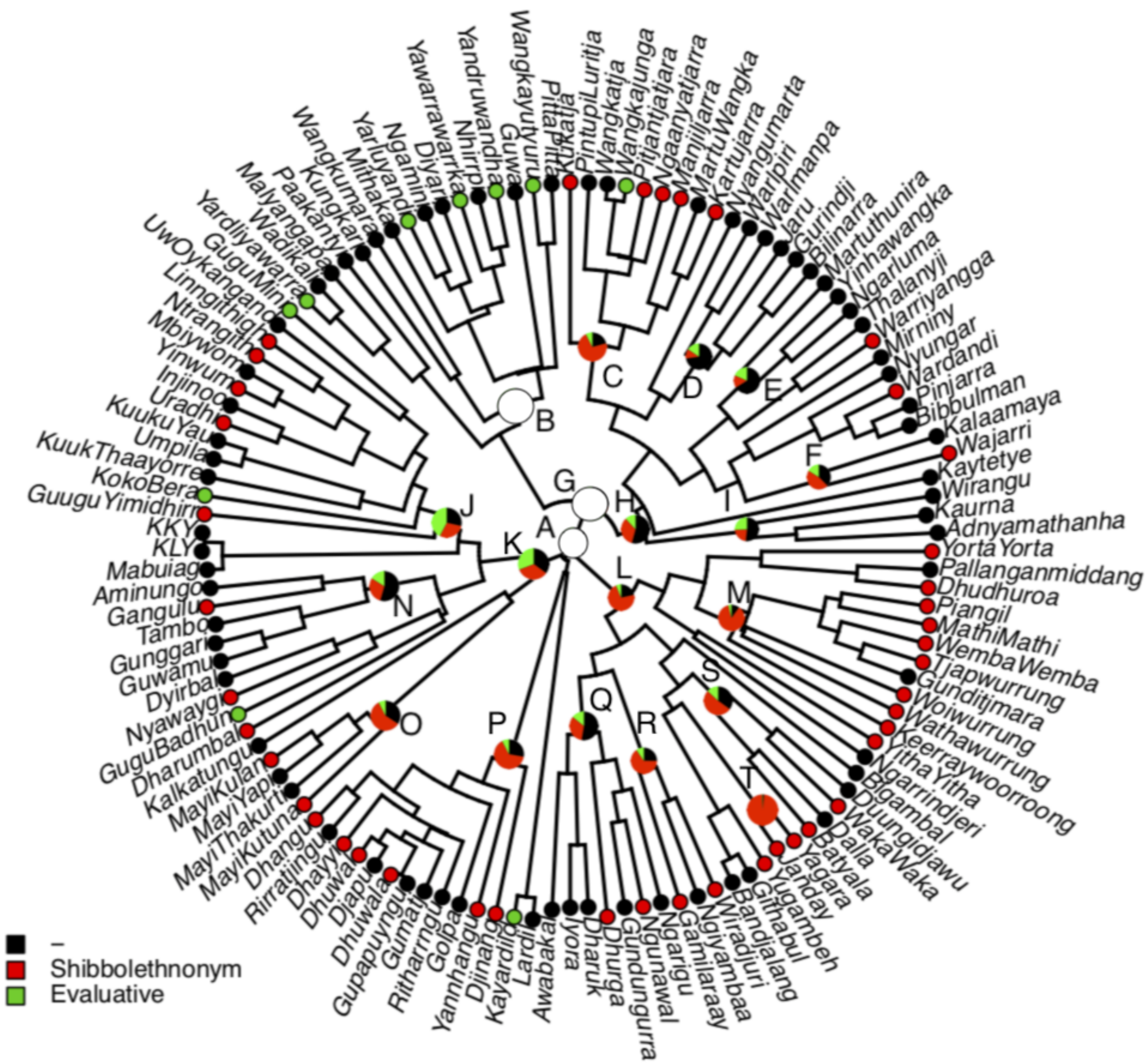


Figure 28: Reconstruction of Shibbolethnonym Type

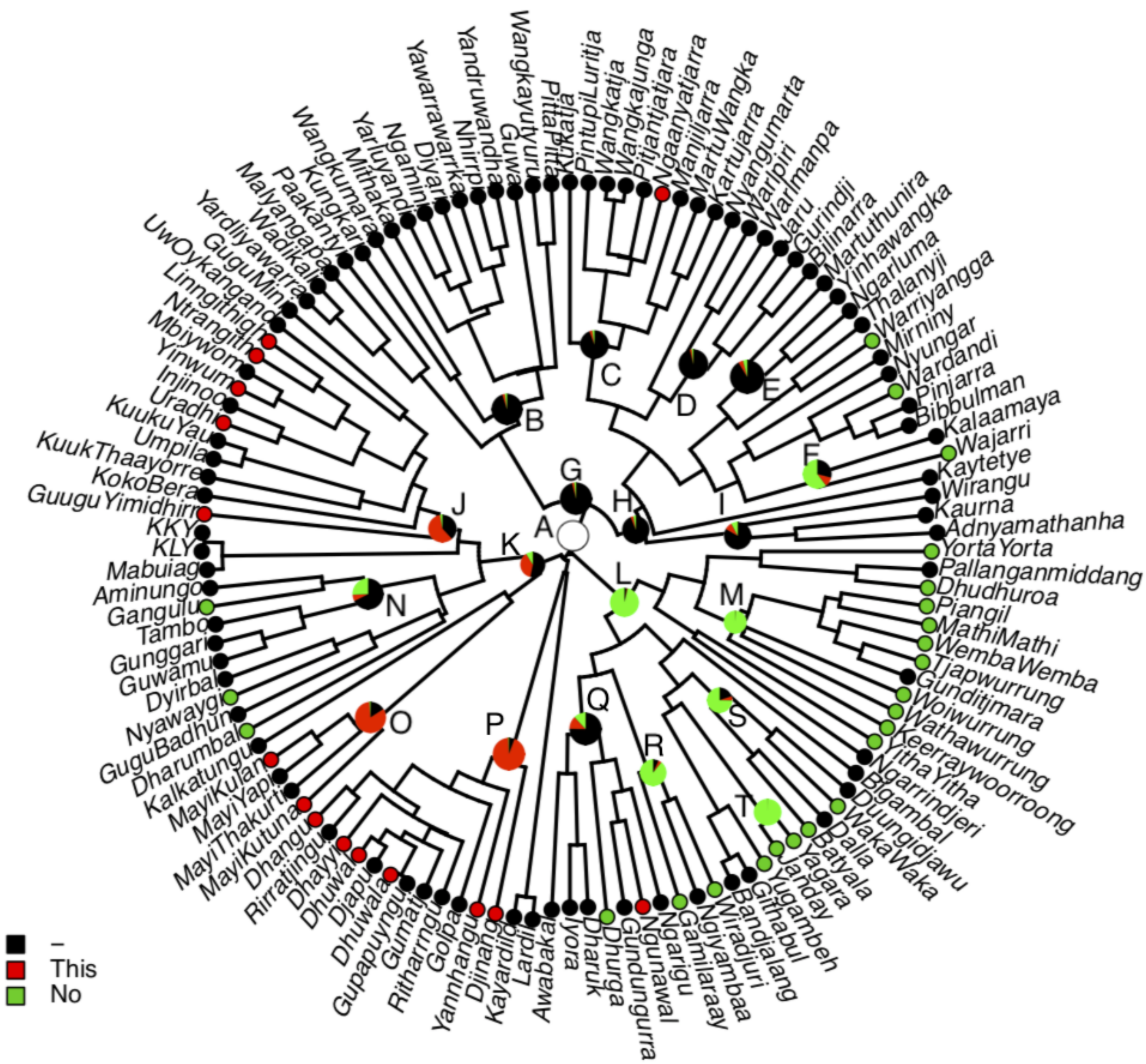
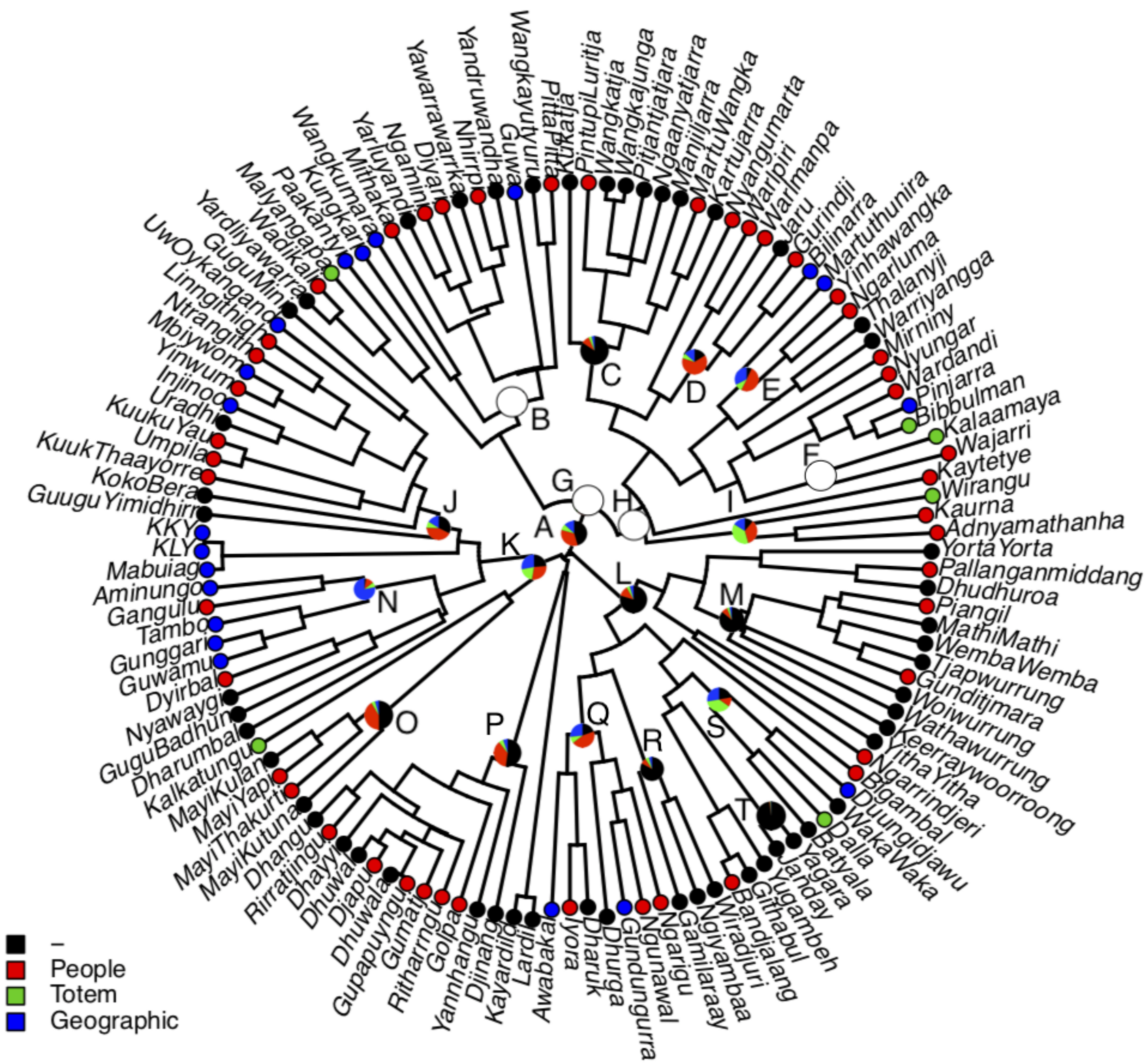


Figure 29: Reconstruction of Speaker Descriptive Strategies



In keeping with the hypothesis that older language names would be less reconstructable, the root only settled on a single analysis in two cases. Even these results, however, are not meaningful, as the probability of Proto-Pama-Nyungan having a shibbolethnonym is 0.468, while the probability of it using a different naming strategy is 0.532. These are basically even odds, suggesting that there was not enough signal to come to a conclusion. Likewise, the speaker-descriptive analysis has basically even probabilities of the original name being a clan name or not being speaker-descriptive. This is slightly more meaningful, as it has mostly ruled out two of the speaker descriptive strategies, but it is still not possible to determine what type of naming strategy Proto Pama-Nyungan used to refer to itself. Similarly, at the higher level nodes, the reconstructed probabilites were usually closer to equal than the lower level nodes, which also fits the hypothesis.

Contrary to the prediction that structural strategies would be the most phylogenetic, and the prediction of the D statistic, the structural analysis had the highest number of nodes (six) that had to be excluded. Of the remaining fourteen nodes, however, eight had high probabilities for a specific state. Ngumpi, Maric, Yolngu, Yuinkuri, and Durubal were all reconstructed as not using any of the three specific structural strategies, while Wati and Central NSW most likely used *-having*, and Mayi used *language*. Interestingly, the strategy of reduplication was never reconstructed, which suggests that it developed very recently.

The other distinction that I thought would be highly phylogenetic was the distinction between names that describe the language and those that describe its speakers. Despite the results of the D-statistic to the contrary, this did turn out to be highly reconstructable: every node but the root settled on a single analysis, and all but four had a very high probability of one or the other. Wati, Kaltu, Paman, East, Kulin, Central NSW, and Durubal all probably used a language descriptive name, while Karnic, Ngumpi, Pilbara, Central, Western, Thura Yura, Maric, and Yuinkuri most likely had names describing their speakers; North, Mayi, Yolngu, and Wakakabi showed much less clear results. Interestingly, this means that the name Paman, which comes from its reconstructed word for *people* (Hale 1966: 162) is probably inaccurate, as Proto-Paman most likely used a language-descriptive name.

Whether or not a name was a shibbolethnonym was almost as reconstructable as the language-speaker distinction. While four nodes, including the root, had roughly 50-50 odds, Durubal, Kulin, Ngumpi, Pilbara, Central, Thura Yura, and Karnic all had high probabilities of not being shibbolethnonyms, and Central NSW, East, and Kaltu most likely were based on shibboleths. Of the remaining five nodes, Mayi and Yolngu were twice as likely to be shibbolethnonyms as not, while Maric, Western and Yuinkuri were twice as likely to not be. It is unsurprising that shibbolethnonyms were less common than every other possible strategy combined, but for an individual strategy it is fairly common.

Interestingly, within the language-descriptive strategies, no nodes were reconstructed as likely to be evaluative. This was probably influenced by the high ratio of shibbolethnonyms to evaluative names, and the high amount of speaker-descriptive languages; additionally, Karnic, which contains many evaluative names but no shibbolethnonyms, was one of the nodes that did not settle on an analysis. The majority of the results had a high probability of Karnic having an evaluative name; however, there was also a peak at 0 for that probability, and so it had to be discounted. Given that the language- or speaker-descriptive analysis settled on Karnic having a speaker-descriptive name, there is clearly some conflicting signal. Durubal, Kulin, East, and Wati all had high probabilites of being shibbolethnonyms; Ngumpi and Pilbara (each of which only have speaker-descriptive descendants) had higher probabilities of not being language-descriptive. Interestingly, these probabilities were both around 0.7, which means that although they were most likely speaker-descriptive, there is a surprisingly large chance that they might have been language-descriptive; this is because the transition rates from either type of language-descriptive name to neither both round to 0.4, which is much higher than the rates of transition from speaker-descriptive to either of the two shibboleths, suggesting that language-descriptive names may have developed fairly long ago but disappear rapidly. This is not, however, supported by the results of the language vs. speaker analysis, which has tiny transition rates for both directions. Maric, Yuinkuri, and Western all had about a 0.5 probability of being neither evaluative or shibbolethnonyms; this makes them being speaker-descriptive the most likely of the three possibilities, but still leaves a 50% chance that they used one of the language-descriptive strategies.

The analysis of shibbolethnonym type—which I had predicted to be less reconstructable than the reconstruction of language-descriptive strategies, but which had higher levels of signal according to the D-statistic—was in fact more reconstructable than the distinction between evaluative and shibbolethnonym, and related nodes often had similar values. This fits with the high level of signal found by the D-statistic. Only the root did not settle on an analysis, and many of the nodes had very high reconstructed probabilities. The Central group and almost all its subgroups (with the exception of Kaltu) had very high probabilities of not being shibbolethnonyms, while the Eastern group and all its subgroups except Yuinkuri had very high probabilities of being *no*-based shibbolethnonyms. Kaltu had a 0.602 probability of being a *no*-based shibbolethnonym, while Yuinkuri had a 0.752 probability of not using either shibboleth. Mayi and Yolngu were both very likely to use *this* as a shibboleth, while Paman had a slightly lower probability (0.6) of being *this*-based, but all three soundly rejected the possibility of being *no*-based, which is unsurprising given that none of their descendants featured *no*. Maric was most likely neither, but possibly *no*-based, while North had a 0.533 probability of being neither and a 0.378 probability of being built on *this*.

The speaker-descriptive strategies were not as reconstructible. While the root did settle on an analysis, it had roughly equal probabilities of either fitting the *People* category or not being speaker-descriptive. Mayi, Yolngu, and Paman were likewise split between *People* and not being speaker-descriptive, with "other" being slightly more likely than People for Mayi and Yolngu. Wati, the Eastern group, Kulin, Central NSW, and Durubal were all most likely not speaker descriptive.The analyses of Karnic, Central, Western, and Kaltu all failed. The only actually speaker-descriptive node that could be reconstructed with confidence was Maric, which used a geographic-based name. The Northern group and Wakakabi were almost evenly divided among all four possibilities; in Ngumpi, Pilbara, and Yuinkuri, People was most likely, but by no means certain.

4 Discussion

While it is possible to partially reconstruct past language names like I did in the previous section, only Durubal, Yolngu, and Central NSW were entirely reconstructable, as *no*, *this*, and *no-having*, respectively. Based on most of the results, Mayi seems likely to have used a shibbolethnonym meaning *this-language*, but this conflicts with the result from the language vs. speaker descriptive analysis, which returned a 50-50 probability. On the other extreme, the root was not at all reconstructable, and Karnic, Central, and Western were all reconstructable only as speaker-descriptive. The fact that Durubal was reconstructable was unsurprising: it has two descendants, both of which have names that mean *no*, and so it is fairly straightforward that the ancestor had the same value.

While it was not possible to reconstruct many past structural or speaker-descriptive strategies, it was eminently possible to reconstruct whether past language names described the actual languages or the people who spoke them, whether they were shibbolethnonyms, and what shibboleth they used. This is largely in contrast to the result of the D statistic analysis: while the shibboleths were found to have signal, shibbolethnonyms in general did not, and neither did the distinction between language-descriptive and speaker-descriptive.

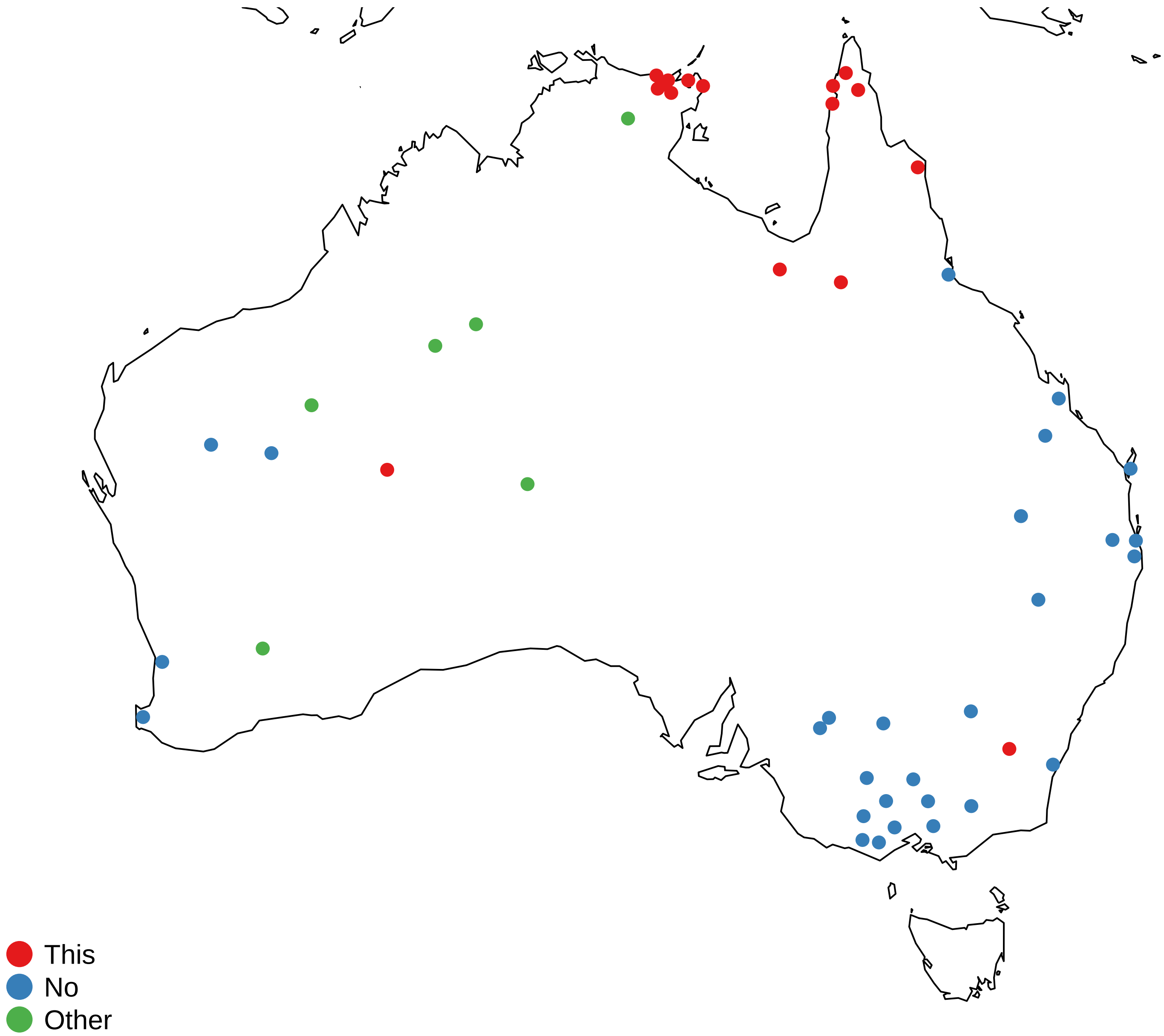
4.1 Potential Areal Influences

The low levels of overall phylogenetic signal, the lack of reconstructability for many strategies, and the conflict between analyses in cases such as Mayi are not entirely unexpected, given that the names of languages are consciously constructed at the time of differentiation between languages, rather than evolving gradually down the tree. The fact that there do seem to be many similarities in naming strategies even with so little signal, however, suggests that there are other factors influencing the development of language names. Given the ideological nature of these names, it seems sensible to expect that there might be a large amount of areal influence on which naming strategies are used. If the primary purpose of language names is to distinguish two neighboring languages, it would be unsurprising if they used similar strategies to name themselves, potentially overriding any inherited naming strategies. Moreover, the naming of languages presumably had to occur at least as often as the diversification events at which geographically close languages were determined to belong to distinct groups, because there were now two languages that wanted to distinguish themselves from each other; this too might influence geographically-close languages to use similar naming strategies.

Since closely related languages are often geographically close to each other as well, this might manifest as something resembling phylogenetic distribution.

An excellent example of this is the analysis of which shibboleth is used for a language name. While there are clear familial groupings of shibboleths, it is clear from the map in Figure 30 that there are just as clearly geographic groupings.

Figure 30: Geographic Distribution of Shibboleths



There is a clear division between a *this*-based cluster in the north and a *no*-based cluster in the east, with a few shibboleths scattered throughout the west. The combination of geography and genetic relationships might have combined to result in such strong patterns both geographically and within the family. The reason I had expected shibboleths to show little signal is that they would be highly contingent on what words the language had in common with nearby languages; if, however, closely related languages were also close geographic neighbors, each language would have a limited set of shibboleths to use, and so their names would end up being similar, leading to the ethnonymic paradigms that McConvell noted. The very presence of a shibbolethnonym suggests that the speakers of that language interacted with, and knew the languages of, their neighbors, because otherwise they would not know which words were distinctive shibboleths.

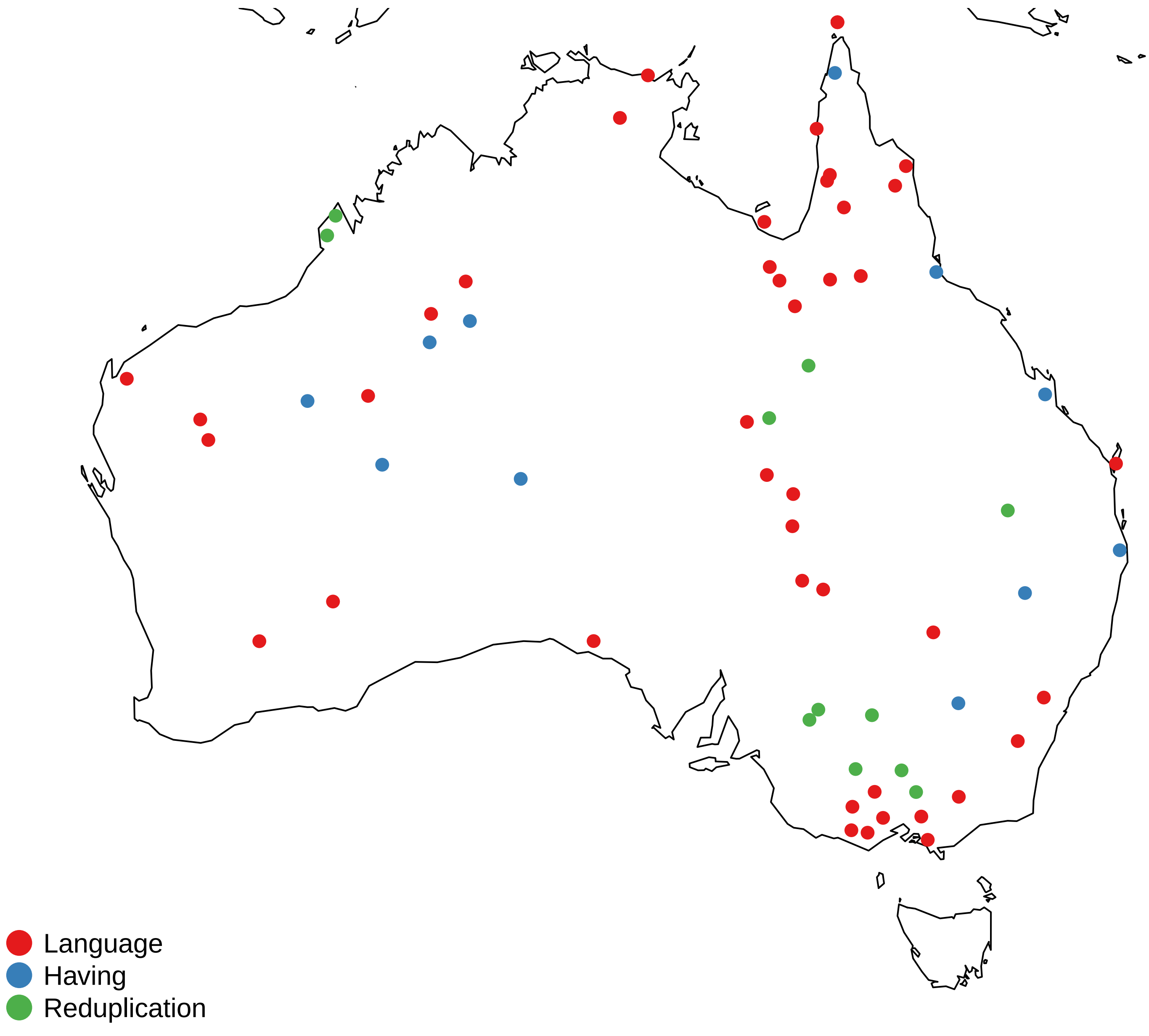
It also seems possible that languages whose neighbors were closely related to them would be more likely to use shibboleths as identifiers; if many aspects of the languages (and cultures) were the same, the shibboleth would stand out as a marker of individuality. This theory, however, limits the reliability of my reconstructed shibbolethnonyms; instead of inheriting a shibboleth from their shared ancestor, languages with a shared ancestor might construct similar shibboleths. This means that, even though a close relationship among languages increases the likelihood of a shared shibboleth, the ancestor of those two languages would not necessarily have used that shibboleth, unless it too was located next to a closely related language that differed by the same word. This is possible, but not guaranteed.

The map of structural strategies in Figure 31 also shows the possibility of areal influences overriding phylogenetic ones.

While *having* is scattered across the continent, there appear to be clusters of names that include *language* and reduplication, particularly in the southeast. These clusters might have caused what seemed to be local phylogenetic relationships that were actually once again the result of geographic influences. This seems to violate my hypothesis that structural components would be less sensitive to areal influences; however, both *-having* and reduplication occur mainly as part of shibbolethnonyms. Any areal influences on the shibboleths might therefore indirectly be influencing the structure of the names.

The logic listed above would predict that speaker-descriptive strategies are less influenced by areal signal, because they do not have the same sort of inherently contrastive meaning. While directional names acknowledge the existence of other groups, for example, they do not require interaction with each other; a name meaning 'people' takes this too the extreme, presupposing that the speakers of that language are the only people, or at least the only people who matter. This suggests that more isolated languages might be more likely to have speaker-descriptive names than language-descriptive ones, or at least that they are less likely to be geographically clustered with similar names. This means that there would be less areal signal interfering with phylogenetic signal; given that most of the speaker-descriptive strategies exhibited low levels of signal and were not conducive to reconstruction, this in turn means that there is little evidence of

Figure 31: Geographic Distribution of Structural Strategies



phylogenetic evolution.

In order to investigate these claims further, it would be necessary to do a more detailed analysis comparing geographic and phylogenetic influences; that is outside the scope of this thesis.

4.2 Other Potential Confounding Factors

There are other possible reasons that might explain why the analyses picked up on very little phylogenetic signal, even if naming strategies truly are subject to evolution. One possibility is that the inclusion of an "other" category complicated the analysis. As I discussed in Section 3.4.1, in order to analyze each trait as compared to the whole tree, most of the characters had to include a category for all the languages that did not fit that strategy. When testing shibbolethnonyms, for example, the data was coded for 'this,' 'no,' and 'other,' in order to acknowledge the possibility that an ancestor language's name might not have been a shibbolethnonym at all. This meant that for each trait, the majority state was 'other,' which included both speaker-descriptive names and evaluative names. Given that these are not really manifestations of the same state, this might have obfuscated any phylogenetic signal that was present. Due to the complexity of the hierarchical classification, however, it was not possible to test every potential state at once, and excluding the "other" category would have led to the false assumption that the every ancestor language was necessarily a shibbolethnonym.

Another issue that might have clouded the amount of signal present was the fact that my data was necessarily incomplete. Because reference grammars often gave me only the semantic content of a name, and ignored the structure, for example, there are many languages that might have potentially fallen into one of the structural strategies, but were instead coded as "other." This would have influenced the results of the analyses. With the data available, however, this was the best that could have been done.

5 Conclusion

By using the hierarchical classification system that I have proposed, it is possible to identify both the similarities and differences among language names, and to discuss both high- and low-level distinctions. The adoption of this principled system allowed me to investigate whether or not the names of Pama-Nyungan languages could be analyzed phylogenetically; I concluded that they show little phylogenetic signal and can only partially be reconstructed. Because most of my root reconstructions failed, it is not possible to know name Proto-Pama-Nyungan used (if, in fact, it had a name); it is possible, however, to determine that Proto-Paman was most likely not called Paman, because the name probably described the language, rather than its speakers.

This sort of discovery allows us to investigate the principles that went into naming specific languages. Because language names are a reflection of the self-identity of the speakers of that language, the specific strategies they choose are important sources of information. Proto-Paman speakers, for example, potentially saw their language as an important part of their individual identity; it would therefore be more accurate to refer to Paman using a language-descriptive term than the speaker-descriptive name we are used to.

Despite this discovery, there was very little overall phylogenetic signal, and many nodes were not reconstructable. While disappointing, because it means that it is not possible to replace arbitrary names like Pama-Nyungan with accurately reconstructed ones, this is in itself a very interesting result. This suggests that the ideological nature of language names, and potentially geographic relationships, are more influential than phylogenetic relationships.

Appendix A: Data

Note: This data includes languages that were not included in my phylogenetic analysis, either because they had English names or because they were not in the tree of Pama-Nyungan that I used. I include them here to provide a more complete dataset that could be used for other purposes. Data without a specific source listed were provided by personal communication with Claire Bowern.

Language Name	**Etymology of Name**	**Source**	**Subgroup**
Adnyamathanha	group name	Sutton 1979: 92	Thura-Yura
Aghu-Tharnggala	totem language	Rigsby n.d.: 4	Paman
Aminungo	place		Maric
Anguthimri	I-having	Crowley 1981: 148	Paman
Antekerrepenh	compass		Arandic
Arabana-Wangkangurru	unanalyzable		Karnic
Aritinngithigh	go X	Hale 1966: 166	Paman
Awabakal	place name	Wafer & Lissarague 2008: 173	Yuin-Kuri
Bandjalang	man	Tindale 1974	Bandjalangic
Baraba Baraba	no-redup	Sutton 1979: 91, Wafer & Lissarague 2008: 68	no subgroup defined
Bardi	unanalyzable		Western Nyulnyulan
Barrow Point	English		Paman
Batyala	no-language	Tindale 1974	Waka-Kabi
Bibbulman	totem	Tindale 1974	Nyungar
Bidhawal	group name	Howitt 1904: 79	Kurnic
Bidjara-Gungabula	unanalyzable	Sutton 1979: 92	Maric
Bigambal	like X	Howitt 1904: 58	Central NSW
Bilinarra	place	Meakins & Nordlinger 2013: 2	Ngumpin-Yapa
Bininj Gun-wok	black man's language		
Boonwurrung	X language	Blake 1991: 47	Kulin
Burarra	unanalyzable		Maningrida
Dadi Dadi	no-redup	Howitt 1904: 53	Lower Murray
Dalabon	go-language		Gunwinjgic
Dalla	totem/place	Tindale, Kite & Wurm 2004: 4	Waka-Kabi
Danggali	group name	Wafer & Lissarague 2008: 266	Paakantyi
Deniliquin	Place		Central NSW
Dhangu	this	Morphy 1983: 3	Yolŋu

Language Name	Etymology of Name	Source	Subgroup
Dharuk	language, speech	Morphy 1983: 5	Yuin-Kuri
Dharumbal	no X	Terrill 2002: 14	Maric
Dhay'yi	this	Morphy 1983: 3	Yolŋu
Dhudhuroa	no redup language	Wafer & Lissarague 2008: 69	Dhuduroa
Dhurga	no	Wafer & Lissarague 2008: 111	Yuin-Kuri
Dhuwal	this	Morphy 1983: 3	Yolŋu
Dhuwala	this	Morphy 1983: 3	Yolŋu
Diyari	unanalyzable		Karnic
Djadjawurung	no-language		Kulin
Djangu	this	Morphy 1983: 3	Yolŋu
Djapu	clan		Yolngu
Djinang	this	Morphy 1983: 3	Yolŋu
Djinba	this	Morphy 1983: 3	Yolŋu
Duungidjawu	place	Kite & Wurm 2004: 4	Waka-Kabi
Dyirbal	clan		Dyirbalic
Eastern and Central Arrernte	English		Arandic
Flinders Island	English		Paman
Gabi-Gabi	no-redup	Kite & Wurm 2004: 4	Waka-Kabi
Galibal	this speakers	Geytenbeek & Geytenbeek 1971: 1	
Gamilaraay	no-having	Ash 2002: 181	Central NSW
Gangulu	no X	Tindale 1974	Maric
Gija	speak		
Githabul	right speakers–or this language Or this	Geytenbeek & Geytenbeek 1971: 1	Bandjalangic
Golpa	clan		Yolŋu
Gooniyandi	compass		Bunuban
Grauadjinggalung	compass	Howitt 1904: 76	Kurnic
Gugu-Badhun	good language	Sutton 1973: 15	Maric
Gugu-Mini	good language	Sutton 1979: 91	Paman
Gugu-Warra	bad language	Sutton 1979: 91	Paman
Gumatj	clan		Yolŋu
Gunditjmara	place men	Howitt 1904: 69	Bungandidj
Gundungurra	compass	Wafer & Lissarague 2008: 110	Yuin-Kuri
Gunggari	compass	Wafer & Lissarague 2008: 325	Maric

Language Name	Etymology of Name	Source	Subgroup
Gunwinygu	X place	Evans 2003: 11	no subgroup defined
Gupapuyngu	clan		Yolŋu
Gureng Gureng	no-redup	Kite & Wurm 2004: 4	Waka-Kabi
Gurindji	unanalyzable		Ngumpin-Yapa
Guugu-Yimidhirr	this-having language		Paman
Guwa	compass	Breen 1990: 109	Maric
Guwamu	compass	Wafer & Lissarague 2008: 325	Maric
Hunter River and Lake Macquarie	English		
Injinoo	place		Paman
Iyora	person	Wafer & Lissarague 2008: 146	Yuin-Kuri
Jabirr-Jabirr	redup (unanalyzable)		Western Nyulnyulan
Janday	no	Tindale 1974	Durubulic
Jaru	language, speech		Ngumpin-Yapa
Jawi	unanalyzable		Western Nyulnyulan
Jukun	unanalyzable		Eastern Nyulnyulan
Kala Lagaw Ya	compass place language		
Kalaamaya	fire language		Nyungar
Kalaw Kawaw Ya	compass		
Kalkatungu	totem	Sutton 1979: 91	Kalkatungic
Kartujarra	go-having		Wati
Kaurna	people	Amery 2000: 3	Thura-Yura
Kayardild	good language		
Kaytetye	clan		Arandic
Keerraywoorroong	no-language		Kulin
Kok-Pap?ngk	good language	Black 2004: 253	Paman
Koko-Bera	good language	Black 2004: 253	Paman
Kukatja	meat having		Wati
Kuku-Mu'inh	X language	Sutton 91	Paman
Kuku-Thaypan	totem language	Rigsby n.d.: 3	Paman
Kuku-Yalanji	this language	Breen 1981: 2	Paman

Language Name	Etymology of Name	Source	Subgroup
Kune	varieties are Kune Narayek, na-rayek = masc.pref+hard and Kune Dulerayek, dule-rayek = word, language; unclear what Kune means	Evans 2003: 11	
Kungkari	compass		Karnic
Kuuk Thaayorre	X mouth language or ethnonym language	Hall 1972: 2 Gaby 2006: 2	Paman
Kuuk-Yak	totem language	Gaby 2006: 3	Paman
Kuuku-Ya'u	X language	Thompson 1988	Paman
Ladji-Ladji	no-redup	Blake et al. 2011: 136, Wafer & Lissarague 2008: 68	Kulin
Lardil	strong customs/law (la(ka) yardil(d))		Tangkic
Linngithigh	this X	Hale 1966: 164	Paman
Lower Southern Aranda	English		Arandic
Luthigh	this X	Hale 1966: 164	Paman
Mabuiag	place		Western Torres
Malyangapa	totem Place	Wafer & Lissarague 2008: 296	Yardli
Manjiljarra	go-having		Wati
Martu Wangka	person's language		Wati
Martuthunira	place	Dench 1987: 1-2	Ngayarta
Mathi-Mathi	no-redup	Wafer & Lissarague 2008: 68	Kulin
Mayali	mind, language		
Mayi-Kulan	this language	Breen 1981: 2	Mayi
Mayi-Kutuna	this language	Breen 1981: 2	Mayi
Mayi-Thakurti	X language	Breen 1981: 2	Mayi
Mayi-Yapi	X language	Breen 1981: 2	Mayi
Mbiywom	place	Hale 1966: 166	Paman
Miriwoong	this X		Miriwoongic
Mirniny	unanalyzable		Mirniny
Mithaka	unanalyzable		Karnic
Miwa	place	McGregor 1993: 5	Northern Worrorran

Language Name	Etymology of Name	Source	Subgroup
Muk-Thang	good language	Wafer & Lissarague 2008: 96	
Muliarra	go-having		
Murrinh-patha	good language		
Nakara	this-having	Eather 2005: 80	Maningrida
Nari Nari	no-redup	Wafer & Lissarague 2008: 68	Kulin
Ndjebbana	this X	McKay 2000: 156	Maningrida
Ngaanyatjarra	this-having	Glass & Hackett 2003: 1	Wati
Ngalia	I		
Ngaliwurru	X-having	Schultze-Berndt 2000: 11	no subgroup defined
Ngamini	unanalyzable		Karnic
Ngara:ngwal	what speakers	Geytenbeek & Geytenbeek 1971: 1	
Ngarigu	clan name	Howitt 1904: 78	Yuin-Kuri
Ngarluma	clan name	Hall 1971: 15	Ngayarta
Ngarrindjeri	man	Wafer & Lissarague 2008: 34	Lower Murray
Ngiyambaa	language, speech	Wafer & Lissarague 2008: 222	Central NSW
Ngkoth	this X	Hale 1966: 165	Paman
Ngumbarl	unanalyzable		Eastern Nyulnyulan
Ngunawal	this X	Wafer & Lissarague 2008: 110	Yuin-Kuri
Nhirrpi	unanalyzable	Bowern 2000	Karnic
Nimanburru	totem		Western Nyulnyulan
Ntra'ngith	this X	Hale 1966: 165	Paman
Ntrwa'ngayth	this X	Hale 1966: 165	
Nyangumarta	X-comparative	Sharp 2004: 2	Marrngu
Nyawaygi	no-having	Dixon 1983: 433	Nyawaygic
Nyikina	unanalyzable		Eastern Nyulnyulan
Nyulnyul	redup (unanalyzable)		Western Nyulnyulan
Nyungar	man	Tindale 1974	Nyungar
Ogunyjan	eating language		
Paakantyi	place name	Wafer & Lissarague 2008: 266	Macro NSW

Language Name	**Etymology of Name**	**Source**	**Subgroup**
Pallanganmiddang	X language		Gippsland
Pantyikali	Place name or group name	Wafer & Lissarague 2008: 266	Paakantyi
Parrintyi	Place name	Wafer & Lissarague 2008: 266	Paakantyi
Piangil	no redup	Wafer & Lissarague 2008: 68	Nyungar
Pikwurrung	no-language		Kulin
Pinjarra	place name	Tindale 1974	Nyungar
Pinjarup	place name	Tindale 1974	
Pintupi-Luritja	unanalyzable-stranger		Wati
Pitjantjatjara	go-having		Wati
Pitjantjatjara-Yankunytjatjara	go-having		Wati
Pitta-Pitta	Redup (unanalyzable)		Karnic
Ramindjeri	place		Lower Murray
Rimanggudinhma	place compound	Godman 1993: 6 FIX CITATION	Paman
Ringu-Ringu	redup (unanalyzable)		Karnic
Rirratjingu	clan		Yolŋu
Ritharrngu	clan		Yolŋu
Southern Paakantyi	English		Paakantyi
Southern Walmajarri	English		Ngumpin-Yapa
Tambo	place		Maric
Thalanyji	language, speech	Austin 1992	Kanyara-Mantharta
Tiwi	man	Sutton 1979: 92	
Tjapwurrung	no-language		Kulin
Umpila	unanalyzable	Sutton 1979: 92	Paman
Upper Clarence River	English		Bandjalangic
Uradhi	this-having	Crowley 1983: 309	Paman
Uw Olkol	x language		
Uw Oykangand	X place language	Sommer 2006: 1	Paman
Wadikali	X language	Hercus & Austin 2004: 228	Yardli
Wajarri	no X	Tindale 1974	Kartu
Wajuk	no X	Tindale 1974	Nyungar

Language Name	Etymology of Name	Source	Subgroup
Waka-Waka	no-redup	Tindale 1974	Waka-Kabi
Wangkajunga	good language	Jones 2011: 11	Wati
Wangkatja	language, speech		Ngayarta
Wangkayutyuru	good language		Karnic
Wangkumara	compass	Wafer & Lissarague 2008: 296	Karnic
Wardandi	no X		Nyungar
Warlmanpa	unanalyzable		Ngumpin-Yapa
Warlpiri	unanalyzable		Ngumpin-Yapa
Warriyangga	no-language	O'Grady 2009	Kanyara-Mantharta
Warrwa	unanalyzable		Eastern Nyulnyulan
Wathawurrung	no-language		Kulin
Wayilwan	no X	Wafer & Lissarague 2008: 222	Central NSW
Waywurru	no-language	Wafer & Lissarague 2008: 69	Gippsland
Wellington	English		Lower Murray
Wemba-Wemba	no-redup	Wafer & Lissarague 2008: 68	Kulin
Western Arrarnte	English		Arandic
Western Desert	English		
Western Torres	English		
Wik-Mungkan	eating language		Paman
Wik-Ngathan	X language		Paman
Wiradjuri	no-having		Central NSW
Wirangu	sky/cloud language	Hercus 1999: 4-5	Thura-Yura
Wiriyaraay	no-having	Wafer & Lissarague 2008: 223	
Woiwurrung	no-language	Howitt 1904: 41	Kulin
Wotjobaluk	man, group	Howitt 1904: 41, 54	Maric
Wunumara	unalaysable		Mayi
Yabula-Yabula	no-redup	Wafer & Lissarague 2008: 68	Yotayotic
Yagara	no	Tindale 1974	Durubulic
Yan-nhangu	this language		Yolŋu
Yandruwandha	good language		Karnic
Yankunytjatjara	go-having		Wati
Yapurarra	compass		
Yardliyawarra	good language		Yardli
Yari-Yari	no-redup	Wafer & Lissarague 2008: 33	

Language Name	Etymology of Name	Source	Subgroup
Yarluyandi	good language		Karnic
Yawarrawarrka	good language		Karnic
Yawuru	unanalyzable		Eastern Nyulnyulan
Yinhawangka	person's language		Pilbara
Yinwum	this X	Hale 1966: 164	Paman
Yir-Yoront	this language		Paman
Yitha-Yitha	no-redup	Wafer & Lissarague 2008: 33	Lower Murray
Yorta Yorta	no-redup	Wafer & Lissarague 2008: 68	Yotayotic
Yu Yu	no-redup	Wafer & Lissarague 2008: 33	Lower Murray
Yugambeh	no-having or no		Bandjalangic

Appendix B: Reconstructions

Reconstructions with an asterisk did not actually settle on a single analysis.

Table 7: Averaged Reconstructions of Structural Strategies

Key:
0: None
1: Language
2: Having
3: Reduplication

	Mean	Effective Sample Size
Lh	-124.873	17800
Harmonic Mean	-127.896	13
q01	0.179	17213
q02	4.63E-02	17800
q03	2.46E-02	17800
q10	0.213	17800
q12	8.69E-02	17299
q13	8.88E-02	17800
q20	0.179	17375
q21	0.206	17800
q23	0.118	17800
q30	0.176	16649
q31	0.206	17753
q32	0.144	17800
Root P(0)*	0.614	17288
Root P(1)*	0.173	16466
Root P(2)*	9.94E-02	17385
Root P(3)*	0.114	16808
Durubal P(0)	0.975	17518
Durubal P(1)	9.71E-03	17800
Durubal P(2)	7.53E-03	17457
Durubal P(3)	7.33E-03	17455
CentralNSW P(0)	2.71E-02	17800
CentralNSW P(1)	0.112	17104
CentralNSW P(2)	0.765	17105
CentralNSW P(3)	9.53E-02	17800
Maric P(0)	0.845	17681
Maric P(1)	6.14E-02	17800
Maric P(2)	4.75E-02	17428
Maric P(3)	4.61E-02	17626
Kulin P(0)	3.30E-02	17054
Kulin P(1)	0.378	17679
Kulin P(2)	9.76E-02	17800
Kulin P(3)	0.491	17531

Western P(0)*	0.607	17531
Western P(1)*	0.144	17800
Western P(2)*	0.164	17528
Western P(3)*	8.59E-02	17005
East P(0)*	0.228	17475
East P(1)*	0.19	17800
East P(2)*	0.176	17800
East P(3)*	0.406	17554
North P(0)	0.316	16969
North P(1)	0.367	17772
North P(2)	0.184	17376
North P(3)	0.133	16946
Ngumpi P(0)	0.765	17587
Ngumpi P(1)	0.13	17800
Ngumpi P(2)	5.33E-02	17486
Ngumpi P(3)	5.20E-02	16799
Wakakabi P(0)	0.275	16800
Wakakabi P(1)	0.233	17800
Wakakabi P(2)	0.134	17800
Wakakabi P(3)	0.358	15690
Wati P(0)	9.15E-03	17572
Wati P(1)	3.60E-02	17405
Wati P(2)	0.912	17373
Wati P(3)	4.29E-02	17800
Mayi P(0)	2.83E-02	17197
Mayi P(1)	0.899	17800
Mayi P(2)	3.61E-02	16013
Mayi P(3)	3.65E-02	17250
Pilbara P(0)	0.592	17551
Pilbara P(1)	0.228	17800
Pilbara P(2)	9.07E-02	17440
Pilbara P(3)	8.88E-02	16660
ThuraYura P(0)	0.46	17411
ThuraYura P(1)	0.264	17749
ThuraYura P(2)	0.139	17356
ThuraYura P(3)	0.137	16710
Central P(0)*	0.709	17661
Central P(1)*	0.121	17800
Central P(2)*	0.101	17561
Central P(3)*	6.88E-02	17220
Kaltu P(0)	0.352	16991
Kaltu P(1)	0.37	17206
Kaltu P(2)	0.14	17282
Kaltu P(3)	0.138	16806
Yolngu P(0)	0.766	17651
Yolngu P(1)	0.118	17800

Yolngu P(2)	5.84E-02	17478
Yolngu P(3)	5.70E-02	16823
Yuinkuri P(0)	0.833	17782
Yuinkuri P(1)	8.15E-02	17800
Yuinkuri P(2)	4.35E-02	17535
Yuinkuri P(3)	4.23E-02	17524
Paman P(0)	0.213	16908
Paman P(1)	0.566	17211
Paman P(2)	0.117	17310
Paman P(3)	0.103	17800
Karnic P(0)*	0.638	17621
Karnic P(1)*	0.166	17299
Karnic P(2)*	7.77E-02	17632
Karnic P(3)*	0.118	16649

Table 8: Averaged Reconstructions of Language Descriptive vs. Speaker Descriptive

Key:
0: Speaker Descriptive
1: Language Descriptive

	Mean	Effective Sample Size
Lh	-103.186	16599
Harmonic Mean	-104.952	6
q01	9.26E-02	17800
q10	8.97E-02	16540
Root P(0)*	0.564	16872
Root P(1)*	0.436	16872
Durubal P(0)	1.74E-03	17795
Durubal P(1)	0.998	17795
CentralNSW P(0)	3.45E-02	17800
CentralNSW P(1)	0.965	17800
Maric P(0)	0.914	16843
Maric P(1)	8.64E-02	16843
Kulin P(0)	4.22E-03	17800
Kulin P(1)	0.996	17800
Western P(0)	0.986	17134
Western P(1)	1.43E-02	17134
East P(0)	0.13	16318
East P(1)	0.87	16318
North P(0)	0.54	16880
North P(1)	0.46	16880
Ngumpi P(0)	0.99	16663
Ngumpi P(1)	1.03E-02	16663
Wakakabi P(0)	0.513	16908

Wakakabi P(1)	0.487	16908
Wati P(0)	1.96E-02	16512
Wati P(1)	0.98	16512
Mayi P(0)	0.504	16908
Mayi P(1)	0.496	16908
Pilbara P(0)	0.987	16649
Pilbara P(1)	1.28E-02	16649
ThuraYura P(0)	0.973	16642
ThuraYura P(1)	2.67E-02	16642
Central P(0)	0.994	16924
Central P(1)	6.46E-03	16924
Kaltu P(0)	3.24E-02	17800
Kaltu P(1)	0.968	17800
Yolngu P(0)	0.404	17041
Yolngu P(1)	0.596	17041
Yuinkuri P(0)	0.923	16970
Yuinkuri P(1)	7.67E-02	16970
Paman P(0)	0.172	16329
Paman P(1)	0.828	16329
Karnic P(0)	0.988	17059
Karnic P(1)	1.22E-02	17059

Table 9: Averaged Reconstructions of Shibbolethnonym Presence

Key:
0: Not a Shibbolethnonym
1: Shibbolethnonym

	Mean	Effective Sample Size
Lh	-82.749	89000
Harmonic Mean	-83.969	24
q01	0.223	86603
q10	0.341	89000
Root P(0)	0.532	85464
Root P(1)	0.468	85464
Durubal P(0)	1.10E-02	88047
Durubal P(1)	0.989	88047
CentralNSW P(0)	0.271	85860
CentralNSW P(1)	0.729	85860
Maric P(0)	0.677	89000
Maric P(1)	0.323	89000
Kulin P(0)	7.77E-02	88195
Kulin P(1)	0.922	88195
Western P(0)	0.676	89000
Western P(1)	0.324	89000
East P(0)	0.224	85244

East P(1)	0.776	85244
North P(0)	0.559	85693
North P(1)	0.441	85693
Ngumpi P(0)	0.895	89000
Ngumpi P(1)	0.105	89000
Wakakabi P(0)	0.412	85727
Wakakabi P(1)	0.588	85727
Wati P(0)	0.18	85890
Wati P(1)	0.82	85890
Mayi P(0)	0.375	85922
Mayi P(1)	0.625	85922
Pilbara P(0)	0.843	89000
Pilbara P(1)	0.157	89000
ThuraYura P(0)	0.729	89000
ThuraYura P(1)	0.271	89000
Central P(0)	0.873	89000
Central P(1)	0.127	89000
Kaltu P(0)	0.167	89000
Kaltu P(1)	0.833	89000
Yolngu P(0)	0.302	86281
Yolngu P(1)	0.698	86281
Yuinkuri P(0)	0.651	86947
Yuinkuri P(1)	0.349	86947
Paman P(0)	0.538	85993
Paman P(1)	0.462	85993
Karnic P(0)	0.858	89000
Karnic P(1)	0.142	89000

Table 10: Averaged Reconstructions of Language Descriptive Strategies

Key:
0: Not Language-Descriptive
1: Shibbolethnonym
2: Evaluative

	Mean	Effective Sample Size
Lh	-115.506	17800
Harmonic Mean	-119.603	10
q01	0.242	17765
q02	8.23E-02	17496
q10	0.368	17800
q12	3.55E-02	17317
q20	0.429	17800
q21	0.109	17800
Root P(0)*	0.317	16563
Root P(1)*	0.414	17800

Root P(2)*	0.269	17800
Durubal P(0)	1.38E-02	17663
Durubal P(1)	0.981	17601
Durubal P(2)	5.02E-03	17800
CentralNSW P(0)	0.251	17800
CentralNSW P(1)	0.653	17800
CentralNSW P(2)	9.55E-02	17800
Maric P(0)	0.545	17800
Maric P(1)	0.294	17800
Maric P(2)	0.161	17800
Kulin P(0)	9.94E-02	17800
Kulin P(1)	0.863	17777
Kulin P(2)	3.71E-02	17800
Western P(0)*	0.527	17800
Western P(1)*	0.279	17800
Western P(2)*	0.194	17800
East P(0)	0.216	17800
East P(1)	0.721	17800
East P(2)	6.35E-02	17800
North P(0)	0.352	17800
North P(1)	0.341	17800
North P(2)	0.307	17800
Ngumpi P(0)	0.727	17800
Ngumpi P(1)	0.118	17800
Ngumpi P(2)	0.155	17800
Wakakabi P(0)	0.35	17800
Wakakabi P(1)	0.52	17800
Wakakabi P(2)	0.13	17800
Wati P(0)	0.169	17800
Wati P(1)	0.766	17800
Wati P(2)	6.50E-02	17800
Mayi P(0)	0.342	17800
Mayi P(1)	0.582	17800
Mayi P(2)	7.59E-02	17800
Pilbara P(0)	0.662	17800
Pilbara P(1)	0.153	17800
Pilbara P(2)	0.185	17800
ThuraYura P(0)	0.518	17800
ThuraYura P(1)	0.226	17800
ThuraYura P(2)	0.256	17800
Central P(0)*	0.413	17800
Central P(1)*	0.169	17800
Central P(2)*	0.418	17800
Kaltu P(0)	0.373	17098
Kaltu P(1)	0.464	17800
Kaltu P(2)	0.164	17800

Yolngu P(0)	0.278	17800
Yolngu P(1)	0.645	17800
Yolngu P(2)	7.63E-02	17800
Yuinkuri P(0)	0.527	17800
Yuinkuri P(1)	0.331	17800
Yuinkuri P(2)	0.142	17800
Paman P(0)	0.286	16626
Paman P(1)	0.292	17800
Paman P(2)	0.422	17800
Karnic P(0)*	0.263	17800
Karnic P(1)*	6.82E-02	17752
Karnic P(2)*	0.669	17800

Table 11: Averaged Reconstructions of Shibbolethnonym Type

Key:
0: Not a shibbolethnonym
1: This
2: No

	Mean	Effective Sample Size
Lh	-98.458	88220
Harmonic Mean	-101.741	27
q01	5.84E-02	89000
q02	6.23E-02	89000
q10	0.177	88215
q12	7.83E-02	89000
q20	0.161	85844
q21	3.72E-02	85584
Root P(0)*	0.6	89000
Root P(1)*	0.237	89000
Root P(2)*	0.164	89000
Durubal P(0)	9.84E-04	89000
Durubal P(1)	1.80E-03	89000
Durubal P(2)	0.997	89000
CentralNSW P(0)	6.92E-02	88859
CentralNSW P(1)	3.84E-02	89000
CentralNSW P(2)	0.892	88953
Maric P(0)	0.676	88318
Maric P(1)	6.71E-02	89000
Maric P(2)	0.257	86590
Kulin P(0)	2.77E-03	89000
Kulin P(1)	5.65E-03	88474
Kulin P(2)	0.992	88904
Western P(0)	0.938	89000
Western P(1)	3.22E-02	89000
Western P(2)	2.95E-02	89000

East P(0)	2.75E-02	87618
East P(1)	9.23E-03	89000
East P(2)	0.963	88648
North P(0)	0.533	89000
North P(1)	0.378	88450
North P(2)	8.89E-02	89000
Ngumpi P(0)	0.962	86953
Ngumpi P(1)	2.08E-02	88127
Ngumpi P(2)	1.73E-02	89000
Wakakabi P(0)	0.2	88300
Wakakabi P(1)	5.43E-02	89000
Wakakabi P(2)	0.746	88320
Wati P(0)	0.942	86378
Wati P(1)	3.33E-02	86616
Wati P(2)	2.49E-02	89000
Mayi P(0)	0.159	87928
Mayi P(1)	0.83	88035
Mayi P(2)	1.12E-02	84446
Pilbara P(0)	0.915	86950
Pilbara P(1)	4.60E-02	88001
Pilbara P(2)	3.94E-02	89000
ThuraYura P(0)	0.832	86988
ThuraYura P(1)	9.02E-02	87815
ThuraYura P(2)	7.77E-02	89000
Central P(0)	0.96	86896
Central P(1)	2.20E-02	88063
Central P(2)	1.84E-02	89000
Kaltu P(0)	0.29	88584
Kaltu P(1)	0.107	89000
Kaltu P(2)	0.602	88686
Yolngu P(0)	5.77E-02	89000
Yolngu P(1)	0.936	89000
Yolngu P(2)	6.35E-03	89000
Yuinkuri P(0)	0.752	89000
Yuinkuri P(1)	0.128	89000
Yuinkuri P(2)	0.12	89000
Paman P(0)	0.383	87782
Paman P(1)	0.589	87879
Paman P(2)	2.72E-02	86176
Karnic P(0)	0.951	86897
Karnic P(1)	2.69E-02	88000
Karnic P(2)	2.24E-02	89000

Table 12: Averaged Reconstructions of Speaker-Descriptive Strategies

Key:
0: Not Speaker-Descriptive
1: People
2: Totem
3: Geographic

	Mean	Effective Sample Size
Lh	-148.308	89000
Harmonic Mean	-151.909	26
q01	0.211	86157
q02	3.30E-02	89000
q03	5.14E-02	89000
q10	0.215	89000
q12	5.69E-02	84142
q13	0.144	88489
q20	0.15	88546
q21	0.155	86930
q23	0.158	89000
q30	0.139	89000
q31	0.173	88507
q32	0.115	89000
Root P(0)	0.455	88800
Root P(1)	0.338	89000
Root P(2)	7.74E-02	87307
Root P(3)	0.129	89000
Durubal P(0)	0.98	89000
Durubal P(1)	9.35E-03	89000
Durubal P(2)	5.55E-03	88702
Durubal P(3)	4.97E-03	89000
CentralNSW P(0)	0.83	89000
CentralNSW P(1)	7.72E-02	89000
CentralNSW P(2)	4.80E-02	88759
CentralNSW P(3)	4.45E-02	89000
Maric P(0)	2.43E-02	89000
Maric P(1)	0.122	88479
Maric P(2)	7.21E-02	89000
Maric P(3)	0.781	88862
Kulin P(0)	0.858	89000
Kulin P(1)	8.08E-02	89000
Kulin P(2)	3.14E-02	89000
Kulin P(3)	2.98E-02	89000
Western P(0)*	0.188	88146
Western P(1)*	0.557	87586
Western P(2)*	0.151	88043
Western P(3)*	0.104	88610
East P(0)	0.796	89000
East P(1)	0.121	89000

East P(2)	3.83E-02	88819
East P(3)	4.47E-02	89000
North P(0)	0.237	89000
North P(1)	0.29	89000
North P(2)	0.193	87848
North P(3)	0.28	89000
Ngumpi P(0)	0.163	88284
Ngumpi P(1)	0.635	87950
Ngumpi P(2)	6.36E-02	86826
Ngumpi P(3)	0.138	89000
Wakakabi P(0)	0.222	89000
Wakakabi P(1)	0.126	87713
Wakakabi P(2)	0.376	86628
Wakakabi P(3)	0.277	88475
Wati P(0)	0.843	89000
Wati P(1)	9.01E-02	89000
Wati P(2)	3.45E-02	88876
Wati P(3)	3.27E-02	89000
Mayi P(0)	0.499	89000
Mayi P(1)	0.414	89000
Mayi P(2)	4.28E-02	88337
Mayi P(3)	4.41E-02	89000
Pilbara P(0)	7.01E-02	89000
Pilbara P(1)	0.505	88314
Pilbara P(2)	9.98E-02	86272
Pilbara P(3)	0.325	88711
ThuraYura P(0)	0.11	87440
ThuraYura P(1)	0.347	87720
ThuraYura P(2)	0.383	88126
ThuraYura P(3)	0.161	88555
Central P(0)*	0.124	89000
Central P(1)*	0.488	87683
Central P(2)*	0.159	88017
Central P(3)*	0.229	88304
Kaltu P(0)*	7.83E-02	86410
Kaltu P(1)*	0.23	86165
Kaltu P(2)*	0.547	88274
Kaltu P(3)*	0.145	88722
Yolngu P(0)	0.522	89000
Yolngu P(1)	0.372	89000
Yolngu P(2)	5.24E-02	88350
Yolngu P(3)	5.33E-02	89000
Yuinkuri P(0)	0.195	88972
Yuinkuri P(1)	0.45	89000
Yuinkuri P(2)	8.74E-02	86846
Yuinkuri P(3)	0.267	89000

Paman P(0)	0.324	88637
Paman P(1)	0.429	89000
Paman P(2)	8.48E-02	87212
Paman P(3)	0.162	89000
Karnic P(0)*	0.101	88800
Karnic P(1)*	0.305	88672
Karnic P(2)*	0.101	87595
Karnic P(3)*	0.493	88424

Bibliography

Allerton, David J. 1987. The linguistic and sociolinguistic status of proper names: What are they, and who do they belong to? *Journal of Pragmatics* 11(1). 61–92.

Amery, Rob. 2000. *Warrabarna Kaurna! reclaiming an Australian language.* Exton, PA: Swets and Zeitlinger Publishers.

Ash, Anna. 2002. Placenames in Yuwaalaraay, Yuwaaliyaay and Gamilaraay languages of North-West New South Wales. In Luise Hercus, Flavia Hodges & Jane Simpson (eds.), *The land is a map*, chap. 13, 181–185. Canberra: Australian National University Press.

Austin, Peter. 1992. *A dictionary of Thalanyji, Western Australia.* Bundoora, Australia: La Trobe University, Department of Linguistics.

Black, Paul. 2004. The failure of the evidence of shared innovations in Cape York Peninsula. In Claire Bowern & Harold Koch (eds.), *Australian languages: Classification and the comparative method* (Current Issues in Linguistic Theory 249), 241–268. Philadelphi: John Benjamins Publishing Company.

Blake, Barry J. 1988. Redefining Pama Nyungan: Towards the prehistory of Australian languages. In Nicholas Evans & Steve Johnson (eds.), *Aboriginal linguistics*, 1–85. Armidale, Australia: University of New England, Department of Linguistics.

Blake, Barry J. 1991. Woiwurrung, the Melbourne language. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 4, 30–122. John Benjamins Publishing Company.

Blake, Barry J., Luise Hercus, Stephen Morey & Edward Ryan. 2011. *The Mathi group of languages* (Pacific Linguistics 628). Canberra: Australian National University.

Blench, Roger & Matthew Spriggs. 1998. *Archaeology and language: Correlating archaeological and linguistic hypotheses.* Psychology Press.

Blomberg, Simon P. & Theodore Garland. 2002. Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* 15(6). 899–910.

Bouckaert, Remco R., Claire Bowern & Quentin D. Atkinson. 2018. The origin and expansion of Pama-Nyungan languages across Australia. *Nature Ecology & Evolution* .

Bouckaert, Remco R., Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960. doi: 10.1126/science.1219669.

Bowern, Claire. 2000. Nhirrpi sketch grammar. Unpublished manuscript.

Bowern, Claire. 2010. Historical linguistics in Australia: Trees, networks and their implications. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1559). 3845–3854. doi:10.1098/rstb.2010.0013.

Bowern, Claire & Quentin D. Atkinson. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88(4). 817–845. doi:10.1353/lan.2012.0081.

Breen, Gavan. 1981. *The Mayi languages of the Queensland Gulf Country* (Aias New Series 29). Acton, Australia: Australian Institute of Aboriginal Studies.

Breen, Gavan. 1990. *Salvage studies of Western Queensland Aboriginal languages* (Pacific Linguistics 105). Canberra: Australian National University.

Bromham, Lindell. 2017. Curiously the same: Swapping tools between linguistics and evolutionary biology. *Biology & Philosophy* doi:10.1007/s10539-017-9594-y.

Bryant, David & Vincent Moulton. 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265.

Buckley, Christopher D. 2012. Investigating cultural evolution using phylogenetic analysis: The origins and descent of the Southeast Asian tradition of warp ikat weaving. *PLOS ONE* 7(12). e52064. doi:10.1371/journal.pone.0052064.

Crowley, Terry. 1981. The Mpakwithi dialect of Anguthimri. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 2, John Benjamins Publishing Company.

Crowley, Terry. 1983. Uradhi. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 3, John Benjamins Publishing Company.

Currie, Thomas E. 2013. Cultural evolution branches out: The phylogenetic approach in cross-cultural research. *Cross-Cultural Research* 47(2). 102–130.

Dench, Alan Charles. 1987. *Martuthunira: A language of the Pilbara region of Western Australia*. Canberra: Australian National University Phd.

Dixon, Robert M. W. 1983. Nyawaygi. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 3, John Benjamins Publishing Company.

Dixon, Robert M. W. & Barry J. Blake (eds.). 1979–2000. *Handbook of Australian languages*, vol. 1-5. John Benjamins Publishing Company.

Dixon, Robert M. W. & Barry J. Blake (eds.). 1981. *Handbook of Australian languages*, vol. 2. John Benjamins Publishing Company.

Eather, Bronwyn. 2005. *A first dictionary of Na-Kara*. Winnellie, Australia: Maningrida Arts and Culture.

Evans, Nicholas. 2003. *Bininj Gun-wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune*, vol. 1. Pacific Linguistics.

Fritz, Susanne A. & Andy Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24(4). 1042–1051.

Gaby, Alice Rose. 2006. *A grammar of Kuuk Thaayorre*. Melbourne, Australia: University of Melbourne dissertation.

Geytenbeek, Brian & Helen Geytenbeek. 1971. *Gidabal grammar and dictionary* (Australian Aboriginal Studies 43). Canberra: Australian Institute of Aboriginal Studies.

Glass, Amee & Dorothy Hackett. 2003. *Ngaanyatjarra & Ngaatjatjarra to English dictionary*. Iad Press.

Godman, Irene. 1993. *A sketch grammar of Rimanggudinhma, a language of the Princess Charlotte Bay region of Cape York Peninsula*. The University of Queensland MA thesis.

Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1559). 3923–3933. doi:10.1098/rstb.2010.0162.

Gray, Russell D., Alexei J. Drummond & Simon J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *science* 323(5913). 479–483.

Gray, Russell D., Simon J. Greenhill & Robert M. Ross. 2007. The pleasures and perils of Darwinizing culture (with Phylogenies). *Biological Theory* 2(4). 360–375. doi:10.1162/biot.2007.2.4.360.

Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B: Biological Sciences* 276(1665). 2299–2306. doi:10.1098/rspb.2008.1944.

Hale, Kenneth. 1966. The Paman group of the Pama-Nyungan phylic family. *Anthropological Linguistics* 8(2).

Hall, Allen Harry. 1972. *A study of the Thaayorre language of the Edward River Tribe Cape York Peninsula, Queensland*. Brisbane, Australia: University of Queensland dissertation.

Hall, Harold Aubrey. 1971. *A partial vocabulary of the Ngalooma Aboriginal tribe* (Australian Aboriginal Studies 18). Canberra: Australian Institute of Aboriginal Studies.

Haynie, Hannah J. & Claire Bowern. 2016. Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Sciences* 113(48). 13666–13671. doi:10.1073/pnas.1613666113.

Hercus, Luise. 1999. *A grammar of the Wirangu language from the West Coast of South Australia* (Pacific Linguistics 150). Canberra: Australian National University.

Hercus, Luise & Peter Austin. 2004. The Yarli languages. In Claire Bowern & Harold Koch (eds.), *Australian languages: Classification and the comparative method*, John Benjamins Publishing Company.

Holland, Barbara R., Katharina T. Huber, Andreas Dress & Vincent Moulton. 2002. δ plots: A tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution* 19(12). 2051–2059.

Howitt, Alfred William. 1904. *The native tribes of South-East Australia*. New York: The Macmillan Company.

Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267.

Jones, Barbara. 2011. *A grammar of Wangajunga–A language of the Great Sandy Desert of North Western Australia* (Pacific Linguistics 636). Canberra: Australian National University.

Jordan, Fiona M. 2013. Comparative phylogenetic methods and the study of pattern and process in kinship. In Patrick McConvell, Ian Keen & Rachel Hendery (eds.), *Kinship Systems: Change and reconstruction*, 43–58. Salt Lake City, UT: University of Utah Press.

Jordan, Fiona M. & Brad R. Huber. 2013. Evolutionary approaches to cross-cultural anthropology. *Cross-Cultural Research* 47(2). 91–101.

Kite, Suzanne & Stephen Wurm. 2004. *The Duungidjawu language of south-east Queensland: Grammar, texts and vocabulary* (Pacific Linguistics 553). Canberra: Australian National University.

Koch, Harold. 2014. Historical relations among the Australian languages: genetic classification and contact-based diffusion. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia: A comprehensive guide*, vol. 3 The World of Linguistics, chap. 2, 23–90. De Gruyter, Inc.

Koopman, Adrian. 2016. Ethnonyms. In Carol Hough (ed.), *The Oxford handbook of names and naming*, Oxford, UK: Oxford University Press.

Léglise, Isabelle & Bettina Migge. 2006. Language-naming practices, ideologies, and linguistic practices: Toward a comprehensive description of language varieties. *Language in Society* 35(3). 313–339.

Lusakalalu, Pedro. 2003. What is Rukavango? *The Nordic Journal of African Studies* 12(1). 92–104.

Maddison, David R. & Wayne P. Maddison. 2000. *Macclade 4*. Sunderland, MA: Sinauer Associates, Inc.

McConvell, Patrick. 1996. Backtracking to Babel: The chronology of Pama-Nyungan expansion in Australia. *Archaeology in Oceania* 31(3). 125–144. doi:10.1002/j.1834-4453.1996.tb00356.x.

McConvell, Patrick. 2006. Shibbolethnonyms, ex-exonyms and eco-ethnonyms in Aboriginal Australia: The pragmatics of onymization and archaism. *Onoma* 41. 185–214.

McConvell, Patrick & Claire Bowern. 2011. The prehistory and internal relationships of Australian languages. *Language and Linguistics Compass* 5(1). 19–32. doi:10.1111/j.1749-818X.2010.00257.x.

McGregor, William B. 1993. *Gunin/kwini* Languages of the World. LINCOM publishers.

McKay, Graham. 2000. Ndjébbana. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 5, John Benjamins Publishing Company.

Meakins, Felicity & Rachel Nordlinger. 2013. *A grammar of Bilinarra: An Australian Aboriginal language of the Northern Territory* (Pacific Linguistics 640). De Gruyter Mouton.

Mesoudi, Alex. 2011. *Cultural evolution: How Darwinian theory can explain human culture & synthesize the social sciences*. Chicago: University of Chicago Press.

Moore, Robert. 2017. Language and materiality in the renaming of indigenous North American languages and peoples. In Jillian R. Cavanaugh & Shalini Shankar (eds.), *Language and materiality: Ethnographic and theoretical explorations*, 204–225. Cambridge University Press.

Morphy, Frances. 1983. Djapu, a Yolngu dialect. In Robert M. W. Dixon & Barry J. Blake (eds.), *Handbook of Australian languages*, vol. 3, John Benjamins Publishing Company.

Nunn, Charles. 2011. Basic phylogenetic concepts and "tree thinking". In *The comparative approach in evolutionary anthropology and biology*, 20–51. Chicago: University of Chicago Press.

O'Grady, Geoffrey. 2009. Warriyangka wordlist.

Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on pjhylogenies. *Systematic Biology* 53(5). 673–684.

Rambaut, Andrew, Alexei J. Drummond, D. Xie, Guy Baele & Marc A. Suchard. 2018. Tracer v1.7. Available from http://beast.community/tracer.

Rigsby, Bruce. n.d. Kuku Thaypan/Awu Alaya. Unpublished manuscript.

Rumsey, Alan. 1993. Language and territoriality in Aboriginal Australia. In Michael Walsh & Colin Yallop (eds.), *Language and culture in Aboriginal Australia*, 191–206. Canberra: Aboriginal Studies Press.

Schultze-Berndt, Eva Friederike. 2000. *Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language*: Katholieke Universiteit Nijmegen dissertation.

Sharp, Janet. 2004. *Nyangumarta: A language of the Pilbarra region of Western Australia* 556. Pacific Linguistics.

Sommer, Bruce A. 2006. *Speaking Kunjen: An ethnography of Oykangand kinship and communication* Pacific Linguistics. Canberra: Australian National University.

Sutton, Peter. 1979. Australian language names. In Stephen Wurm (ed.), *Australian linguistic studies*, 87–105. Canberra: Pacific Linguistics.

Sutton, Peter John. 1973. *Gugu Badhun and its neighbors*. Sydney: Macquarie University M.A.

Terrill, Angela. 2002. *Dharumbal: The language of Rockhampton, Australia* (Pacific Linguistics 525). Canberra: Australian National University.

Thompson, David. 1988. *Lockhart River "sand beach" language : an outline of Kuuku Ya'u and Umpila* Work papers of SIL-AAIB. Darwin, Australia: Summer Institute of Linguistics.

Tindale, Norman. 1974. Tindale's catalogue of Australian Aboriginal tribes.

Tsunoda, Tasaku. 2011. *A grammar of Warrongo* (Mouton Grammar Library 53). De Gruyter Mouton.

Wafer, Jim & Amanda Lissarague. 2008. *A handbook of Aboriginal languages of New South Wales and the Australian Capital Territory*. Muurbay Aboriginal Language and Culture Co-operative.

Wohlgemuth, Jan. 2015. Language naming principles. In *Diversity Linguistics: Retrospect and Prospect*, .