

DENASALIZATION IN EARLY AUSTRONESIAN

by

WILLIAM J. LACKEY

(Under the Direction of Jared Klein)

ABSTRACT

Recent and ongoing work in the reconstruction of the Proto-Austronesian lexicon has shown an unusual paucity of stems that have an initial nasal consonant. There are a number of possible reasons for this, but one is denasalization, or the loss of the nasal contrast due to sound change. I will present evidence from the typology of denasalization, from internal variation within Austronesian, and from external comparanda that have initial nasals that are lacking in the corresponding Proto-Austronesian word. These together will provide strong support for the hypothesis of sound change as an explanation for the distribution of nasals in Proto-Austronesian. The presence of denasalization in Austronesian's early history has important implications both for the study of sound change in general, and for comparative linguistics in Southeast Asia as well.

INDEX WORDS: Proto-Austronesian, Austronesian, Formosan languages, Denasalization, Nasals, Sound change, Historical linguistics, Southeast Asian languages

DENASALIZATION IN EARLY AUSTRONESIAN

by

WILLIAM J. LACKEY

B.A., University of Georgia, 2017

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2019

© 2019

William J. Lackey

All Rights Reserved

DENASALIZATION IN EARLY AUSTRONESIAN

by

WILLIAM J. LACKEY

Major Professor:	Jared Klein
Committee:	Margaret Renwick
	Mark Wenthe

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

ACKNOWLEDGMENTS

Thank you to my family and my friends for their continuing support, without which I could not have achieved such an important step in my life.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
CHAPTER	
1 Introduction.....	1
2 Denasalization Cross-linguistically	9
3 Internal Variation in the Presence of Initial Nasals.....	13
*m-initial words	19
*n-initial words	28
*ñ-initial words	32
*ŋ-initial words	33
4 External Evidence for Initial Nasals	36
5 Discussion	43
6 Conclusion	48
REFERENCES	50
APPENDIX	
Proto-Austronesian Nasal-initial Stems in the ACD.....	54

CHAPTER 1

Introduction

When it comes to the reconstruction of Proto-Austronesian (PAN), the phoneme inventory is well-established; as Malcolm Ross states, the vowel system and the majority of the consonant system are more or less uncontroversial (Ross 1992: 30-31). However, a phonology does not consist merely in a list of phonemes, but also contains a number of important phonotactic or distributional constraints. This aspect of historical phonology has long been a topic of interest in the reconstruction of Proto-Austronesian. For instance, Chrétien (1965) takes a statistical approach to the identification of Proto-Austronesian phonotactics. This work has been very important in the development of our understanding of Austronesian history, but is insufficient for modern Austronesian studies, because of its usage of Dempwolff's reconstruction of PAN, which is now obsolete. In particular, the inability to identify morphological boundaries from the languages used for reconstruction meant that Chrétien could not identify some interesting constraints in PAN phonotactic structure. One that is fairly interesting is the rarity of initial nasals in reconstructed words (Blust 2013: 214), which was not possible to identify in early reconstructions because many instances of initial nasals are actually morphologically complex, prefixed forms (see below as well for more discussion of this issue).

What conclusions can we make from highly distinctive phonotactic or distributional constraints such as this? Cross-linguistically, nasals are very common

(Maddieson 2013), and so the absence of nasals in some particular position seems to demand an explanation. The form of this explanation could plausibly be in terms of a hypothesis about sound change that occurred in the language before the point we can access through comparison of available data. Within Austronesian, this move has been made in explaining the distribution of *g and *j in the proto-language: because these two phonemes are in complementary distribution, with *g only appearing initially and *j only appearing post-vocally (*j appears to have been a voiced velar), the most likely explanation is that they were originally the same phoneme (Ross 1992: 32). In this case we do not have the 'pre-Proto-Austronesian' data to verify this hypothesis, but we can deduce its validity from the phonotactic/distributional facts we have available in the Proto-Austronesian data. Another example we may consider is the case of Sanskrit. Suppose that instead of having the enormous amount of knowledge we have about Indo-European history, we were not aware of the genetic origin of the Sanskrit language. Even if we did not know this information, some distributional facts about Sanskrit phonemes would demand explanation. If anyone took a cursory look at the Sanskrit lexicon, they would notice the extreme overrepresentation of the phoneme /a/, and they may hypothesize that this situation arose from a merger of many different syllable nuclei. Indeed, this hypothesis would be correct--in the real world we do know a lot about the history of Sanskrit, and as a result of its Indo-Iranian heritage, the vowel /a/ derives from several different Proto-Indo-European nuclei including *(h₁)e, *h₂e, h₃e, *o, and syllabic *m and *n (Lubotsky 2018: 1876). Such a comprehensive merger is sure to result in overrepresentation of its outcome, and it is not illogical to run this logic backwards and infer from the distributional facts of Sanskrit that a merger of many nuclei must have occurred,

even if we cannot necessarily know what the value of the original vowels were from just the Sanskrit data. Inferences like these can be understood as a kind of 'internal reconstruction' approach to historical phonology, although that technique is usually applied to morphological idiosyncrasies.

We can make inferences like this because we have knowledge about what kinds of sound changes tend to occur in languages, and we also know what languages tend to look like. Thus, if a language differs phonologically from the typological norm we may reasonably make an inference that this was a result of sound change, and some earlier stage had a distribution of sounds that was more typical. For Proto-Austronesian, we would then propose a change where nasals changed into non-nasal sounds word initially. In proposing this change, there are a number of issues we need to deal with. First of all, the proposal of this change conflicts with the typological tendencies we have previously discussed with regard to the presence of nasals, since it would reduce their frequency in a language. Additionally, it would add more plausibility to the proposal of such a change if we could actually make conclusions about what the outcome of the change is.

To address these issues I will begin by discussing other cases where we know that denasalization of earlier nasal consonants has occurred. External cases of denasalization by nature make a proposal of denasalization in Proto-Austronesian more plausible; however, by looking at the similarities between these cases we can bolster this increased plausibility by finding a phonetic/phonological rationale for initial denasalization to occur. This allows us to circumvent the typological issues, because we can say that even if the change might be rare it has a reasonable basis in the way language behaves generally.

After this, I will proceed to analyze the lexical data that provides direct evidence of denasalization in early Austronesian. The first step in this analysis is an examination of the words reconstructed in PAN with initial nasals. We will find that in these words, there are a number of explanations for the presence of an initial nasal. For a few words, I will claim that they were probably not present in PAN and so do not bear on the issue of early PAN phonological structure. For others, the initial nasal is due to their status as morphologically derived. Another group is words that seem to have a sound-symbolic or onomatopoeic character; for these words the presence of initial nasals is probably due to restoration of their earlier form (cf. the section on words with an initial velar nasal). The remaining forms almost all show variation in the presence of an initial nasal, which alternates with an initial voiced stop or, in the case of coronals, occasionally alternates with initial *l. This last group points to denasalization as a sound change in early Austronesian. In this comparative study of a segment of the PAN lexicon, I will often find it useful to employ the distinction between Formosan languages and Malayo-Polynesian languages (abbreviated MP, with PMP = Proto-Malayo-Polynesian). The former term refers to all the Austronesian languages of Taiwan except for Yami, whereas Malayo-Polynesian is a single subgroup including everything else (Ross 1992: 25). The term Formosan is useful because subgrouping there is unclear except for some very close-knit branches like Atayalic; regardless, any reconstruction of a word in Proto-Austronesian must make use of at least some Formosan evidence. Any subgroups of Austronesian mentioned, unless otherwise specified, will be subgroups of Malayo-Polynesian.

The second step in this analysis is to examine some words in PAN which are shared with other language families, in which the external representative has an initial

nasal and the Austronesian forms have an initial voiced stop. It is generally agreed that there are a number of words shared between Austronesian and other language families such as Kra-Dai, and the similarities between their lexicons are not due to chance (Thurgood 1994: 348). When too many lexical items are shared between two families for chance to be a valid explanation, the two remaining explanations are genetic relationship or borrowing. However, it is important to note that my proposal that denasalization occurred in early Austronesian does not logically depend on any claim for higher-order genetic relationships between Austronesian and other language families. If any kind of 'macrophyllum' existed between Austronesian and other language families of Southeast Asia, then the Austronesian words would be native and sound change would simply apply normally. If the nearby language families are not related to Austronesian, then the alternative of borrowing does not pose serious issues either, since the words could have been borrowed before denasalization occurred. The hypothesis of very early contact (i.e. before the breakup of Proto-Austronesian) between Austronesian and Kra-Dai, for example, is accepted by Thurgood (1992: 361). After this borrowing occurred, denasalization would occur in Austronesian and apply to the lexical item just as it would to a word that was ultimately inherited rather than borrowed. The purpose of the external comparanda then are to provide evidence otherwise unavailable that the original Austronesian form must have been nasal-initial at a stage earlier than Proto-Austronesian.

Addressing the external evidence for denasalization will require us to make use of reconstructed forms for language families that are nearby. In general, I will draw my lexical data from reference texts for each language family that is relevant. For the Kra-Dai family, there is no contemporary reconstruction of the whole family's proto-language,

so we will cite reconstructions from individual subgroups to make the case for much earlier forms. For the Tai subgroup, the reference will be Pittayawat Pittayaporn's 2009 dissertation *The Phonology of Proto-Tai*; for the Kra subgroup the reference will be Weera Ostapirat's 2000 book *Proto-Kra*; and for the Hlai subgroup we will use Peter Norquest's recent work in 2016, *A Phonological Reconstruction of Proto-Hlai*. For the Austroasiatic family, we will use *A Mon-Khmer Comparative Dictionary* by Shorto (2006). Otherwise, I will use the lexical data in online databases, like the STEDT for Sino-Tibetan reconstructions, the ABVD for contemporary forms when reconstructions are not available (in the bibliography as Greenhill et al. 2008), and Proto-Austronesian reconstructions will be taken from the online Austronesian Comparative Dictionary (ACD). In taking words from these sources, we will preserve the orthography that they use in each separate text, although the sources conflict with each other on some points of orthography. In particular, texts on Kra-Dai languages often use letters A, B, C, and D after a word to indicate that it had some particular tone. They also use the capital C within a word to indicate an unspecified consonant. In contrast, *C is used in the context of Proto-Austronesian to refer to a specific reconstructed proto-phoneme. The one exception to this choice is that we will differ from the ACD by writing *L instead of *N for one proto-phoneme. This is because Blust's *N is more or less universally agreed upon by Austronesian scholars to *not* have originally been a nasal phoneme, and since the difference between a sound being nasal or not a nasal is directly at issue in this thesis, it is important to avoid confusion about a sound on this topic. Finally, all of these sources are either searchable databases, or have an appendix where all the lexical data is listed, and so they will not be cited explicitly in the text. The two exceptions are when lexical

data is cited from an alternate source (if it is not available in the just listed texts), or for any Proto-Hlai forms listed, because Norquest's book lacks an index of reconstructions.

After considering this evidence that the sound change occurred, I will take a broad scope look at the change and its context, concluding that the form of the sound change was specifically a shift of initial nasals to initial voiced stops. There are some cases of correspondence between initial coronal nasals and *l, among other issues, but these will not be found to be seriously problematic, as there are other factors at play for these words. I will also discuss the likely time frame for the loss of the nasality contrast initially. I will conclude that since the data indicates that the sound change must have been in operation relatively recently, before the split-up of Proto-Austronesian, morphological processes reintroducing nasal-initial bases may have stopped the sound change from occurring.

Finally, I will conclude by discussing the implications of these results for historical Austronesian studies in general. Some earlier works suggest that distinctions apparently reconstructible for Proto-Austronesian may actually be later developments, or that validly reconstructible distinctions in PAN derive from allophony that has been obscured by morphological developments. The argument I present in this thesis provides a similar contribution to historical Austronesian studies, as it explains an unusual distributional fact of the reconstructed language. We thus discover that an earlier stage of the language had a more even distribution of phonemes. Being able to explain many of the odd facts about a proto-language's phonology is interesting in its own right, but this information is also relevant to hypotheses about deep comparative relationships. Comparison between Austronesian and other language families is of interest to many

scholars, and the similar lexicon is a clear reason why. Any comparison between the Austronesian language family and another family must either take into account the earliest stage of Proto-Austronesian, or propose that the other language family is a subgroup of Austronesian, not a sibling. Thus, careful investigation of these features is interesting for many aspects of historical and comparative Austronesian study.

CHAPTER 2

Denasalization Cross-linguistically

Nasal consonant phonemes are overwhelmingly represented in the world's languages, being completely lacking in only about 1.8% of them (Maddieson 2013). Thus we should not expect changes to result in the loss of nasal phonemes to be very common, because that would mean the expected proportion of languages lacking nasal phonemes should be higher. However, there are a number of processes resulting in the loss of a nasal feature in a phoneme, most of which are conditioned sound changes; conditioned loss of nasality would preserve nasals as phonemes but reduce their frequency. Indeed, when we survey the cases of wholesale loss of nasality that are known in linguistic history, there are very few cases, the two most notable being the areal shift of nasal consonants to voiced stops in the Pacific Northwest (Kinkade 1985), and the same shift of nasals to voiced obstruents in the Central variety of Rotokas (Robinson 2006: 207-208). Instead, what we find to have occurred in the Southeast Asian linguistic area, and to some degree in Asia as a whole, is a frequent loss of the oral-nasal contrast only in syllable initial position. This has happened in a couple of Sinitic languages; for instance, in Southern Min varieties all initial nasals became non-nasal unless the final was also nasal (Norman 1991: 352), and in the Hong Kong variety of Cantonese, there is an ongoing shift of initial /n/ to /l/ (Zee 1999: 121).

The hypothesized change for early Austronesian matches up better with Southern Min, but we do not have the same synchronic data about variation, phonetics, etc. that we

do with the Cantonese change, because the change in Southern Min occurred quite some time ago. However, we do have a case of ongoing denasalization in some varieties of Korean. As we will see, this is a particularly helpful case for the hypothesis examined here, because unlike other languages of East Asia in which we can find precedent, Korean does not have a generally 'monosyllabic' word shape, and so is structurally more similar to Proto-Austronesian. In addition, the evidence from Korean shows a strong correlation between prosodic context and the degree of loss of nasality.

The center of our analysis of Korean will be the article "Phonetic implementation of Korean 'denasalization' and its variation related to prosody" (Yoshida 2008). One difference between Korean and Proto-Austronesian pertaining to this issue is that Korean lacks an initial velar nasal (p.1), but the initial velar nasal in Proto-Austronesian almost always appears in the context of a specific sound symbolic phenomenon (see below). It is therefore not central to our analysis of early Austronesian history. Yoshida observes that in Korean, nasality occurs on a spectrum, and the degree of loss of nasality is correlated strongly with prosodic context. In particular, denasalization/nasal weakening does not occur in word-internal position (p.12). In contrast, it occurs somewhat more frequently in word-initial and verb-phrase initial position, and very commonly in utterance and clause-initial position (p.12). In this regard, the degree of loss of nasality corresponds to the location of the nasal after progressively stronger prosodic boundaries; the least degree of nasality thus occurs after a pause, in utterance-initial position. However, since by the time of Proto-Austronesian denasalization was complete, whereas in Korean it is not complete by any means--Yoshida claims that no tokens in the data observed completely lose their nasality (p.12)--we may want to investigate possible reasons why the sound change

appears to be mainly represented in word-initial position, which in a Korean context is the least strong environment where denasalization/nasal weakening may occur. That is, for denasalization to occur at all in a Korean context, the nasal must at least be word-initial.

In regard to this issue, it is important to recognize how these different categories overlap in Korean versus Proto-Austronesian. Korean, being an SOV language, always has the verb phrase in post-subject position (Yoshida 2008: 4). The syntactic profile of Proto-Austronesian, in the canonical understanding, was quite different from this. While the relationship of syntax to argument structure is complex, there are several basic facts about PAN syntax that are important here. First of all, the basic word order is verb-initial (Starosta, Pawley, & Reid 2009: 339), and nominal or prepositional predicates were also in the initial position of clauses (id., 351). The implications of these syntactic facts are interesting for the previously discussed correlation of prosodic strength with loss of nasality. Since any content word regardless of word class would plausibly be in strict utterance-initial position, the prosodic factors motivating loss of nasality would apply strongly to the lexicon in general. This explains why even though there are signs that the sound change was still in progress after the separation of Proto-Austronesian into different languages, the direct lexical evidence is low in number. Because the strongest prosodic context that motivates nasal loss and increases its frequency would have been a typical context for (content-bearing) lexical items of all types to occur, the sound change should have spread fairly rapidly and left only a small residue identifiable in the modern Austronesian languages.

The varying outcomes of denasalization in Korean and Sinitic varieties mean that we cannot automatically deduce from the fact that the input of a sound change is nasal and the output is not, that the sound change must result in some particular type of segment. It is possible that the output of the sound change could be stops, fricatives, approximants, etc., or these types of sounds could be mixed. We have a mixed situation in Southern Min varieties, where *m and *ŋ became stops, but *n became a lateral approximant /l/ (Norman 1991: 352). Rather than being a single specific sound change, denasalization should be understood as an umbrella term for a set of related sound changes, and so the following sections examining the specific evidence in Austronesian will do double duty in my argument. To begin with, they will provide general evidence of a loss of the nasality distinction in word-initial position. Secondly, they will provide evidence establishing what type of sound change occurred in early Austronesian, namely whether it was of a Korean type or a Sinitic type.

CHAPTER 3

Internal Variation in the Presence of Initial Nasals

The regularity of sound change is an important principle of historical linguistics, and so if we intend to propose that a sound change occurred in a language, but the proposed sound change is apparently contradicted by the modern reflexes of a sound, then we need to explain what the source of this contradiction is. It is known that the principle of regularity can be violated in certain specific ways. For instance, sound changes as events in the real world do not occur instantaneously; they take a certain amount of time to come to completion and disseminate throughout the lexicon and the speech community. If a sound change was in the middle of spreading through the lexicon, and some other factor caused the sound change to stop spreading, then it may appear to be irregular from the perspective of the whole lexicon. Onomatopoeia often appears to resist sound change as well, although a more likely explanation is that onomatopoetic forms are restored after sound change causes them to lose their iconic and sound symbolic qualities (Hock & Joseph 2009: 223). If some words with initial nasals in Proto-Austronesian are onomatopoetic or sound-symbolic in nature, this would then be a good explanation of why they have 'resisted' this sound change--they were simply restored to their original form or reinnovated.

We will show that for the word stems reconstructed as nasal-initial in the ACD, there is some variation in modern Austronesian languages with regard to the presence of an initial nasal. Interestingly, the explanation for this variation seems to be different

depending on the place of articulation--words with initial *m that are genuinely reconstructible for Proto-Austronesian (I disagree with some reconstructions--see below) are covertly morphologically complex, or have other phonological idiosyncrasies. In contrast, words with initial nasals at other places of articulation appear to show genuine denasalization in a sizeable number of cases. Due to this difference, we will separate the lexical evidence by place of articulation. However, it should not be taken from this that there is a concrete difference here in the *phonological* history with respect to place of articulation (see below).

This morphological complexity is certainly one part of why the constraint against initial nasals has been difficult to identify. The last systematic, detailed analysis of PAN phonotactics was that of Chrétien. Because he used Dempwolff's early reconstruction of PAN, he was not able to extract morphology like the stative prefix *ma- from a large number of bases. When bases are extracted from these morphologically complex stative forms, this removes a very large number of the nasal-initial reconstructions proposed by Dempwolff. However, there are a number of other morphological processes that result in nasal-initial base allomorphs; this would remove signs of an earlier distributional constraint to some degree. A useful inventory of these processes is provided in Robert Blust 2004, "Austronesian Nasal Substitution: A Survey". Its main topic is the process usually referred to as nasal substitution (henceforth abbreviated as NS in this thesis), in which a stem-initial obstruent is replaced by a nasal at the same place of articulation when a word is prefixed with *maŋ- or *paŋ- (Blust 2004: 73).

One issue with identifying NS as a process that applies at the Proto-Austronesian (PAN) level is that it seems not to be productively present in any Austronesian language of Taiwan (the 'Formosan languages'). As Blust puts it,

"Such examples of apparently fossilized NS in languages that have no active counterpart are rare, and generally difficult to interpret. Is Puyuma *maŋayaw* 'go headhunting' the lone survivor of a once active process of NS in Formosan languages (hence in PAN)? Or did speakers of PMP [Proto-Malayo-Polynesian] reanalyze a structurally atypical base *maŋayaw as containing a shorter base *kayaw and an active verb prefix?...the evidence that CEMP [Central-Eastern Malayo-Polynesian] languages once had an active process of NS appears far stronger than the evidence for a similar fossilization in Formosan languages." (Blust 2004: 75).

The problem with this argument is that Puyuma *maŋayaw* is not completely isolated as a form attesting NS in Formosan languages. For example, we have PAN *naRa vs. *taRa(h), with the former attested in Formosan languages, PAN *dakiS 'climb, ascend' > Saisiyat *manakiʃ*, and we also find cases like PAN *taqu 'know' > Puyuma *manaqu*, and PAN *tukub 'cover' > Kavalan *mnukub*. The same source also provides evidence that the morphological origin of nasal substitution is attested in Formosan languages, but it is incorporated into larger clusters that have merged together into 'portmanteau' formations (Kaufman 2009: 9). We may conclude from this that NS was present in at least some later stage of the development of the Austronesian family, but not necessarily as late as the Malayo-Polynesian branch.

The most significant other process occurring in some Austronesian languages that Blust mentions is what he refers to as 'pseudo-nasal substitution', which is a means of repairing words that violate the 'no dissimilar labials' constraint when the infix *-um- is applied to a word (Blust 2004: 76). For instance, we have words like Thao p<um>atash > *matash* 'to write'. One may worry that the existence of processes like nasal substitution and pseudo-nasal substitution in Austronesian languages would conflict with the possibility of identifying denasalization in early Austronesian. This worry can be dismissed by noting several important differences in how denasalization and the morphological processes just discussed apply. The most important difference is that by hypothesis denasalization occurred before the breakup of Proto-Austronesian, quite early in the history of the Austronesian languages. However, both nasal substitution and pseudo-nasal substitution seem to be later developments, per our earlier discussion. In addition, these processes only apply to initial *voiceless* obstruents in the vast majority of languages, so cases where they apply to other forms are probably the results of later analogy. By contrast, the normal outcome of denasalization is an initial voiced obstruent. Lastly, because nasal substitution and pseudo-nasal substitution are derivational morphological processes that apply only to verbal roots, they always co-occur with a resulting change in meaning. Sound changes do not change the meaning of the words they apply to. These numerous differences between the contexts of denasalization versus nasal substitution and pseudo-nasal substitution, as well as the differences in how these processes occur, mean that we do not have to worry about misidentifying a sound change as a morphological alternation.

A few words are reconstructed with initial voiced stops without any doublets being noted but show isolated reflexes with initial nasals. Since there are a number of processes that may re-introduce initial nasals to words, these isolated examples are not ideal, but they could possibly be the residue of the change we are discussing. For example, PAN *buRay 'flower' has reflexes like Saisiyat *boLay*, Atayal *buay*, but Kavalan *muray*. We also have nasal-initial variants of *balabaw 'rat', like Kavalan *m̥la:báw* and Amis (Fata'an) *malafaw* but the initial syllable of this word varies more significantly than this as well so this may not be a good example. Other cases are even more unclear: the word *babuy 'pig' has some reflexes like Murik *mabi* 'wild pig' (Blust 2002: 91), and Bare'e *mbawu*. Prenasalized stops seem to be marginal in Bare'e for inherited vocabulary, but in this word it may be connected to the Murik initial /m/. However, there is another form in Bare'e *wawu*, and I am not aware of any explanation for this internal variation. The pair of reconstructions *muken 'omen dove' and *buken 'id.' appear only in Malayo-Polynesian and only appear with prefixes, but there also appear to be prenasalized reflexes that we can understand as a midway point on the scale of denasalization, e.g. Karo Batak *limbukan* 'a kind of dove'. A spectrum of reflexes suggests ongoing sound change, despite the small amount of evidence. These words are very few, but the overall trend is attested in a good bit more lexical evidence.

The last issue we will address in our preliminary discussion is 'roots'. The word 'root' has a different meaning in Austronesian studies than it does in the general linguistic context, and refers to monosyllabic units that are not necessarily identifiable as morphemes, but seem to appear in various semantically similar words anyway. A few of these are indeed nasal-initial, but we will not address them in any detail in our analysis

because the question of where they stand in the history of Austronesian is still unclear. In addition, many of the roots have an onomatopoeic or sound symbolic character, like *-tuk 'knock, pound, beat', in which case they would not be good examples for the analysis of historical phonology. However, for the very few nasal-initial roots that are reconstructible for Proto-Austronesian, we see the same variation as is found in properly reconstructible disyllabic morphemes when the nasal appears in word-initial position. For instance, *nemnem 'think' has a doublet *demdem; *n ~ *d is exactly the kind of alternation we would expect, and the nasal seems to be original here since this may have been borrowed from a Tibeto-Burman source; compare Written Tibetan *snjam-pa*, Old Chinese *n[ə]mʔ, both meaning 'think'. We do have a triplet of forms *ñawñaw, *rawraw, and *SawSaw, all meaning 'wash, rinse', and these variants are idiosyncratic to this form. However, we will discuss this form later in some detail. In conclusion, roots will only occasionally be mentioned because of their unclear history, and will only be examined in detail when they show unusual behavior, or other evidence for denasalization at some place of articulation is not readily available (this will be the case for velars). For more information on roots, the standard reference text is Blust 1988.

In any case, I will now proceed to the analysis of the nasal-initial lexicon of PAN as reconstructed in the ACD. This will not include function words/grammatical particles, as they would never be in a prosodically strong context, and so per our earlier discussion would never be subject to denasalization. Although there are a number of different processes at play in the content words that we will examine, a good portion of them provide evidence for denasalization in a word-initial context.

***m-initial words**

*maCa 'eye'

For our hypothesis, this word is probably the most challenging to account for. It is well distributed in all subgroups of Austronesian, being absent only in Kanakanabu, Saaroa, and Atayalic. There is no reason to believe that the absence of this term in two closely related languages and a group that is clearly highly lexically innovative indicates that it is a later lexical development in Austronesian history.

The possibility that this word was originally morphologically derived with the *ma- prefix is also not plausible. Juliette Blevins proposes such an analysis in her argument for a genetic connection between Ongan and Austronesian, claiming that *maCa is derived from *ma-aCa/s; with a secondary segmentation of *aCas 'high, tall' (Blevins 2007: 167). Blust presents several reasons why this analysis is seriously flawed. Namely, there is no good Austronesian-internal evidence that either the presence here of *ma- or the existence of an *-s suffix is present (Blust 2014: 320). We might also note that the semantics here are very strained. The semantic shift proposed here is apparently 'high' > 'high up on the face' > 'eye(s)'. This seems implausible on its face in my opinion, and I have not been able to find any other examples of such a semantic shift in the languages of the world.

Furthermore, similar forms for 'eye' are present in several other language families nearby:

Proto-Tai: *p.ta: A

Proto-Kra: *m-ʈa A

Proto-Austroasiatic *mat

Proto-Tibeto-Burman *s-myak

The evidence all points to an original initial *m-, which for some reason has not been lost. Several possibilities could explain this. First of all, it's possible that the word could have been borrowed into Austronesian, e.g. from a Tibeto-Burman source. Although it is basic vocabulary, Hmong-Mien has apparently borrowed its word for 'eye' from Chinese: compare Proto-Hmong-Mien *muei to Old Chinese *C.m(r)[u]k, Mandarin *mù*. If it was borrowed after the denasalization sound change occurred, then it would still preserve an initial nasal. Perhaps this word is a Kra-Dai borrowing, and the contrastive length present in that family (see the Proto-Tai form above) was interpreted as stress in Proto-Austronesian. In this case, per our discussion of why the *ma- prefix was not denasalized, the *maCa word would not be either.

Nevertheless, it is not preferable to hypothesize borrowing for such a basic vocabulary term, and it is possible to retain the above argument with *maCa being native. Although Blust does not reconstruct stress for Proto-Austronesian, some other scholars do (cf. Wolff 2010, Ross 1992). Some modern languages indicate that an earlier contrastive stress pattern must have been present. For example, Tsou *mcoo* 'eye' < *maCa requires final stress for vowel syncope; compare *walu > *voru*, which preserves the first vowel. In the forms the ACD lists from the Philippines, final stress is also by far the most common for the 'eye' word when stress is indicated. Since Tsou (Formosan) and the languages of the Philippines match up for this lexical item, it is plausible that it was originally stressed on the final syllable in PAN. And indeed Ross reconstructs it as *maCá (Ross 1992, 24). This might lead us to believe there should be even more nasal-initial words preserved, unless there is some unknown correlation between initial stress and the

presence of a nasal, but since most modern Formosan languages do not reflect contrastive stress (Ross 1992: 54), it may just be that contrastive stress and its distinctions were mostly obsolescent by the PAN stage.

As we can see, a number of different hypotheses have been offered to explain why this word is phonologically anomalous among Proto-Austronesian vocabulary items. Those that seem plausible are all consistent with nasal substitution, but Ross' explanation in terms of stress seems most preferable for my hypothesis.

**maliqi* 'pregnant'

Not much needs to be said about this word because the ACD provides a clear explanation for why there is no morpheme boundary indicated for a word that evidently contains the **ma-* prefix: "This form almost certainly contains the stative prefix **ma-*. However, it is unclear whether the PAN base was **aliqi* or **liqi*." Since there is no way to decide from the reflexes where the morpheme boundary is, it was not indicated even though there is plenty of reason to believe it was indeed originally present.

**mamah* 'father's brother'

No reflexes of this word vary in the presence of an initial nasal, however we have good reason to believe that the initial **m* is not original. In several cases, reflexes of **mamah* are used for the vocative of the 'father' word, **ama(x)*. In that case, **mamah* may be derived from the father term via reduplication. Reflexes of **mamah* are also used for the meaning 'father'. If it was a later derivation, then it would not have originally had

an initial nasal and thus would not have been subject to the sound change that we hypothesize.

**manuk* 'chicken'

This word is reconstructed for PAN by Blust on account of the fact that reflexes with the meaning 'bird' exist in Basai and Trobiawan, two different varieties in the Ketagalan dialect cluster. Although these are the only non-MP languages to reflect this form, Blust nevertheless reconstructs it back to PAN, on the grounds that a borrowing relationship does not explain the semantic difference between 'chicken' and 'bird' in the reflexes. The most plausible historical source for a borrowing would be a Philippine language, but in almost all modern Philippine languages the basic reflexes of **manuk* refer specifically to chickens, and not birds in general; for the more general term we see reduplicated reflexes of **manuk* (Blust & Trussel).

There are a few reasons why this argument is not very convincing. Under the supposition that 'chicken' was the original meaning of **manuk*, we have several independent examples of the semantic shift 'chicken' > 'bird' in Austronesian, including several languages of Borneo, Javanese, and in Oceanic. We have no a priori reason to believe that the same semantic shift that has happened many times in Austronesian could not happen in Ketagalan.

In addition, the languages of Taiwan mostly have a different word for 'chicken':

Bunun *tulukuk* 'rooster'

Saaroa *toroko:kə* 'chicken'

Puyuma *turukuk* 'fowl, poultry'

Tsou *troʔua*

Kanakanabu *tarukúuka*

Kavalan *traquq*

Proto-Rukai *tarokoko

[PAN *tuRukuk]

Also see Paiwan *koka*, Siraya *kuka*. Some scholars feel that the *turu-/tara-* forms may be derived from this smaller form via prefixation (Sagart 2005: 3). Regardless, these forms clearly have considerable antiquity in Formosan languages, as they show regular reflexes of sound changes: $k > ʔ$ in Tsou, $k > q$ in Kavalan before back vowels, and in contrast to /r/ everywhere else, we find /l/ in Bunun, where /l/ is the normal reflex of *R. By the 'majority rules' principle of historical linguistics, we should say that this form is the original 'chicken' word in Austronesian, and the *manuk word, which is only present in Ketagalan and MP, was a later innovation.

Considering that *manuk is very well distributed in Continental Southeast Asia, we have a plausible source of this innovation as being a borrowing from some source there during the expansion of Malayo-Polynesian.

PHM *m-nək 'bird'

Proto-Kra *ŋək D 'bird'

Proto-Tai *Ç.nok D 'bird'

Langjia Buyang *maOnuk11* 'bird' (ABVD)

*maLaŋ 'sharp'

Blust & Trussel report reflexes from two languages in Taiwan: Pazeh *ka-malaŋ* 'sharp (blade)', *k<in>a-mala-malaŋ* 'very sharp' Kavalan *manaŋ~maŋan* 'sharp (point, blade)'. Searching the ABVD does not provide any other apparent cognates in Taiwan, and a monosyllabic base *-Laŋ is not observed elsewhere in the family, but considering that this has an adjectival meaning it is probable that the initial *ma- is the stative prefix.

*maqaw 'a plant: *Litsea cubea* (Lour.)'

The species name is actually *Litsea cubeba*. In Li (1994: 260-261), we find this term showing some reflexes that indicate species of alder, one of which is Amis *p-in-aqaw* 'black alder, *Alnus*'. This distribution indicates that the base was probably originally *paqaw, and the modern forms are various nominalizations of this earlier base. Although the alternation between /m/ and /p/ in forms can sometimes be a result of Nasal Alternation, it is more likely that here we have the usage of *-um- followed by the effects of the constraint against non-identical labials. This would provide a good analogy to the morphology of the Amis form, which shows the *-in- infix. It has multiple functions in the modern languages, but its basic usage in Amis is as a nominalizer for the object of the verbal stem (Starosta, Pawley, & Reid 1981: 422). All forms would then be nominalizations, with some languages taking different strategies than others, and this would indicate that the initial nasal was non-original.

*-matek 'jungle leech, *Haemadipsa* spp.' and *-meCaq 'paddy leech'

These forms are considered together because the semantic and formal similarity makes it very probable that these two bases are connected to each other and originally went back to the same form. Most likely, *-meCaq is original because it preserves distinctions that are often lost in Austronesian languages (in particular, the *t vs. *C distinction).

However, when we look at the distribution of forms, both are well distributed, but not so among the Formosan languages. The ACD reports that in Taiwan, *-matek forms are attested in Kanakanabu, Rukai, and Paiwan, but in the Kanakanabu form *ʔanimətək-a* 'creek leech' the vowels seem to match a reconstruction of *metek instead of *matek. Interestingly enough Kanakanabu also has a reflex of the other form: *ni-məcaʔə* 'paddy leech'. The only other Formosan language with a reflex of *-meCaq is Amis, with *la-lintaq* 'mountain leech' (reduplicated). The Amis form shows a trend that is common for reflexes of this form--since it is always prefixed with *qali or a variant, there is typically loss of *e in a medial syllable:

Isneg *alimtá*

Malagasy *dínta*

Malay *(ha)lintah*

In these cases, the /m/ simply is not in an onset, and so it would not be subject to the change. The Amis form shows that this vowel loss occurs in Taiwan as well, so it must have considerable antiquity. Forms for other words show that medial *e varies with null quite often in Formosan languages, for instance compare Pazeh *belebel*, Puyuma *belbel*, both 'banana'. When a variant form with /a/ rather than /e/ arose in the family, it

could easily maintain its /m/ by analogy to the original form. The fact that in languages like Kanakanabu doublets are present only makes it more likely that speakers would be aware that these forms are related.

*-medaw 'dizzy, giddy'

For this form, we can use the same argument as for the 'leech' forms above: the word for 'dizzy' or 'giddy' is also always prefixed in modern languages with some variant of the *qali- prefix. And indeed we see vowel ellipsis in reflexes of this form as well, e.g. Ilokano ali-ndaw. To repeat, since the sound change being proposed is initial denasalization, the correct phonological environment is not present in this lexical item.

*mujiŋ 'face'

This is obviously a prefixed form of the 'nose' word *ijun~ujun, as many reflexes of *mujiŋ reported in the ACD preserve the original 'nose' meaning, but it is unclear what the original function of the prefixed *m- was. There is a prefix *mi- for 'possession of an object' reported in the ACD as well but it has a very poor distribution. In any case, the initial nasal here was apparently not original.

*mula 'to plant'

Within Austronesian, this term usually alternates with *CaNem, but the latter's basic meaning was probably 'bury' in general, rather than referring to seeds in particular. In addition, similar forms are attested in the Kra-Dai family. A form *Cuhra: is reconstructible for Proto-Hlai, and some similar forms appear in Kam-Sui.

Proto-Hlai: *Cuhra: (Norquest 2016: 98)

Mulam: *mya2* (ABVD)

Northern Dong: *mja22* (ABVD)

This makes a reconstruction of *mula fairly solid. However, in the Formosan languages the evidence is somewhat less clear. Saisiyat and Malayo-Polynesian do reflect a reconstruction of *mula regularly, but we have attestations from most other Formosan languages reflecting a metathesized form. The ACD reports Kavalan *m-ruma* 'to plant, to grow, to cultivate', but we also have (from the ABVD):

Basay (Ketagalan) *pa-luma*

Amis *paluma'*

Pazeh *xuruma*

Since strictly speaking, Proto-Malayo-Polynesian is only a single witness for a proto-Austronesian reconstruction, the 'metathesized forms' equal or outnumber the 'non-metathesized forms' from the perspective of Proto-Austronesian. In addition to these forms Seediq has *mhuma*. This is not regular for Seediq, but it is for the adjacent language Bunun (compare *lima 'five' > hima, *luCuŋ 'monkey' > hutuŋ, etc.). Considering all of the available evidence, we have more independent reflexes of a base *luma than a base *mula.

From this distribution it is probably better to take the base *luma as original. Metathesis is generally held to be a sporadic and irregular change in phonology except in very particular cases, and if we held that the original base was *mula, we would have to propose that more reflexes metathesized the proto-form than kept it the same. This is not a likely scenario. Given the 'luma first' account, we would only have to propose that

metathesis happened from there in the individual languages and then the Kra-Dai forms are due to later borrowing.

However, regardless of which position is taken on the shape of the original form, there would be a few examples of metathesis for this word. It is possible that early Austronesians heard the word *mula and interpreted this as a base **la with a *mu-prefix. As we have discussed already, this prefix eventually became an infix in most parts of the Austronesian family, and we would thus expect the analogy to follow this morphological shift, producing forms going back to *luma as we see in several different branches. Since bound morphology was resistant to denasalization, this reanalysis of *mula would result in the preservation of the initial *m.

*mumu 'Formosan blind mole: *Talpa micrura insularis* (Swinhoe)'

This form has only two reflexes, and the correspondence is not regular, as the Kavalan form seems to show reduplication. Thao and Kavalan are not currently adjacent, but it is evident from their history that Kavalan speakers were probably nearby Thao territory at some point; considering the irregularity of the correspondence, this word may have been a relatively recent, localized innovation.

***n-initial words**

*najam 'accustomed to, familiar with; tame'

This has a good number of reflexes but there are some forms that vary in the initial consonant. A stem *dajam with the same meaning also has a good number of reflexes:

Pazeh *daxam*

Favorlang/Babuza *ma-darram*

Siraya *ma-dagam*

There are also several other words whose relationship to the stem **najam* is not obvious on the surface, but historically were likely to be much closer. Tsou *a-hmohmo* and Saaroa *ma-lalamə* are derivable as reduplications of a stem **Lam*, as is Pangasinan *lamlám*, assuming it's a case of the occasional word-initial preservation of **L* in Malayo-Polynesian. This seems pretty far from **najam*, but as later lexical items will show, there is some variation between the sounds **L* and **n* in other PAN words, and they later merged in PMP. As for the monosyllabic structure of the form, **j* is regularly lost in Tsou and Saaroa. **j* > Ø would give us a stem **nam* and then denasalization produces **Lam*. Pangasinan normally preserves **j* as /g/, but if the **Lam* stem became lexically separated, it may have been preserved in this language of the northern Philippines.

Lastly, another variant **laja* is found in Oceanic. Blust reconstructs **lasa*, but this is incompatible with the Polynesian reflexes like Tongan *lata* 'be tame' and Hawai'ian *laka* 'tame'. These point towards a Proto-Polynesian **lata*. The sources of Proto-Polynesian **t* include only **t*, **d*, and **j* (Otsuka 2005: 17), and so this form should be connected with a Proto-Oceanic **laja* rather than **lasa*. Although none of the reflexes give evidence of a final consonant, the correspondence in meaning and form suggests that this is most likely yet another denasalized variant of PAN **najam*. Puyuma has a form *ladam* which descends regularly from a proto-form **lajam*, and this matches the Oceanic forms, although there is considerable distance between Oceanic and Puyuma.

*naLaq 'pus'

This word does not seem to show any denasalized variants, except for isolated forms like Maranao *danaʔ*. The ACD also connects Pazehe *laŋa*, but we would have to make exception for several irregularities to line these forms up. The only discussion in the literature that seems to point to an explanation of this is Laurent Sagart's connection of *naLaq with Tibeto-Burman, where the corresponding lexical item has metathesized onset consonants (Sagart 2002: 5). If the nasal was not originally initial then there is no issue, but without good evidence of internal variation or a more firm set of external representatives of this etymon, it is probably best to just chalk this word up as exceptional with regard to my hypothesis.

*naRa 'to wait'

This lexical item shows a doublet in *taRah according to the ACD, but the evidence is actually contradictory with regard to the presence of the final *h. Only the Bunun reflex of *taRah, *mal-ta-talah* 'to wait', shows a final /h/, which makes its absence in the Amis reflex *tala* 'to wait' problematic; compare Amis *faʔloh* 'new; renew' with the reconstructed *baqeRuh. Since the languages involved reflect the absence of a final consonant rather than a final *h, we can say that the final /h/ in the modern Bunun form may simply be an idiosyncrasy. Based on the data in the ACD some varieties of Bunun do seem to occasionally insert unetymological glottal sounds word finally, so this has some plausibility.

If the *-h is unetymological, then we appear to have a case of nasal substitution. However, Blust in the ACD remains skeptical of this analysis because nasal substitution

is not productive in any modern Formosan language, and even as a fossilized morphological process the evidence for its presence outside of MP is very slim. We have already discussed how doublets like *taRa(h) and *naRa can be compatible with nasal substitution even if it is not productive in any modern Formosan languages, so we will regard this pair of forms as following from that process, *pace* Blust's argument described above. With the earlier discussion in mind, we can conclude that the nasal-initial form is morphologically derived. Further evidence for this is the fact that all representatives of *naRa in the ACD have additional prefixes, which may be the source of the nasal. For example Squliq Atayal has a reflex *mə-naga*, which could be derived from *maŋ-taRa.

*nasuk 'cook by boiling'

This form is connected with a doublet *Lasu, 'to boil.' At first it is difficult to know what to make of the comparison that leads Blust to reconstruct this form for the ACD. The Paiwan and Kelabit forms are regular, but the Gedaged form is not: the latter language never preserves Proto-Oceanic (and thus, Proto-Austronesian) final consonants; see the many reflexes of Proto-Oceanic words given in the section on Ngero-Vitiaz languages in Ross 1988. In the ABVD we find a form *t-um-ahuk 'to cook' for Proto-Atayal, and we can project this back to an original PAN stem **tasuk, from which Paiwan *nasuk can be derived via NS. Thus the appearance of a doublet here is in fact a result of convergent evolution of two semantically similar words. Furthermore, the nasal-initial here appears not to be original.

*nimas 'bailer in a canoe'

This form has a doublet in *limas, and both are well-distributed in Austronesian. It is one of the clearest cases we have of alternation with regard to a nasal vs. a non-nasal-initial.

*nipen 'tooth'

This word has avatars showing a range of initial consonants, but *lipen is a common alternative, showing the expected variation. Another variant form, *ɲipen, will be discussed below, as it is clearly a result of sound symbolic change.

*nunuh 'female breast'

As noted in the ACD, this term is in competition with a much better-supported lexical item *susu with the same meaning. Blust notes that it is most likely a 'nursery word', and this interpretation makes sense when we take note of other forms he notes like Mono-Alu *lulu* and Makatea *mumu*.

***ñ-initial words**

Other than a couple of monosyllabic roots, there is only one good PAN reconstruction with an initial *ñ:

*ñamuR 'dew'

We have a good 'denasalized' variant, but several other variants of this stem are difficult to explain. Tagalog *hamog* and the Oceanic language Pokau's form *lamu* go back to a proto-form *lamuR, which is what our theory predicts. However, there are strange

alternates like *dahemuR > Manobo *dehemug*, Balinese *damuh*, and Ibaloy *amol* seems to show the loss of the initial entirely. If the palatal nasal is original here, the wide variety of reflexes may simply have to do with the fact that it was a rare sound to begin with in Proto-Austronesian and had a very low functional load; some doubt that it was a part of the original phoneme inventory at all. Regardless, this is the only good example we have of a word with an initial *ñ and it seems to show the same results of denasalization as *n does.

***ŋ-initial words**

Words with initial *ŋ, contrary to what we might expect given the general cross-linguistic constraint against initial velar nasals, are the largest class of words here. However, the vast majority of them fall into a particular semantic class, and only two need to be examined on their own.

***ŋajan 'name'**

Other than words with a sound symbolic character, this is the only lexical item reconstructed for PAN that seems to have originally begun with an initial velar nasal. It shows several variants without an initial nasal. One variant is reported in the ACD as *ajan with no initial at all. It is quite common and is well-represented across the different sub-branches of MP. However, the evidence from Oceanic points to an initial *q- (Ross, Pawley, & Osmond 2016, 206). Since *q- is lost in many languages, the reflexes would not usually distinguish between it and a null onset.

Another variant seen in several languages suggests an initial *g: the ACD reports Kenyah *karan*, Simalur *kaxan*, Mentawai *gagan*, Bare'e *gara*, and Kemak *galan*. This is what we would predict as the result of denasalization, but the *q initial variant is clearly more common. The devoicing of *g in the Kenyah and Simalur forms points to a possible explanation of this but the explanation for why *g devoices to *q instead of *k in these forms is unclear. Another example of initial *g ~ *q may be seen in *gemgem 'fist; hold in the fist' vs. *qemqem 'hold in the closed mouth'.

*ŋayaw 'headhunting'

For all languages represented in the ACD with the exception of Puyuma *ŋayaw* 'headhunting', the reflexes are compatible with an initial *k in this form, suggesting that the Puyuma form is a result of nasal substitution. As we have previously discussed, Blust's skepticism in regard to the presence of NS in Formosan languages seems unwarranted, so we can safely conclude that the nasal here is a result of morphological derivation. The Puyuma word may perhaps be a result of backformation, since this language also has a morphologically complex form *ma-ŋayaw* 'to go headhunting' (said of many people), consistent with the morphological context of typical NS. The rarity of NS in Formosan languages may have contributed to this backformation, since Puyuma speakers would eventually not be aware of the morphological process that derived the form *ma-ŋayaw* from an earlier *kayaw, if this process was not ultimately productive in their language.

*ɲipen 'tooth'

This form, yet another variant of *nipen, is clearly a result of the sound symbolic association between velar nasals and the mouth/nose area common to Austronesian languages (cf. Blust 2003). In fact, except for the two discussed above, all other words reconstructible for proto-Austronesian with an initial velar nasal have meanings within this semantic domain. This explains their phonological stability in contrast to other initial nasals in Austronesian, because onomatopoeia, ideophones, and similar examples of iconicity in language often are exceptional with regard to phonological rules generally. Many examples exist in English, but the frequent pronunciation of a syllabic /z/ in the onomatopoeia 'buzz' for flying insects is apt enough. Due to these issues it is not surprising that we find many words with initial velar nasals when initial nasals in general have been shown to be so rare.

Interestingly, we also find a variant *ipen listed in the ACD, but Blust points out that Blaan has *kifen*, suggesting earlier *qipen; he also says a *q-initial form could be a "potential source for many forms in [the vowel initial] comparison." If the correct reconstruction is *qipen, then this form would suggest the same development of *ɲ > *q as we find in *ɲajan ~ *qajan 'name'. Per our earlier discussion, this is probably a secondary development from *ɲ > *g, as we find the same variation in the 'name' word and in *gemgem ~ *qemqem 'hold in the fist; hold in the mouth'. Considering this matrix of variants, the word for 'tooth' is unique in that it has two different nasal-initial variants, *both of which* have denasalized variants.

CHAPTER 4

External Evidence for Initial Nasals

We have seen so far that the words reconstructed with initial nasals in Proto-Austronesian almost all show a high degree of variability in their initial consonants with regard to the presence of a nasal. The most likely explanation for this is that the words with initial nasals we see that cannot be chalked up to later morphological innovations are the residue of an ongoing change that simply was not complete at the time of the split-up of Proto-Austronesian. However, in order for this explanation to be completely plausible we would want a number of examples of lexical items that used to have initial nasals, but do not show initial nasals in any Austronesian languages. The comparative method would not allow us to identify any such words, for obvious reasons. However, Southeast Asia has a long history of interaction between different cultures and different linguistic groups, and so there is a large amount of shared vocabulary between languages as well, especially with the languages in the Kra-Dai family. We would predict that if denasalization occurred, then a good proportion of this shared vocabulary should show nasals in other languages where in Proto-Austronesian they show initial voiced stops, or in the case of coronals, they may correspond to initial PAN *l or *L. As we will see, this does prove to be the case in many words.

It would be thinkable that all examples of initial voiced stops in Proto-Austronesian were originally initial nasals. This does not seem to be true, as we have some words with an initial voiced stop in PAN that show initial obstruents in other nearby

languages. For example, PAN *biRaq 'wild taro' corresponds well to terms for 'taro' in the Kra-Dai family:

Proto-Tai *pruəkD

Proto-Hlai *hra:k (Norquest 2016: 251)

Proto-Kra *p-ɣakD

These show an initial obstruent, so there is no evidence of an initial nasal here. However, a number of terms with initial voiced obstruents in PAN do show nasals elsewhere in Southeast Asia. Many of these are even in basic vocabulary. For example, the word for 'woman' in PAN is reconstructed in the ACD as *bahi. This term has been connected to m-initial nursery words for 'mother' in the literature (Krupa 1994: 233), and when we look at words for 'mother' or 'woman' in nearby languages, the similarity is uncanny.

Proto-Tai *me:B 'mother; woman'

Proto-Hlai *hmi:ʔ (Norquest 2016: 342)

Jiamao *ma:j*5 (ABVD)

Proto-Kra *maiC

The Proto-Tai form is especially significant because it shows that the semantic shift from 'mother' to 'woman' is attested. Although this seems intuitively reasonable, an attested example of this kind of polysemy strengthens the argument for such a semantic shift in Proto-Austronesian. In other cases however, finding evidence of an original initial nasal requires closer analysis.

PAN *beRas 'unhusked rice'

Representatives of this word in Kra-Dai are lacking, except for Proto-Kra *mla(u)C 'rice (cooked)', but this has a semantic difference from the PAN form, and it lacks a final consonant. More phonologically and semantically similar forms can be found in Tibeto-Burman, where we find Written Tibetan *hbras* 'rice' (prenasalization in transcriptions of Written Tibetan is indicated by 'h') and Old Chinese *[r]ʰat-s 'dehusked but not polished grain'. The unspecified initial in the Old Chinese form is compatible with something like **mr- (see Sagart 2011: 126-127), and the prenasalization of the Written Tibetan form provides even more evidence of this. The evidence thus points to the Proto-Kra form and the PAN form both being taken from some Tibeto-Burman source with an initial nasal.

PAN *bulaL 'moon'

Apparently connected forms appear throughout Kra-Dai, but they have long been noted as having unusual correspondences (Thurgood 1994: 350). However, when we consider the possibility of an initial nasal in this word, the set of reflexes makes much more sense. Proto-Tai *ɓluənA 'moon, month' attests an initial voiced obstruent, but Proto-Kra *m-djanA, Proto-Hlai *C-ɲa:n 'moon/month', and forms like Mulam *mya:n2* (ABVD) point to an initial nasal.

PAN *qabaRa 'shoulder'

This word does not appear at first glance to have an initial voiced obstruent, but *qa- may be a fossilized prefix, although its function is unclear. We also find a form in

PMP *baRa 'hand, arm', *qapejux 'gall, gall bladder, bile', and *qalima vs. *lima, both meaning 'hand'. In the ACD Blust considers that *qaqay 'foot' may derive from **qaqai. Similar forms for the meaning 'shoulder' in Kra-Dai mostly show initial obstruents, but the reconstruction Proto-Kra *m-ba shows nasalization. The difference in the distribution of nasalization here may be a result of the word being a later introduction to Kra-Dai when the process of denasalization was mostly complete.

PAN *pudu, *burun, *buyuq, Proto-Oceanic *pu-pulu; 'betel pepper'

These forms are highly variable, but they all show a family resemblance to each other and to the Mon-Khmer reconstruction *mləw 'betel pepper', which has an initial nasal. Since betel chewing is a significant part of the culture of all of continental and insular Southeast Asia, this irregularity may be the result of a number of separate borrowings into Austronesian, although this faces archaeological difficulties (Zumbroich 2007: 108). If these terms are ultimately related then this is another good case of denasalization.

PAN *beli 'buy, brideprice(?)'

This could possibly be connected to the Old Chinese word *m^hraj?, but the final vowel is somewhat different and there is a semantic difference that may be problematic for the connection of these forms. The ACD notes that there is no evidence of money in early Austronesian economy, and that the specific meaning of brideprice was likely the original designation of this term. One could alternatively argue that the lack of a monetary system in early Austronesian economy would motivate the borrowing of such a

term, pointing more towards a valid connection between Tibeto-Burman terms and the PAN reconstruction, and validating this as an example of denasalization.

As one can see there are a good number of words where external evidence points to an original nasal-initial for the labial place of articulation; however, identifying similar evidence for initial nasals at other places of articulation is difficult for several reasons. One of which is that for the alveolar position there is variation in what the outcome of denasalization is. Another is that it is probable that the process of denasalization happened earlier and more quickly at non-labial places of articulation (see above). In this scenario, if a form was borrowed from early Austronesian into another language family, the original nasal may have already been lost, and so could not be preserved in external comparanda that did not undergo denasalization.

PAN *daki 'dirt on skin; dandruff'

This reconstruction can be compared with *Caki / *Caki 'feces, excrement'; the phonological forms are very similar and the semantic connection between 'dirt' and 'feces' is attested in other languages such as English. In the literature, phonologically and semantically similar forms in other languages of Southeast Asia have been connected to the *C-initial terms *Caki/Caki (see for example Ostapirat 2005: 122), but this does not match up with the reconstructed forms.

Proto-Hmong-Mien *N-KəjX

Proto-Tai *C.qɯj

The Proto-Hmong-Mien form directly points to an earlier nasal, as does the Proto-Tai form, which represents an unspecified initial voiced consonant. Because of these

considerations, it is most reasonable on phonological grounds to pair these words up with PAN *daki rather than PAN *Caki or *Caki, and the Hmong-Mien reconstruction points to an original initial nasal in this word.

I have not been able to identify any full words with a velar onset showing denasalization when external evidence is compared; however, one case of a 'root' does. From the word *gemgem 'hold in the fist' we can extract a monosyllabic root *gem and compare this with PTB *ɲam 'chew, hold in the mouth' (for the semantics see above), which is noted as possibly connected with another form PTB *gam 'put into mouth, seize with mouth'. For the latter, the Written Tibetan reflex is *hgam-pa*, showing prenasalization. As with the rice word, this allows us to "bridge" between the nasal-initial reflexes and the obstruent-initial one in Proto-Austronesian.

These facts make it clear that there is only external evidence for denasalization at the labial place of articulation, but at other places of articulation there is a mixture of external and internal evidence. This asymmetry seems to suggest that denasalization occurred earlier at the labial place of articulation than elsewhere. This seems to differ from the general trend with regard to denasalization as a sound change. Cantonese clearly shows the opposite order, and 'labial-last' in denasalization is the expected order (Yoshida 2008: 10). However, because the set of words identified as demonstrating denasalization in (proto-)Austronesian is small in number, consisting of no more than about 15-20 lexical items, these may not actually be indicative of any real difference relating to place of articulation, and would not be likely to yield strong statistical results. Therefore, based on the current state of our knowledge, we can say that this stratification in the place of articulation where denasalization appears to occur is at best a tendency. More research

may find additional words showing denasalization, which would clarify whether the chronology of denasalization in Proto-Austronesian differs from that seen in other languages. Nevertheless, in the current state of our knowledge we should not overhastily conclude that denasalization has occurred more quickly at the labial place of articulation.

CHAPTER 5

Discussion

In the previous section, I provided many words which vary between nasal-initial forms, and forms with other initial consonants. Additionally, there were forms which showed initial nasals in external representatives, and showed other, non-nasal-initial consonants within Proto-Austronesian. These both provide evidence of denasalization, but in different ways. However, a number of other processes have produced some noise in the eventual outcomes of denasalization. Nevertheless, this noise can mostly be accounted for by noting some other processes that are common, within Austronesian or generally. Taking inventory of the forms we have seen so far, *m varies only with *b, so there are no issues to address. For the velar point of articulation, *ŋ varies with *g and *q, the uvular stop. This is not a serious issue, as loss of voicing is common at the velar place of articulation in Austronesian languages (Blust 1996: 145). Although in the source just cited the variation is between *g and *k, variation between *g and *q, and between *ŋ and *q seems represented strongly enough to say that this is a process related to what Blust describes.

For the alveolar and palatal positions, the reflexes are somewhat more chaotic. Although we have 4 words that represent variability between *n or *ñ and *d, there are many words which vary between coronal nasals and *l or *L. Some of these are words where we find *d forms as well, but two others show only *l: *n/limas 'bailer', and *n/lipen 'tooth'. For both of these words, there are good explanations for this unexpected

distribution. For *limas, contamination with *lima 'hand' has likely occurred. The formal similarity here is clear. In addition, one can easily see the similarity in appearance between a scoop used for removing water from a canoe and a scooped hand. For the 'tooth' word, the lack of representatives for initial *d could perhaps be a result of the fact that it is basic vocabulary. In an understanding of sound change as spreading throughout the lexicon over time, the words that are likely to undergo a sound change more quickly are those that are more common. Therefore, the secondary change of *d > *l may have idiosyncratically occurred more quickly in this word simply because, being a term for a body part, it is a basic vocabulary item. A merger of *d and *l in Austronesian is also common at least conditionally, which adds plausibility to this account. Since the reflexes other than voiced obstruents can be explained as secondary variation, and this variation is common in Austronesian outside of the loss of the nasal distinction, the default reflexes appear to be voiced oral stops, and so the basic form of the sound change we have identified over these pages is $m\ n\ \eta > b\ d\ g/\#_$, or if a feature-based approach is preferred, we could write $[+nasal, +voice] > [-nasal]/\#_$.

One may notice that the palatal nasal was not written in the segment-based change. The reason for this is that there are some problems with accounting for the outcome of denasalization for the palatal nasal, but when we sort through these issues, it ends up behaving exactly the same as the alveolar nasal. For instance, *ñamuR has the same distribution of reflexes that we find for words with an initial *n. In contrast, the reduplicated root *ñawñaw shows a distribution of variants that is not found in any other reconstruction, being in a triplet with *rawraw and *SawSaw. Several issues need to be addressed here. First of all, not all scholars accept an opposition between *r and other

phonemes at every location within a reconstructed word. Ross claims that *r is simply a variant of what he writes as *d1 when it appears in non-initial position (Ross 1992: 43). Furthermore, Ross's *d1 is equivalent to Blust's *d in the ACD for at least some lexical items, e.g. *d1áRaq can be compared to the ACD's *daRaq. This points to a possible earlier root **daw preceding the observed *raw, but this is inconsistent with the evidence provided in the ACD that roots with initial *d and initial *r have distinct reflexes. Nevertheless, Blust at least seems to affirm that *r is somewhat marginal word-initially, because he only reconstructs a very small number of words with initial *r, two of which are reduplicated roots.

Probably the reason why *ñawñaw shows idiosyncratic variation, including alternation with a proto-phoneme that is marginal at best, is because the distribution of similar forms in nearby languages suggests that it is a loanword into PAN. Upon a search for external comparanda the following forms were identified:

Proto Tai *za:wA 'to wash'

Proto-Hmong-Mien *ntsæwX 'to wash (hands)' cf. OC *mə.[ts]ʰaw? 'wash (hands)'

A prenasalized affricate *nts (< *məts-) could easily be the source of all of the Austronesian forms, depending on the sound change that occurred; *ñaw and *Saw can be interpreted as cluster reduction (with assimilation of the nasal to the following affricate in the first case), and if it went through a stage *zaw like the Proto-Tai form, then *z > *r would simply be rhotacism. Varying strategies of phonological adjustment are common and expected in a borrowing scenario, so this would explain the unusual set of forms we see for this reduplicated root.

The last point of discussion with regard to the palatal nasal is that scholars disagree on whether it can be reconstructed to Proto-Austronesian. The number of lexemes in PAN that represent it in the ACD is vanishingly small, and in Formosan languages it seems to show the same reflexes as *L when a distinction can be made. Malcolm Ross also argues from distributional facts that supposed palatals in PAN are actually derived from coronals (Ross 1992: 37). If Ross and others are right that "Proto-Austronesian" palatals were actually a development within Proto-Malayo-Polynesian instead, then the fact that *n and *ɲ have the same outcome from denasalization is no mystery because they were indeed the same thing when the sound change occurred. On the other hand, if they were different, then the low functional load of the latter would place little importance on the distinction to begin with, and thus little importance on preserving the distinction after denasalization occurred.

Now that we have tied together the loose ends on what the sound change specifically was, an additional question to be asked is when it occurred. We can make only a rough guess at this preliminary stage, but the facts seem to point to the loss of the nasal distinction initially occurring only very shortly before the breakup of Austronesian. In accordance with our earlier discussion of the causal factors involved in denasalization of the type that Proto-Austronesian had undergone, these factors would have been quite strong in the PAN context and thus the change probably would have occurred fairly quickly. On the other hand, the variation internal to Austronesian with regard to the presence of initial nasals implies that the sound change was not complete when breakup occurred. At the same time and most likely afterwards as well, progress in the grammaticalization of a number of nasal-initial prefixes in different branches of the

family would have obscured the relationship between word-initial position and any ongoing sound change. The result of this is that we would have a good amount of 'residue' left over from the change where nasal-initial and non-nasal-initial forms coexisted in the speech community.

CHAPTER 6

Conclusion

The claim I have aimed to demonstrate in this thesis is that a short time before Proto-Austronesian broke up into several branches, a sound change occurred in which all word-initial nasals merged with voiced stops at the same place of articulation. Typological evidence was presented for why this is plausible. After this, I showed that variation in some words internal to the Austronesian family, as well as the presence of nasal-initial external comparanda, gives good direct evidence that this sound change did indeed happen, although to some degree it has been obscured by morphology that subsequently or contemporaneously developed along with the sound change. Nothing about this changes the precise phonotactics or phonological inventory of Proto-Austronesian as reconstructed, but it does have wider implications for comparative and historical Austronesian studies. As we saw earlier, some other languages in East Asia have demonstrably undergone the same change, and although Korean is losing initial nasals now and not thousands of years ago, there could be some kind of areal factor involved. In addition, there are very serious implications for work attempting to identify deep genetic relationships between the Austronesian family and other language families. If many initial voiced stops in Proto-Austronesian indeed used to be nasals at some point, which is probable, given that we have examined several such cases in the previous pages, then many cases of lexical items that appear to be connected between Proto-Austronesian and some other language family would have to be later borrowings if they unambiguously

represent an initial voiced stop, and we know that they formerly had an initial nasal. Another issue is that if we know a merger occurred, then we need at least a few more comparisons to be sure that an apparent relationship is real. Since an initial voiced stop in Proto-Austronesian could be connected to at least two different sounds in any other language family, this increases the probability of similarities that occur simply due to randomness as compared to a simple one-to-one correspondence where a merger has not occurred.

Lastly, this thesis contributes to the ongoing investigation in Austronesian studies as to whether various curiosities in the phonology or lexicon point to some earlier stage deep in the history of the language family. For example, as we have seen, Malcolm Ross does a similar sort of 'internal reconstruction' in phonology to propose that at an earlier stage of Proto-Austronesian, the pairs *g and *j, and *d and *r were not distinct (Ross 1992: 32, 43). In a more recent article, Robert Blust (2011) also discusses some aspects of the Austronesian lexicon that are quite chaotic from a comparative phonological standpoint. In a broad scope perspective, what I have shown in this thesis is that at least some of these apparent irregularities can eventually be demonstrated to go back to earlier consistencies that are only synchronically irregular due to later interference.

REFERENCES

- ABVD: See Greenhill et al., 2008.
- ACD: See Blust & Trussell, ongoing.
- Blevins, Juliette. 2007. A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands. *Oceanic Linguistics* 46(1). 154-98.
- Blust, Robert & Stephen Trussell. ongoing. The Austronesian Comparative Dictionary. <http://www.trussell2.com/acd/> .
- Blust, Robert. 1988. *Austronesian Root Theory: an essay on the limits of morphology*. Amsterdam: Benjamins.
- Blust, Robert. 1996. *The Neogrammarian hypothesis and Pandemic irregularity*. In Malcolm Ross & Mark Durie (eds.), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York: Oxford University Press.
- Blust, Robert. 2002. The history of faunal terms in Austronesian languages. *Oceanic Linguistics* 41(1). 89-139.
- Blust, Robert. 2003. The phonestheme ŋ- in Austronesian languages. *Oceanic Linguistics* 42(1). 187-212.
- Blust, Robert. 2004. Austronesian nasal substitution: a survey. *Oceanic Linguistics* 43(1). 73-148.
- Blust, Robert. 2011. The problem of doubleting in Austronesian languages. *Oceanic Linguistics* 50(2). 399-457.

- Blust, Robert. 2014. Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics* 53(2). 300-91.
- Chrétien, C. Douglas. 1965. The statistical structure of the Proto-Austronesian morph. *Lingua* 14. 243-70.
- Greenhill, S.J., R. Blust, & R.D. Gray. 2008. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evolutionary Bioinformatics* 4. 271-83.
- Hock, Hans Henrich & Brian D. Joseph. 2009. *Language history, language change, and language relationship: an introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- Kaufman, Daniel. 2009. On *pa-, *pa<R>-, and *pa<ŋ>-. Paper presented at ICAL XI, Aussois, France.
- Kinkade, M. Dale. 1985. More on nasal loss on the Northwest Coast. *International Journal of American Linguistics* 51(4). 478-80.
- Krupa, Viktor. 1993. Aspects of Austronesian kinship semantics. In A.K. Pawley & M.D. Ross (eds.), *Austronesian terminologies: continuity and change*. Canberra: Australian National University. 231-40.
- Lubotsky, Alexander. 2018. Proto-Indo-Iranian phonology. In Jared Klein, Brian Joseph, Matthias Fritz, and Mark Wenthe (eds.), *Handbook of Comparative and Historical Indo-European Linguistics* Vol. 3. Berlin: Mouton de Gruyter. 1875-88.
- Maddieson, Ian. 2013. Absence of Common Consonants. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max

Planck Institute for Evolutionary Anthropology. (<http://wals.info/chapter/18>, Accessed on 2019-03-20.)

Matisoff, James A. ongoing. The Sino-Tibetan Etymological Dictionary and Thesaurus. <http://stedt.berkeley.edu/>

Norman, Jerry. 1991. The Min dialects in historical perspective. *Journal of Chinese Linguistics Monograph Series* 3. 323-58.

Norquest, Peter. 2016. *A phonological reconstruction of Proto-Hlai*. Leiden: Brill. eBook.

Ostapirat, Weera. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23(1). 1-251.

Pittayaporn, Pittayawat. 2009. *The phonology of Proto-Tai*. Ithaca: Cornell University Press.

Robinson, Stuart. 2006. The phoneme inventory of the Aita dialect of Rotokas. *Oceanic Linguistics* 45(1). 206-9.

Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Australian National University

Ross, Malcolm D. 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31(1). 23-64.

Ross, Malcolm D, Andrew Pawley, & Meredith Osmond. 2016. *The Lexicon of Proto-Oceanic* Vol. 5. Canberra: Australian National University.

Sagart, Laurent. 2005. Sino-Tibetan-Austronesian: an updated and improved argument. In Alicia Sanchez-Mazas, Roger Blench, & Laurent Sagart (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge. 161-76.

- Sagart, Laurent. 2011. How many independent rice vocabularies in Asia? *Rice* 4. 121-33.
- Shorto, Harry S. 2006. *A Mon-Khmer comparative dictionary*. Canberra: Australian National University.
- Starosta, Stanley, Andrew Pawley, & Lawrence A. Reid. 2009. The evolution of focus in Austronesian. In Elizabeth Zeitoun (ed.), *Formosan Linguistics: Stanley Starosta's Contributions*. Taipei: Institute of Linguistics, 329-481.
- STEDT: see Matisoff, ongoing.
- Thurgood, Graham. 1994. Tai-Kadai and Austronesian: the nature of the historical relationship. *Oceanic Linguistics* 33(2). 345-68.
- Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. Ithaca, NY: Cornell Southeast Asia Program Publications.
- Yoshida, Kenji. 2008. Phonetic implementation of Korean "denasalization" and its variation related to prosody. *IULC Working Papers* 8(1). Bloomington: Indiana University Press. 1-23.
- Zee, Eric. 1999. Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics* 27(1). 120-67.
- Zumbroich, Thomas J. 2007. The origin and diffusion of betel chewing: a synthesis of evidence from South Asia, Southeast Asia, and beyond. *Journal of Indian Medicine* 1. 87-140.

APPENDIX

Proto-Austronesian Nasal-initial Stems in the ACD

Reduplicated roots are not included here, due to their problematic character (see above). However, anything else that is listed as morphologically simple in the index of Proto-Austronesian reconstructions in the ACD is.

*maCa 'eye, focal point, center or most prominent part'

*maliqi 'pregnant'

*mamah 'father's brother'

*manuk 'chicken'

*maNaŋ 'sharp'

*maqaw 'a plant: *Litsea cubea* (Lour.)'

*-matek 'jungle leech, *Haemadipsa* spp.'

*-meCaq 'paddy leech'

*-medaw 'dizzy, giddy'

*mujiŋ 'face (human)'

*mula 'to plant'

*mumu 'Formosan blind mole: *Talpa micrura insularis* (Swinhoe)'

*najam 'accustomed to, familiar with; tame'

*naNaq 'pus'

*naRa 'to wait'

*nasuk 'cook by boiling'

*nimas 'bailer in a canoe'

*nipen 'tooth'

*nunuh 'female breast'

*ñamuR 'dew'

*ŋajan 'name'

*ŋajay 'saliva, drivel'

*ŋayaw 'headhunting'

*ŋidaw 'toothless'

*ŋilu 'painful sensation in teeth, as from eating something sour'

*ŋipen 'tooth'

*ŋusuq 'nasal area; snout'

*ŋuSuN 'nosebleed'

*ŋuSuR 'nasal mucus'