HELSINGIN YLIOPISTO

The oldest layer of Amuric-Tungusic lexical contacts

Martijn G. T. M. Knapen
Master's thesis
Linguistic Diversity in the Digital Age
University of Helsinki
November 2021

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Abstract

Faculty: Faculty of Arts

Degree programme: Linguistic Diversity in the Digital Age

Study track: Diversity Linguistics

Author: Martijn Gerardus Theodorus Maria Knapen

Title: The oldest layer of Amuric-Tungusic lexical contacts

Level: Master's thesis

Month and year: November 2021

Number of pages: 73 + 43

Keywords: Amuric languages, Tungusic languages, Nivkh, language contact, loanwords

Supervisor or supervisors: Ekaterina Gruzdeva

Where deposited: HELDA digital repository – Helsinki University Library

Additional information:

Abstract: Research on the interaction of the Amuric languages (referred to as "Nivkh" or "Ghilyak" when regarded as a single language) with the Tungusic languages was initiated by Grube (1892). His focus on loanwords has been the object of study until the present day. Recently, Janhunen (2010: 292, 296; 2016: 23) has suggested that contact between the two families already started between their ultimate ancestors: Pre-Proto-Amuric and Proto-Tungusic. This thesis investigates whether some of the lexical parallels proposed by earlier research belong to this period.

As the thesis is written from the perspective of language contact, the parallels are regarded as the result of borrowing instead of inheritance. The distinction between these two modes of transmission formed the theoretical basis for the methodology that was employed. To prove ancient contact, it had to be shown that the Amuric and Tungusic languages inherited their shared lexemes from their respective ancestors and that these ancestors may have borrowed from each other. As the methodology relied on the literature on Amuric and Tungusic historical phonology, an overview of this topic is also included.

First, fifty parallels were drawn from those listed in previous research. These could be reconstructed to Proto-Amuric and Proto-Tungusic using the Comparative Method and thus could have been inherited from them or an earlier ancestor in the case of Proto-Amuric. Additionally, they exhibited phonological similarities that could reasonably be expected from borrowing between Pre-Proto-Amuric and Proto-Tungusic. Afterwards, a direction of borrowing had to be established, the principal evidence of borrowing. For that purpose, nine criteria were developed. These criteria considered morphology, diachronic and synchronic phonology, extent of attestation, semantics and extra-linguistic factors. Finally, the data was separated into older and younger strata, since in the selection phase only the Proto-Amuric stage was considered, while the target was Pre-Proto-Amuric. These layers were classified on the basis of phonological developments.

For most of the fifty parallels the direction of borrowing could be determined. In this stage of analysis, fifteen of them were ultimately dismissed as recent or doubtful. The remaining thirty-five were examined for properties that could have resulted from the sound changes that followed Pre-Proto-Amuric that were proposed in earlier research. Ultimately, it could only be proven that the absence of vowels in non-initial syllables was a property characteristic of ancient lexemes in the Amuric lineage. Consequently, although a substantially old stratum of Amuric-Tungusic parallels was found, further research is needed to show that any of them date to Pre-Proto-Amuric and Proto-Tungusic times.

Table of contents

List of figures

Abbreviations

Languages

Am.	Amuric
AN	Amur Nivkh
Ch.	Chinese
EEwk.	Eastern Ewenki
ESN	East Sakhalin Nighvng
Ew.	Ewen
Ewk.	Ewenki
Ma.	Manchu
Mo.	Mongolic
Na.	Nanai
NgN	Nogliki Nighvng
NSN	North Sakhalin Nivkh
OAm.	Old Amuric
Or.	Oroch
PAm.	Proto-Amuric
PMo.	Proto-Mongolic
PPam.	Pre-Proto-Amuric
PTg.	Proto-Tungusic
Sol.	Solon
Tg.	Tungusic
SSN	South Sakhalin Nighvng
Ulc.	Ulcha
Uil.	Uilta
WSN	Western Sakhalin Nivkh

Sources

CND	Fortescue (2016)
NRS	Savel’jeva & Taksami (1970)
NRS-S	Sangi & Gašilova (2016)
RNS	Savel’jeva & Taksami (1965)
SSTM1	Cincius (1975)

SSTM2	Cincius (1977)

Terminology

CA	consonant alternation
PA	phonological adaptation
PH	palatal harmony
RTR	retracted tongue root (feature)
RTRH	retracted tongue root harmony
VH	vowel harmony

1 Introduction

Already in Grube's *Giljakisches Wörterverzeichnis* from 1892 it was recognised that Amuric and Tungusic had been in contact. His work constitutes one of the first dictionaries on an Amuric language (or languages) and contains perhaps the earliest mention of Tungusic loanwords in a language of that family. The bulk of subsequent research has focussed on compiling and expanding lists of such borrowings and over the years the accompanying analyses have gradually increased in sophistication. Most recently, Janhunen (2016) has employed them to elucidate internal developments in the Amuric language family and to aid in reconstructing Pre-Proto-Amuric, a stage that preceded Proto-Amuric. In the past few years, morphological borrowings have also started receiving attention (cf. Gusev 2015; Pevnov 2016; Alonso de la Fuente 2017).

Janhunen's publications are especially notable in the regard that he believes Pre-Proto-Amuric and Proto-Tungusic to be contemporaneous (cf. Janhunen 2010: 292, 296; 2016: 23). In other words, Pre-Proto-Amuric and Proto-Tungusic existed as distinct linguistic units at roughly the same point in time. These two languages supposedly interacted with each other. If that is indeed the case, then there should be evidence of these early contacts within the available linguistic data. Searching for such evidence in the lexicons of these two languages represents the primary task of this thesis, which is formulated as the following main research question:

- What lexical items shared between the Amuric and Tungusic languages constitute a layer of borrowing that dates to the Pre-Proto-Amuric and Proto-Tungusic stages of their respective families?

The "oldest layer" of lexical contact alluded to in the title of this thesis is, in this research question, conceived of in a narrow sense. The labels "Pre-Proto-Amuric" and "Proto-Tungusic" are only applied to the oldest reconstructable levels of their respective language families, which means that there is no immediate evidence of a language beyond them that was identifiably of the Amuric or Tungusic lineage.

Therefore, the oldest stage of borrowing that can be found necessarily involves these most ancient forms of Amuric and Tungusic that can be reconstructed.

In the linguistic study of loanwords, there are generally a number of questions that need to be answered in order to classify a lexeme as a loanword. For that reason, my main research question is accompanied by the following minor research questions:

- What lexical items found in both the Amuric and Tungusic languages can reliably be considered to have a common origin?
- What is the direction of borrowing (from Amuric to Tungusic or Tungusic to Amuric)?

The final objective of my thesis is to find the oldest stratum of interaction between these two families, which leaves identifying what strata there are. Thus, I have formulated another minor research question.

- What layers can these loanwords be grouped into?

I begin my thesis by defining the Amuric and Tungusic language families, their extant members and where they are located in the world (§2). Thereafter, the ultimate homelands of Pre-Proto-Amuric and Proto-Tungusic are localised, i.e. the areas where contact between the two protolanguages supposedly took place (§3). This is followed by a discussion of previous research on lexical contact between the two families (§4).

In section 5, I introduce the concept of borrowing and compare it to and contrast it with the concept inheritance. The phonological makeup of lexemes is affected by these two modes of transmission, which impacts my overall analysis. Therefore, I also discuss the phonologies of (Pre-)Proto-Amuric and Proto-Tungusic (mainly from a diachronic perspective) (§6). I first compare their vowel inventories (§6.1), which is followed by a discussion of their consonant inventories (§6.2). After concluding this treatment, I provide a short overview of previous research on phonological adaptation in Amuric, where phonological adaptation presents the phonological dimension of borrowing (§7).

The methodology that I have employed in answering my research questions is presented in Section 8. I give the sources on my data and the manner in which they are presented in this thesis (§8.1). Because I have selected from this dataset only those items that were likely to be of common origin, I also address the principles on which my selection is based (§8.2). In Section 8.3, I describe the criteria I have used to determine the direction of borrowing for each item in my dataset. The underlying principles of classifying loanwords into layers are the topic of Section 8.4.

I apply my methodology in Section 9. First, I give an overview of all the items together with the direction of borrowing that I have established (§9.1). Those words that were most likely to be ancient I have further attempted to assign to different strata (§9.2). After summarising my findings (§9.3), I reflect on my research and methodology in Section 10 and provide suggestions for further research.

2 The Amuric and Tungusic language families

The main sources of information on a language family's common ancestor are its descendants. Therefore, I start with the extant members of the Amuric and Tungusic language families, the descendants of Proto-Amuric and Proto-Tungusic.[1] This section presents them briefly as well as the labels applied to them, their current geographical location and matters of classification, starting with the Amuric language family.

In this thesis I use the label "Amuric" to refer to a language family that consists of two branches: Nivkh and Nighvng. This corresponds to the way in which Gruzdeva and Janhunen (2020) use the term. The immediate ancestor of these languages is Proto-Amuric, while the ultimate ancestor of this family may be termed "Pre-Proto-Amuric" (previously termed "Pre-Proto-Ghilyak"). This is the language that may have existed concurrently with Proto-Tungusic, as per Janhunen (2010: 292, 296; 2016: 23).

[1] Note that I use "Amuric" and "Tungusic" as a shorthand for "the Amuric languages" and "the Tungusic languages" respectively. A common convention within historical comparative linguistics.

The Nivkh and Nighvng branches are often regarded as varieties of a single language instead of branches of a language family. When regarded as a single language, it is generally called “Nivkh” or “Ghilyak”. Since I regard it as a language family, “Amuric” is the most appropriate label. Its division in two branches is adapted from Gruzdeva and Janhunen’s (2020: 3, 8-9) classification, which is shown in Figure 1.

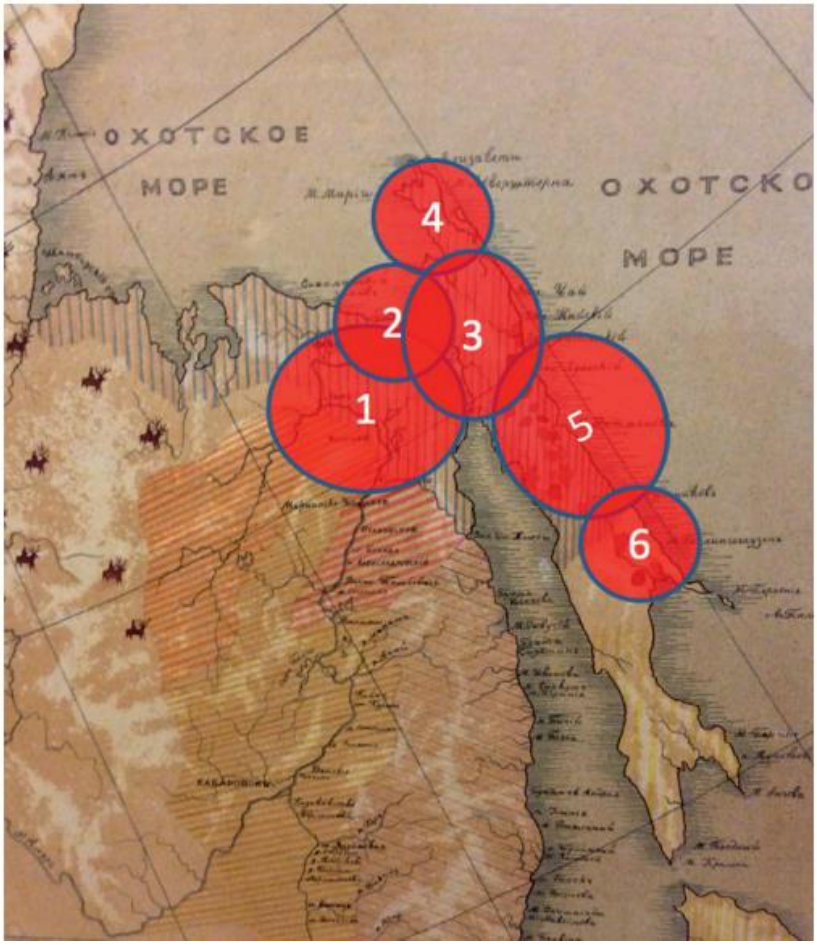

Figure 1: Geographical distributions of Nivkh and Nighvng varieties (Gruzdeva & Janhunen 2020: 3)

In this classification, the main split is between Nivkh and Nighvng. Nivkh (or Amur group) consists of the Lower Amur (1), Amur Liman (2), West Sakhalin (3) and North Sakhalin (4) varieties, while Nighvng (or Sakhalin group) consists of the East Sakhalin (5) and South Sakhalin (6) varieties. The names of these varieties correspond to their geographical locations, while the names of the groups they belong to refer to the word for ‘human’ found within them.

In addition to Gruzdeva and Janhunen (2020: 3, 8-9), there are several other classifications. These differ on how the North and South Sakhalin varieties relate to the Amur group and East Sakhalin variety respectively. The North Sakhalin variety is intermediate between both the Amur group and the East Sakhalin variety, while the South Sakhalin in many ways differs from the other three (Gruzdeva 1998: 7). There are then four separate varieties (cf. Mattisen 2003: 5; Gruzdeva 1998: 7) or the North Sakhalin and South Sakhalin varieties are included in the Sakhalin group (cf. Vovin 2016: 30) or the North Sakhalin variety is included in the Amur group and the South Sakhalin variety is included in the Sakhalin group together with the East Sakhalin variety, as is the case for Gruzdeva and Janhunen (2020: 3, 8-9). For the purposes of this thesis, this last classification is the most practical.

The territory currently occupied by the Amuric languages – the Lower Amur and the Island of Sakhalin – is surrounded by languages belonging to the Tungusic language family. A discussion about terminology can also be had as regards to the Tungusic label (cf. Pevnov 2012: 17), however, in this thesis I maintain this label as it has increasingly become the established term and its connotation seems to be more or less neutral. The internal classification of the language family seems to be a matter that is still far from being settled.[2] Yet there are four subgroupings that are more or less uncontroversial and can reliably be defined: a Northern complex, a Nanai complex, an Udihe complex and a Manchu complex (Oskolskaya & Whaley 2020: 91). Following Janhunen (2019: 6), I refer to these complexes or branches as Ewenic, Nanaic, Orochic and Jurchenic respectively. Ewenic consists of Ewen, Ewenki, Khamnigan Ewenki, Neghidal, Oroqen and Solon. Nanaic consists of Nanai, Uilta (also called Orok), Ulcha, Kili (also called Kur-Urmi Nanai) and Kilen (also called Hezhe). Orochic consists of Oroch and Udihe. Jurchenic consists of Jurchen, Manchu and Sibe. Depending on the analysis, some of these languages may be described as dialects of other languages, which would reduce the number of languages within each subgroup.

In geographical terms, the Ewenic branch represents the greatest territorial expansion of any branch of Tungusic. They are found scattered throughout an area that reaches

[2] See Oskolskaya and Whaley (2020: 81-91) for an in-depth discussion.

as far east as the Sea of Okhotsk coast and as far west as the Yenisei basin. In the south, there are also Ewenic groups living in Northern Manchuria, Sakhalin and along the Amur river (Janhunen 1996: 67-73).

The branch located furthest to the south is the Jurchenic branch, whose speakers were the founders of various polities. Jurchen was the language of the founders of the Jin empire (1115-1234), while its descendent Manchu was the language of the founders of the Qing dynasty (1644-1911). The latter came to encompass all of China, Mongolia and Tibet. As a result, ethnic Manchus could be found throughout the empire. Though the Manchus nowadays form one of the largest minorities in China, there are very few speakers left of the Manchu language, mostly confined to the north-eastern section of China's Heilongjiang province. The Sibe language, itself a descendent of Manchu, is currently spoken outside of Manchuria: in the Ili region of northern Xinjiang (Bradley 2007: 380; Janhunen 1996: 46-50, 98).

Geographically in between the Jurchenic and Ewenic branches, there are Orochic and Nanaic. Orochic is located mainly in the area surrounding the Sikhote Alin range. Nanaic is spread out over much of the Amur valley, along the Ussuri and lower Sungari rivers and Lake Khanka, as well as across the Tartar Strait on Sakhalin (Zgusta 2015: 104-109, 155).

The Lower Amur is the area of the greatest density of Tungusic languages. Figure 2 displays this region as well as Sakhalin island together with points of settlement of Tungusic and Amuric peoples as they were during the 19th century and the beginning of the 20th century. In order, the numbers correspond to following peoples: (1) Nanai, (2) Ulcha, (3) Nivkh[3], (4) Neghidal, (5) Uilta, (6) Oroch, (7) Udihe, (8) Ewenki. This map shows that anywhere the Amuric-speaking Nivkh were located, there was a Tungusic-speaking people in their vicinity. Yet it should be noted that the language people speak and their ethnicity, though interconnected, are not necessarily overlapping. Nevertheless, this map most likely also reflects the linguistic situation of that period accurately.

[3] Used here in the ethnonymic sense. Linguistically, these markers may indicate both speakers of Nivkh and Nighvng varieties.

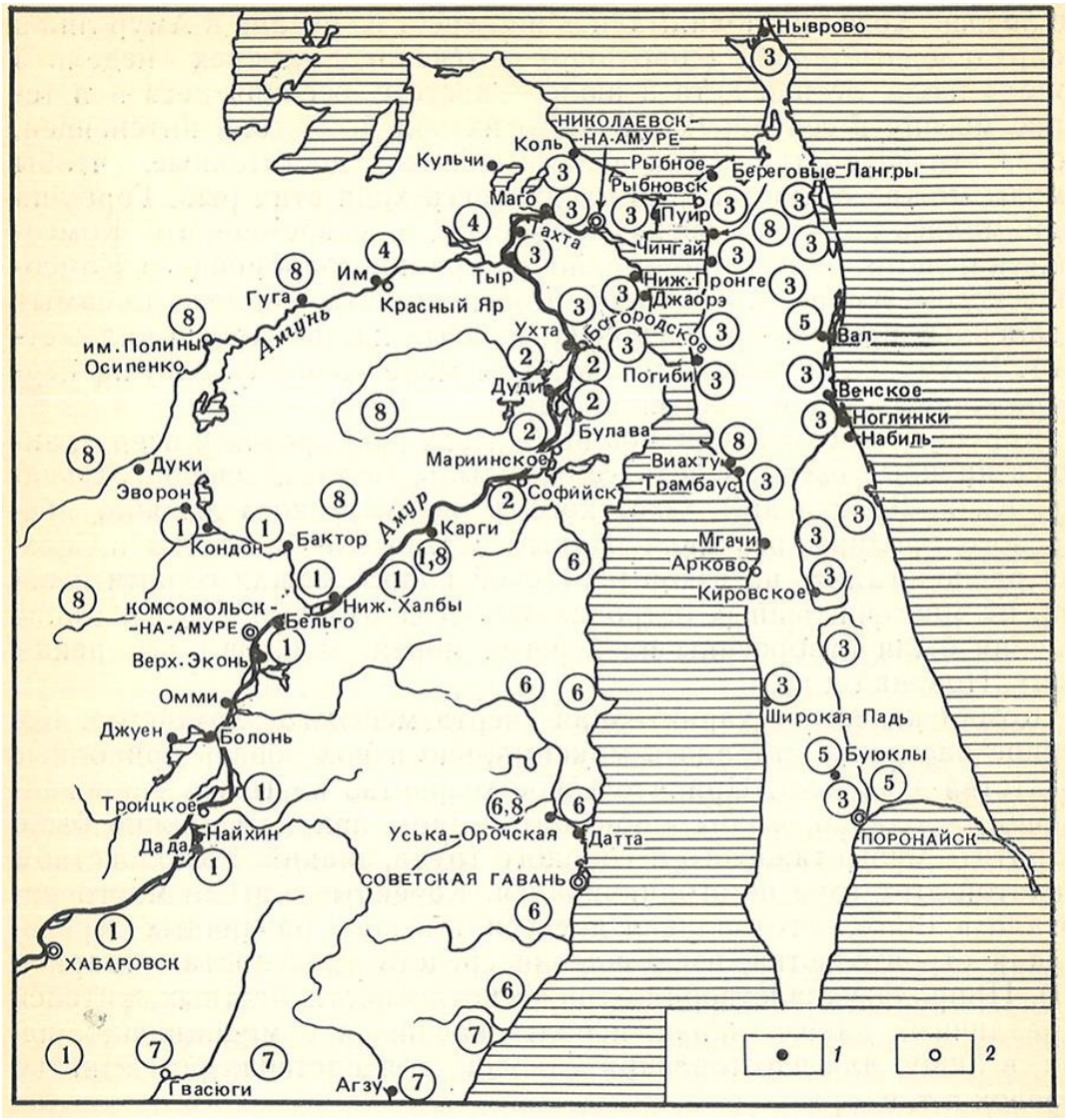

Figure 2: Map of settlements of the peoples of the Lower Amur (19th century till the first half of the 20th century) (Smoljak 1984: 26)

In sum, Tungusic and Amuric-speaking peoples have long inhabited the same regions. On the Lower Amur, Amuric has been surrounded by Nanaic languages to the South and Ewenic languages to the North. These subgroupings have also been neighbours of Amuric on Sakhalin. This geographic proximity needs to be considered for the later stages of Amuric-Tungusic contacts. Yet, for the oldest stage it is most important to locate their place of origin, to which I turn in the next section.

3 The Amuric and Tungusic homelands

The geographical location of a certain language family's ancestor is referred to as a "homeland". This is the place from which the spread of the language family started. There is also an additional differentiation between "immediate" and "ultimate" homelands. The immediate homeland is the area most recently occupied by a given family's protolanguage. If there is additional evidence available that allows the ultimate origin of this language family to be located at a different place at an earlier point in time, then this place is its ultimate homeland (Robbeets et al. 2020: 753). This is the basic terminological distinction made in this section. The ultimate homelands of the Amuric and Tungusic language families as well as other Northern Asian language families are shown in Figure 3. The localisation as well as the map itself is based on Janhunen (1996: 216).

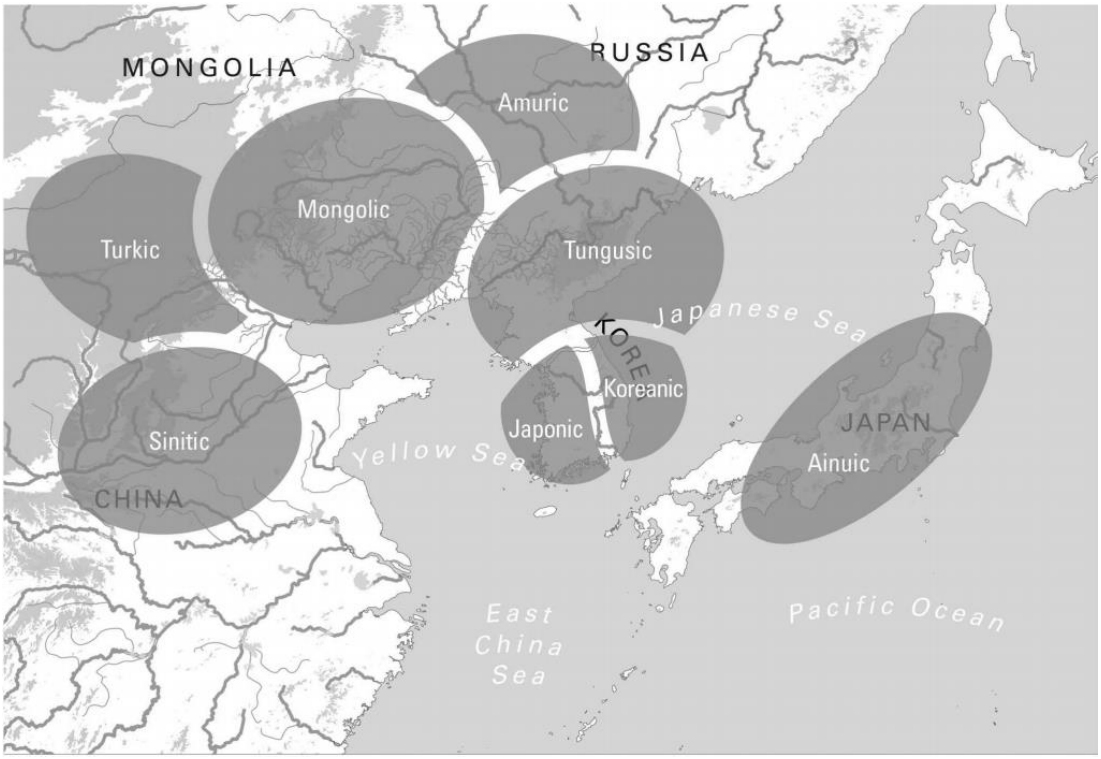

Figure 3: Approximate homelands of the ethnolinguistic groups of Manchuria in the first millennium BC (Robbeets et al. 2020: 754)

In Figure 3, the ultimate homeland of Amuric is placed in the Sungari-Amur basin, as argued for by Janhunen (1996: 231; 2016: 5). Some evidence of this southern origin can be found in the native metallurgical terminology, which would represent

an activity more suited to an environment than is currently inhabited by Amuric speakers (Janhunen 2016: 6). An additional piece of evidence could be found in recent research on toponyms, significantly the hydronym *Sungari* itself, which may contain the Amuric element *eri* 'river' (Gruzdeva & Temina 2020: 173). A third would be the loanwords shared between the Amuric and Tungusic languages (Janhunen 2010: 294). These loans are analysed in this thesis.

South to the Amuric ultimate homeland, Figure 3 presents one of the various proposals for the location of the Tungusic *Urheimat*.[4] This area corresponds to the region of Liaodong, as proposed by Janhunen (2010: 295; 2012: 10). Robbeets et al. (2020: 762-761) accept this region as the ultimate homeland of a stage preceding Proto-Tungusic: Macro-Tungusic. However, according to them, the immediate homeland of Proto-Tungusic would be further to the north, around Lake Khanka. Regardless of which region is the ultimate homeland, both are geographically close to the Sungari-Amur basin and therefore either region would have accommodated contact between Proto-Tungusic and Pre-Proto-Amuric.

As discussed in Section 1, Janhunen (2010: 292, 296; 2016: 23) regards Proto-Tungusic and Pre-Proto-Amuric as linguistic units that must have existed at roughly the same point in time, somewhere between 1,000 to 2,000 years before the present. Though the absolute dating is not particularly exact, a relative dating to the Iron Age for Proto-Tungusic seems to be relatively uncontroversial. The term **sələ*, which refers to not just any metal but iron specifically, can be reliably reconstructed for Proto-Tungusic. Within the archaeological record, the start of Iron Age in this area is represented by the Krounovska culture, which is dated between 600 BC to 200 AD (Robbeets & Wang 2020: 5). Such a specific term for 'iron' can also be reconstructed for (Pre-)Proto-Amuric – given as **wat* by Fortescue (CND: 160). However, the internal diversification of Amuric provides only an estimate of around 500 years before the present for the breakup of Proto-Amuric (Janhunen 2010: 290). This is much more recent than the fragmentation of Proto-Tungusic that Robbeets and Wang (2020: 5), combining different proposals, estimate to lie between 600 BC and 700 AD.

[4] See Robbeets and Wang (2020) for a detailed assessment of the different hypotheses.

To date the ultimate ancestor of the Amuric language family, i.e. Pre-Proto-Amuric, to the same point in time as Proto-Tungusic, Janhunen (2016: 23) has relied on loanwords. On their basis, using a method referred to as “external reconstruction” (cf. §6), he has reconstructed this ancestor and concluded that it and Proto-Tungusic were in contact. Therefore, they must have existed concurrently.

The discussion above shows again the importance of the study of Amuric-Tungusic contacts for the purposes of understanding the history of the languages involved. Hypotheses on Pre-Proto-Amuric as well as a more southern ultimate homeland for the language family that sprung from it have so far required there to be proof of very early contacts with Tungusic. Not just any Tungusic language, but the ancestor of the family specifically. For that reason, it is crucial to study whether any ancient layer within the available lexical parallels can be found. A long list is currently available to me, provided by previous research that I discuss in the following section.

4 Previous research

Section 1 presented a short overview of the history of scholarship on Amuric-Tungusic lexical parallels starting with Grube (1892). This line of inquiry was continued by subsequent scholars, among whom can be counted von Möllendorff (1894), Laufer (1902), Krejnovič (1937; 1955a), Austerlitz (1956; 1977; 1983b; 1983c; 1984a; 1984b; 1985; 1986b), Bouda (1960), Jakobson (1971), Panfilov (1973), Cincius (1975; 1977), Doerfer (1985a; 1958b), Janhunen (1986; 1998; 2005; 2008; 2011; 2014; 2016), Burykin (1987; 1988; 1989; 2005; 2012), Pevnov (1992; 2016; 2018), Blažek (1997), Shternberg (1999), Abondolo (2005), Burykin and Popov (2012), Robbeets (2015), Tangiku (2015; n. d.), Shiraishi and Botma (2016) and Fortescue (CND). This list is extensive but perhaps not exhaustive. Some authors focus on providing long lists of words while other just mention a single item in passing. Nevertheless, this may be the most complete collection gathered so far.

The above cited works represent different theoretical assumptions. The most commonly held position is that the similarities in the lexicons of the Amuric and Tungusic languages are the result of borrowing. Yet some of the scholars in this list, i.e von Möllendorff (1894), Burykin (1987; 1988; 1989; 2005; 2012), Burykin and

Popov (2012) and Panfilov (1973), have regarded them as evidence for common ancestry of Amuric and Tungusic. After careful consideration, I have found their arguments to be unconvincing. Therefore, I follow the former position in this thesis and assume that the words shared between these two language families are the result of borrowing.

The cited sources present the trajectory of development within the field. The first significant point on this timeline is von Möllendorff (1894), who was the first to study Amuric-Tungusic lexical parallels as a specific topic. He largely builds upon Grube (1892), his predecessor, though Grube was mainly concerned with lexicography and language description in his publication. Although von Möllendorff may be the first in this line of research, his work is apparently relatively obscure, as his name does not show up in later publications. Von Möllendorff's main objective seems to have been compiling a list of Amuric words that bear similarity to Tungusic words and this has been the primary concern for much of the subsequent research. Yet sometimes there were also attempts of making morphological analyses (e.g. Bouda 1960) on the basis of these parallels. The first systematic analysis of regular sound correspondences is represented by Krejnovič (1955a), which can be regarded as the foundational text of this field and is cited by most later scholars. Some additional sound correspondences were proposed by Burykin (1987; 1989), though they show varying degrees of regularity. The greatest advancements were made by Janhunen (2016), who again showed that there were regular sound correspondences and started employing them for the purpose of reconstruction (i.e. "external reconstruction"). He also made an attempt of identifying chronological strata of borrowing as well as directions of borrowing. Yet he applied his methodology on a small set of words, which has left many questions still unanswered.

The direction taken by Janhunen (2016) presents the area where most progress is to be made in the study of Amuric-Tungusic lexical parallels. External reconstruction, direction of borrowing and the stratigraphy of borrowing all need to be expanded on. This thesis makes an attempt to add to the latter two by investigating the oldest layer of borrowing.

5 Theoretical background: borrowing versus inheritance

Campbell (2013: 56) defined a loanword as "a lexical item (a word) which has been 'borrowed' from another language, a word which originally was not part of the vocabulary of the recipient language but was adopted from some other language and made part of the borrowing language's vocabulary". "Borrowing" is a commonly used metaphor to describe language contact. In this terminology, the related term "loanword" is applied to a complete unit of phonological material that is borrowed together with its meaning (cf. Haugen 1950), which is how "a lexical item" in Campbell's definition should be interpreted. The word "originally" in his definition encapsulates the distinction between borrowing and inheritance that is important to the description of a language's history: they refer to different modes of transmission.

In the case of inheritance, a morpheme was transmitted from an earlier stage of a certain language to the next. In the case of borrowing, a morpheme was taken from one language and inserted into the morphosyntactic frame of another. The distinction between these two also implies that there is a direction to borrowing. There should always be a language for which the word is inherited and therefore original (the donor language) and another for which it is not (the recipient language). The movement of this word was *from* the donor language *to* the recipient language. These languages do not share the same earlier stage, or alternatively put, the same ancestor.

Differentiating borrowing from inheritance is important to the historical linguist, since only inherited lexemes can prove relatedness while borrowed lexemes cannot. Both Greenberg (1957: 38) and Campbell (2013: 65-66) have observed that loanwords are often restricted to certain semantic domains. This reveals the specific socio-cultural context of borrowing. By way of illustration, the latter may be concluded from some of the findings of Pevnov (1992: 26) and Doerfer (1985b: 196-199) who found that many Amuric-Tungusic lexical parallels fall within semantic domains relating to aquatic fauna and fishing. It is not inconceivable that these animals and ways of catching them were known to speakers of only one of the two linguistic groups. The other then simply may have borrowed both the knowledge as well as the words connected with that knowledge from them. However, while Greenberg has argued that shared vocabulary restricted to semantic domains is proof

of borrowing, Campbell (2013: 65-66) regards it only as “rough indication”. Much stronger proof of borrowing can be found if one considers the differences in formal expression of borrowing and inheritance.

These different modes of transmission, or rather the interplay between the two, have different effects on the developments within the lexicon of a language family. Items that are inherited follow sound laws that know no exceptions. The regularity of these sound laws explain how languages may exhibit regular sound correspondences that are both “productive and predictive” (Austerlitz 1983a: 52-53). This means that if one asserts that segment X in language A corresponds to segment Y in language B, one should be able to predict that the segment X will be present in language A wherever segment Y is present in language B and vice versa. If languages share the same relatively long sequence of segments, i.e. a word, then this is a strong indicator that both of these languages received the word from a single source, since the probability that *any* language has this sequence is very low, as shown by Nichols (1996: 50-54). Although regular sound correspondences do not have to be between identical segments. The segments can also be only similar and the correspondence can still be regular. In that case, the probability increases however. The solution is generally to use a large dataset that proves that the correspondence is regular and this partly mediates this effect, since it then creates a structured set of words, which are unlikely to be found in any language, which may be compared to Nichols’s example of numerals. Yet if it is shown that the two languages acquired a given word from a single source, then that still does not have to be proof of relatedness. As Nichols (1196: 50) notes “[t]he languages need not all be descended from that source; the word could be a loan in some of the languages.” This is because a loanword may show regular sound correspondences when compared to the original that was its model, just as cognates do. However, this is not the result of sound change but of another process: phonological adaptation.

If the donor language and recipient language are phonologically different, then words borrowed from the donor language may contain structures illicit in the phonology of the recipient language. These structures are then amended so that they conform to the restrictions of the borrowing language. This process is referred to as “phonological adaptation” (hereafter PA). Kang (2011) discusses that there is a degree of

systematicity to PA. For example, this systematicity is found in segmental adaptation. “When the foreign input contains a non-native segment, the segment is replaced with the “closest” sound in the native language.” (Kang 2011: 2265). The result are regular sound correspondences between a loanword and a matching word in the donor language. These correspondences are certainly productive and predictive, but this usually only holds true for a section of the lexicon. This is caused by the interplay of sound change and PA.

Sound laws are only operational at a certain point in time. After this point, they cease to be productive. Stages within a language’s development are therefore characterised by different innovations. All words that are inherited from its ancestor partake in all these innovations. Conversely, words that are borrowed during one stage, only participate in the sound laws that are applied after the time when they were borrowed. They do not participate in any of the preceding. If there are several layers of loans and they are compared to the apparent originals in the donor language, then one layer of loanwords will reflect one set of correspondences, while another layer will reflect another set of correspondences. This sets loanwords apart from cognates and provides a means of formal differentiation.

6 Tungusic and Amuric historical phonologies

The process of PA presented in the previous section targets those structures in loanwords that violate the segmental, phonotactic, suprasegmental and morphophonological restrictions of the recipient language (Kang 2011: 2258). Because these restrictions can change over time, a diachronic analysis of loanwords also requires a sufficiently detailed account of how the phonologies of the languages that were in contact developed. For that reason, I discuss what is presently known about the phonologies of (Pre-)Proto-Amuric and Proto-Tungusic in this section.

Janhunen (2016; 1981: 131) has argued that, in this regard, these two languages were very much alike. The consonant inventories of Pre-Proto-Amuric and Proto-Tungusic were most likely identical. Their vowel inventories were also comparable, though not same (the closest equivalent of (Pre-)Proto-Amuric’s vowel inventory is found in the later Tungusic languages of the Lower Amur). On the level of the syllable, Pre-

Proto-Amuric must have allowed basically the same structures as Proto-Tungusic as well. These same points were reiterated in Gruzdeva and Janhunen (2020).

If it was indeed the case that Pre-Proto-Amuric and Proto-Tungusic were basically identical in terms of phonology, then that would greatly impact what proposed items can be regarded as evidence of borrowing. The implication is that no Pre-Proto-Amuric structures would have been illicit in Proto-Tungusic and vice versa. If that were the case, it is not to be expected that loanwords would require PA.[5] Pairs of Pre-Proto-Amuric and Proto-Tungusic words that are neither fully nor nearly identical must then be dismissed.

While Proto-Tungusic can be reconstructed relatively accurately using the standard method of historical comparative linguistics – the Comparative Method – this is not the case for Pre-Proto-Amuric. Interdialectal comparison has only succeeded in reconstructing Proto-Amuric but so far failed to reconstruct Pre-Proto-Amuric, Proto-Amuric's forebear that would have been as ancient as Proto-Tungusic. To reconstruct Pre-Proto-Amuric historical linguists have had to rely on other methods as well. Janhunen (2016: 6) distinguishes three: the aforementioned Comparative Method but also "internal reconstruction" and "external reconstruction". This subdivision is based on the types of data these methods rely on. The Comparative Method uses synchronic data from two or more related languages, internal reconstruction relies on synchronic data within a single language and external reconstruction considers data from two or more unrelated languages. Many of the findings of these methods are still – to a degree – speculative. Yet by the end of this section it should be clear on what basis it can be assumed that Pre-Proto-Amuric and Proto-Tungusic were indeed all but identical in terms of phonology.

6.1 (Pre-)Proto-Amuric and Proto-Tungusic vowel inventories

It has been observed that Amuric and Tungusic vowel systems share the characteristic of vowel harmony (hereafter VH) (cf. Janhunen 1981; Ko et al. 2014:

[5] That is however not to say that this would not result in any alteration to loanwords, since there are also instances of so-called "unnecessary repair" discussed by Kang (2011: 2260). Yet for most of her examples, there are alternative explanations that put into question whether this repair is unnecessary at all (cf. Boersma & Hamann 2009).

167; Joseph et al. 2020). In both languages, vowels are arranged in pairs or sets in which each vowel forms the other's harmonic counterpart. To illustrate, Table 1 displays the vowels of Proto-Tungusic as they have been reconstructed by Benzing (1956: 19-25) and following Janhunen's (1981: 136) interpretation of this reconstruction.

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	*i	*ü	*ï	*u
Low	*ä	*ö	*a	*o

Table 1: Proto-Tungusic vowel phonemes (Benzing 1956: 19-25; Janhunen 1981: 136)

These vowels could be both long and short. Proto-Tungusic must also have had a number of diphthongs with one member always being a high vowel (Benzing 1956: 19-25; Ryzhkov-Shukumine 2020). Rearranging these vowels into two harmonic sets, results in Table 2.

Set 1 (Front)	*i	*ü	*ä	*ö
Set 2 (Back)	*ï	*u	*a	*o

Table 2: Proto-Tungusic harmonic sets

Set 1 and Set 2 contrast in frontness, which is the "active" feature in VH according to this reconstruction. This means that roots in Proto-Tungusic could consist of vowels belonging to either Set 1 or Set 2 only. Suffixes in the languages also had to harmonise to the vowels found in the roots. This system is referred to as "horizontal" or "palatal" harmony (hereafter PH). Nearly all modern Tungusic languages have a different type of vowel harmony referred to as "apertural" or "retracted tongue root" harmony (hereafter RTRH, in which RTR stands for "retracted tongue root").[6] Instead of front vs. back, the opposition is based, in some form, on the relative size of the pharynx in the articulation of these vowels, which is the result of retracting or not retracting the tongue root (Li 1996: 84-122). According

[6] In some only remnants are found. Additionally, western varieties of Ewen seem to be the only idioms with PH instead of RTRH, although this is most likely the result of contact with Yakut (Barrere & Janhunen 2019: 50).

to Janhunen (1981: 136-139) the difference between Proto-Tungusic and its descendants in terms of VH is due to a process he has termed “vowel rotation” in which Set 1 vowels changed from [+front] to [-RTR] and Set 2 vowels changed from [-front] to [+RTR].

Li (1996: 23-24, 84-122) has analysed the active feature of VH in Baiyinna Orochen as RTR. This is a Tungusic language belonging to the same branch as Ewenki. Example (1) shows how VH operates in this language.

(1) a. *nəlki-nii* ‘of spring’

b. *ʤu̯wa-ni̯i* ‘summery’ (Li 1996: 127)

In (1) the suffix *-nII* agrees with the root in terms of RTR. In (a), the root *nəlki* ‘spring’ consists of only [-RTR] vowels (corresponding to Set 1 in Table 2) and takes the allomorph *-nii*, which contains a [-RTR] vowel as well. In (b), its allomorph with a [+RTR] vowel, corresponding to Set 2 in Table 2, is found after *ʤu̯wa* ‘summer’, a root with only [+RTR] vowels.

Proto-Amuric can be described in a similar manner, though compared to the vowel system of Proto-Tungusic, the vowel inventory of Proto-Amuric was somewhat different. The reconstruction of this system found in CND is essentially identical to the one found in Table 3 for the attested Amuric languages.

	Front	Central	Back
High	(*)i	(*)ə	(*)u
Low	(*)e	(*)a	(*)o

Table 3: Vowel phonemes (adapted from Hattori 1962a: 1; 1962b: 68)

No length contrast can be reconstructed. Diphthongs have been reconstructed however. In these diphthongs, the second member was always a high vowel. According to Halm (2017) such sequences were differentiated from vowel-glide sequences.

The active feature in the Amuric system of VH is not frontness, but vowel height instead. Its Set 1 consists then of the high vowels **i, *ə, *u*, while Set 2 consists of **e, *a, *o*. In contrast to Tungusic, where vowel harmony is applied progressively (i.e. from the front of the word to the back of the word), in Amuric it operates regressively (i.e. from the back to the front of the word). Consider examples (2) and (3).

(2) a. *i-rlə-* ‘pull’

b. *e-rʁop-* ‘touch’ (Shiraishi & Botma 2017: 206)

(3) a. *u-ndə-* ‘see each other’

b. *o-ŋala-* ‘resemble each other’ (Shiraishi & Botma 2017: 206)

Example (2) shows the realisations of the third person undergoer prefix **I-*, which has the allomorphs *i-* and *e-* here. Example (3) shows the realisations of the reciprocal prefix **U-* with the allomorphs *u-* and *o-* here. Both of these prefixes also have a consonantal allomorph, (*)*j-* and (*)*w-* respectively, for which no behaviour according to VH is recorded (Gruzdeva 1998: 29-32). As can be seen, roots headed by a high vowel take the allomorphs with a high vowel, while roots headed by low vowels take the allomorphs with a low vowel.

Root-internally, this type of vowel harmony seems to be present only rudimentarily. Hattori (1962a: 78-82) has shown for SSN that the second syllable in a polysyllabic word never contains a low vowel if the first syllable contains a high vowel. The same was also found in AN by Shiraishi and Botma (2016). Both publications have also shown that high vowels are allowed in the second syllable even if the first syllable contains a low vowel. Such behaviour is generally not allowed in a VH system. Shiraishi and Botma (2017: 208-210) explain it as a result of vowel reduction in unstressed syllables, which raises vowels. Alternatively, there may be a historical explanation to this apparent violation of VH. According to Janhunen (2016: 22) violations of VH are reflective of compounding. In his line of reasoning, when a

word is made up of syllables headed by vowels of two different harmonic pairs, these syllables were originally part of two separate words. Because of that they would not have had to follow VH. In the case of (*)*i*, he has suggested that some examples may ultimately go back to **Vj*, for which he has shown evidence in the same publication. As a consonant, **j* would not have to conform to VH. Therefore, Pre-Proto-Amuric may have had fully regular VH, which cannot be reconstructed for Proto-Amuric on the basis of the attested languages.

When comparing the vowel inventories of Proto-Tungusic and Proto-Amuric, it seems clear that they were quite different. Proto-Tungusic had two vowels more than Proto-Amuric as well as different contrasts in vowels. This is also reflected in the types of VH they possessed: Proto-Tungusic apparently had PH, while Proto-Amuric had VH based on vowel height.

It has been suggested these two vowel systems originally were more similar (excepting the direction in which VH operated). Ko et al. (2014) and Joseph et al. (2020) have argued that in Proto-Tungusic as well as other so-called "Altaic" or "Transeurasian" languages the original type of VH was RTRH, not PH. They have presented two arguments: (1) the majority of these languages follow RTRH and (2) there is a clear phonological path from RTRH to PH. Since nearly all Tungusic languages are attested with RTRH[7], the latter argument is not relevant, but the former certainly is. As the majority of the Tungusic languages have this type of vowel harmony, it seems that it should also be reconstructed for Proto-Tungusic.[8] The resulting reconstruction is as given in Table 4 with the corresponding vowels from the more traditional reconstruction given in brackets.[9]

[7] As stated in note 6, western varieties of Ewen are the only exception, which may have a different explanation (Barrere & Janhunen 2019: 50).
[8] Though see Barrere and Janhunen (2019: 59-60) for arguments in favour of PH.
[9] In Table 4, I follow Aralova's (2015) convention of writing RTR vowels with an underdot, since it also common in Russian language literature and as a result is also reflected in the romanisation scheme I have used (see Appendix A). Different conventions are also used. For instance, Li (1996) uses the IPA characters for lax vowels.

	non-back		back	
	nonRTR	RTR	nonRTR	RTR
High	*i (*i)	*ị (*ï)	*u (*ü)	*ụ (*u)
Low	*ə (*ä)	*a (*a)	*o (*ö)	*ọ (*o)

Table 4: Proto-Tungusic vowel phonemes reinterpreted

The forerunner of the Amuric system of VH may also have been RTRH, as suggested by Janhunen (1981: 139-140), Ko et al. (2014: 167), Joseph et al. (2020: 505).[10] In this scenario, low vowels are originally RTR vowels. Shiraishi and Botma (2017) found evidence of remnants of an RTR-contrast in NSN (or Shmidt dialect per their terminology). This dialect has an additional high back vowel /ʊ/, which only occurs after uvulars and stands in opposition with /u/ but corresponds to /o/ in other varieties. They posit that after losing the contrast between /u/ and /ʊ/, the contrast based on tongue root was reinterpreted as a height contrast.

This scenario may also explain some apparent irregular correspondences between Amuric languages. For instance, when comparing AN *laʁo-* 'be frightened' to ESN *laʁu-* 'id.' listed by Savel'jeva and Taksami (NRS: 153), the correspondence of AN /o/ to ESN /u/ is unexpected. A solution may be to propose an additional vowel **ụ* for Proto-Amuric as well as Pre-Proto-Amuric that merged with (*)*o* in AN and with (*)*u* in ESN. NSN, again, provides proof of this vowel, since the relevant cognate listed by Shiraishi and Botma (2017: 211), *laʁʊ-* 'be surprised', has /ʊ/.

This hypothesis would also explain Janhunen's (2016: 21) observation that PAm. **o* sometimes corresponds to (*)*ụ* in Tungusic. Yet his examples – Ma. *ulgʲan* 'pig', Na. *ọlgịan* 'id.', AN *olɣoŋ* 'id.' – may not be sufficient evidence, since he is unclear on the direction of borrowing. It appears that it is a borrowing from Amuric, since it cannot be confidently reconstructed for Proto-Tungusic with attestations limited to the Tungusic languages of the Amur Basin and Manchuria (cf. SSTM2: 259). A more convincing example may be found in Uilta. Pevnov (2016: 58; 2018: 449) and

[10] Barrere and Janhunen (2019: 73) argue that VH in Pre-Proto-Amuric was originally PH, as they analyse the opposition in the dorsals as velar vs. postvelar. In this thesis, I follow the more common analysis of this opposition as velar vs. uvular instead. As uvulars are consonants specified for [+RTR] (cf. Ko et al. 2014: 145-147), Pre-Proto-Amuric VH must therefore be analysed as RTRH.

Austerlitz (1977: 19) both compare Uil. *udala(n)* ‘frog’ to AN *ral* ‘id.’[11]. Given that this word is not attested in any other Tungusic language, it is most likely that Uilta borrowed it from an Amuric language, as suggested by these authors.[12] If VH was based on height in Pre-Proto-Amuric as it is in the attested Amuric languages, then the initial vowel of the words for ‘pig’ and ‘frog’ should be reconstructed as **o*. However, it seems unlikely that **o* would be categorised as /u/ in Uilta, given that the language has two low back rounded vowels that would be closer equivalents. The same holds true for Manchu and probably also for Nanai in an earlier period, making the categorisation of PAm. **o* to (*)*ụ* equally peculiar, assuming that they borrowed their word for ‘pig’ from an Amuric language. Based on this fact, as well as the aforementioned /ʊ/ found in NSN, I would tentatively reconstruct **ụ* for (Pre-)Proto-Amuric, which would be the RTR counterpart of non-RTR **u* in VH.

With the addition of **ụ*, the vowel inventory of (Pre-)Proto-Amuric is all but identical to the inventory given for Tungusic in Table 4. The only missing phoneme is **o*, assuming that later /o/ derives from earlier **ọ* in accordance with the hypothesis that low vowels were originally RTR vowels. Yet given that both protolanguages were likely characterised by RTRH, it is to be expected that **o* existed in (Pre-)Proto-Amuric as well, making the inventories completely identical.

6.2 (Pre-)Proto-Amuric and Proto-Tungusic consonant inventories

As stated in the introduction to Section 6, Janhunen (2016) and Gruzdeva and Janhunen (2020) have argued that Pre-Proto-Amuric had essentially the same set of consonant phonemes as Proto-Tungusic. That would entail that Pre-Proto-Amuric had the same inventory as is presented in Table 5 for Proto-Tungusic.

[11] Ikegami (1980: 47) lists *udala* and *udalamba* as its accusative, suggesting that the the underlying stem is /udalan/.
[12] Cincius (SSTM2: 248) does relate this word to Ewk. *udun* ‘rain’ but there seems to be no plausible explanation for how *udala* would have been morphologically derived from this word. Therefore, this connection is doubtful.

	Labial	Dental	Palatal	Velar
Voiceless plosive	*p	*t	*č	*k
Voiced Plosive	*b	*d	*ʤ	*g
Voiceless fricative		*s		*x
Nasal	*m	*n	(*ɲ)	*ŋ
Liquid		*r, *l		
Glide	*w		*j	

Table 5: Proto-Tungusic consonant phonemes (adapted from Benzing 1956: 27; Whaley & Oskolskaya 2020: 84)

The reconstruction of the Proto-Amuric consonant inventory that can be reached through the Comparative Method is somewhat different. One version of which was provided by Fortescue (CND), which is given in Table 6.

	Labial	Dental	Palatal	Velar	Uvular	Glottal
Fortis plosive	*p	*t	*c	*k	*q	
Lenis plosive	*b	*d	*ɟ	*g	*ɢ	
Lenis fricative	*v	*r	*z	*ɣ	*ʁ	
Nasal	*m	*n	*ɲ	*ŋ		
Liquid		*l				
Glide	*w		*j			*h

Table 6: Proto-Amuric consonant inventories (CND: 5)

Compared to Proto-Tungusic, Proto-Amuric seems to have had almost the same set of plosives, though with the addition of a uvular column. Yet the contrast between fortis and lenis obstruents is also rendered rather deceptively as a contrast on the basis of voice, which is not accurate to the attested Amuric languages. Although Fortescue does not argue that the fortis-lenis contrast was a voice contrast.

Another criticism that can be levelled at this model is the single set of fricatives. This postulate has apparently forced Fortescue to reconstruct PAm. **tə* 'door', which is attested as *ɾ̥ə* – with a voiceless or fortis initial fricative, not a plosive – in all Amuric languages (CND: 153). Additionally, there is the contrast in fricatives that has been

described for SSN. This variety had two sets of fricatives, one that becomes voiced in intervocalic position and one that remained voiceless in the same position. These two sets also behaved differently in morphophonological processes (Hattori 1962a: 102-103; 1962b: 30). Mattissen (2003: 40-41) analyses the first set as underlyingly voiced, which devoices in word-final position. According to Mattissen, this contrast between word-final voiced and voiceless fricatives also exists in the Nivkh branch, but Shiraishi (2006: 70-74) shows that this contrast does not exist in WSN, which belongs to the same branch. He suggests that Mattissen's analysis may have been based on a faulty dataset. Her source is the dictionary by Savel'jeva and Taksami (NRS), which contains many inconsistencies in spelling of final fricatives. Possibly, the contrast was neutralised in the entire Nivkh branch, which makes SSN possibly archaic in this regard. Fortescue (CND: 4) regards these fricatives as secondary but his argumentation cannot account for the behaviour described for SSN.

Halm and Slater (2018: 33-35) have stated that reconstructions with and without this contrast have to be accepted provisionally. Acceptance of the former results in a reconstruction of Proto-Amuric with a phonology that was essentially identical to the one found in all attested Amuric languages. An overview of this inventory is laid out in Table 7.

	Labial	Dental	Palatal	Velar	Uvular	Glottal
Fortis plosive	pʰ	tʰ	cʰ	kʰ	qʰ	
Lenis plosive	p	t	c	k	q	
Fortis fricative	f	r̥	s	x	χ	
Lenis fricative	v	r	z	ɣ	ʁ	
Nasal	m	n	ɲ	ŋ		
Liquid		l				
Glide	(w)		j			h

Table 7: Amuric consonant phonemes (adapted from Gruzdeva 1998: 10; Hattori 1962a: 2; 1962b: 68; Shiraishi 2006: 24)

In the table above, the consonants are presented in the way they are realised, which means that fortis plosives are given as voiceless aspirated plosives and lenis plosives are given as voiceless unaspirated plosives. Gruzdeva (1998: 10-11) also recognises

a contrast between voiceless and voiced lenis plosives in AN, WSN and NSN, although this is a relatively recent development caused by post-nasal voicing. These nasals were subsequently lost, creating a phonemic contrast between voiced and voiceless lenis plosives.

Another detail pertains to the glide /w/. This is given in brackets in Table 7, since it has been merged with /v/ in AN and WSN (Jakobson 1971: 93-94; Halm & Slater 2018: 24-25; Gruzdeva 1998: 11; Shiraishi 2006: 28). According to Halm and Slater (2018: 24-25), the development also occurred in NSN, while Gruzdeva (1998: 11) states that it did not.

Aside from Fortescue (CND), other authors have attempted to reconstruct a more archaic form of this consonant inventory using the Comparative Method as well as internal reconstruction and external reconstruction. In the following subsections, I focus on a number of hypotheses that have been put forward that drastically reduce the number of consonantal oppositions for Pre-Proto-Amuric.

6.2.1 The fricative-plosive contrast in Amuric

As stated before, the characters used in Table 7 reflect the current phonetic value of the consonants listed therein. While there is a voice contrast in the fricatives, this does not exist in the plosives, which are contrasted for aspiration instead. These two contrasts are best reinterpreted as a single opposition, which can be inferred from a morphophonological process termed “consonant mutation” or “consonant alternation” (hereafter CA). The target of this process is the initial consonant of a morpheme at the boundary of two morphemes. The input and output sequences are given in (4) (adapted from Shiraishi 2006: 58).

(4)	Input sequences		Output sequences
	Vowel - Plosive	>	Vowel - Fricative
	Glide - Plosive	>	Glide - Fricative
	Plosive - Plosive	>	Fricative - Fricative
	Fricative - Fricative	>	Fricative - Plosive
	Nasal - Fricative	>	Nasal - Plosive

The first member of the sequence represents the final segment of the morpheme that triggers the alternation. The initial consonant of the following morpheme, the second member of the sequence, is the one that alternates. It applies to combinations of the types affix-root, root-suffix, attribute-noun and object-verb (Gruzdeva 1998: 14). As is shown in (4), alternation only affects the realisation of segment specified for [+obstruent]. Fortis plosives alternate with fortis fricatives and lenis plosives alternate with lenis fricatives. Fortis vs. lenis is therefore the main contrast in Amuric obstruents and this is why fricatives and plosives are classified as either “fortis” or “lenis” in Table 7. The description of these sets of obstruents with the terms “fortis” and “lenis” originates with Hattori (1962a: 2; 1962b: 68). However, the fortis obstruents may be more accurately analysed as [+spread glottis], since this feature unites voiceless aspirated plosives and voiceless fricatives. The opposite, [-spread glottis] is the unifying feature of voiceless plosives and voiced fricatives (Shiraishi 2006: 47-49). Although [±spread glottis] may be a more accurate way of describing these consonants, the terms fortis and lenis are commonly used in the historical linguistic literature on the Amuric languages (e.g. CND) and for that reason I have elected to maintain them here as well. Other apparent peculiarities in Table 7 can be explained similarly by taking CA into account. Though /r̥/ and /r/ are phonetically trills, they behave as the corresponding fricatives of /tʰ/ and /t/ respectively. Likewise, /s/ an /z/ are phonetically alveolar fricatives but alternate with the palatal plosives /cʰ/ and /c/ respectively.

In light of their behaviour in CA, the contrast between initial plosives and fricatives is suspicious. Fricative-fricative alternation only exists in verbs that are fricative-initial. Arguably, these verbs do not even start with a fricative synchronically. Given that all fricative-initial verbs are transitive, they are most likely preceded by the third person undergoer prefix **I-*. This prefix is productively deleted, leaving

spirantisation as its only trace (Shiraishi 2000; Jakobson 1971: 88-90). A similar explanation has been given for fricative-initial nouns. These do not alternate in the fricative-fricative context (Shiraishi 2006: 102-103). Jakobson (1971: 93-94) argues that these nouns, especially kinship terms, also contain the prefix **I-*, which also serves as a possessive prefix. Other cases may be the result of the loss of segments as well, such as *ɻə* 'door', attested in Grube (1892) as <-ktö>, where the initial (*)*k*$^{(h)}$ triggered the alternation.[13] In WSN and AN (possibly also NSN), nouns with the initial fricative /v/ are simply the result of the aforementioned change of **w* > /v/ (Jakobson 1971: 93-94; Halm & Slater 2018: 24-25; Gruzdeva 1998: 11; Shiraishi 2006: 28). Jakobson has also claimed that some nouns are simply fricative-initial because they are onomatopoeic, significantly, AN *ral* 'frog'. This same word has also been discussed in Section 6.1. Given that it is attested as *udala(n)* in Uilta, it may be reconstructed as **ụtalVŋ* [14], where **ụ* provides essentially identical conditioning as the prefix *I-*. Austerlitz (1977: 19) gives the same explanation for *ɻə* 'door', by comparing it to Uil. *utə* 'id.' but this leaves (*)*k*$^{(h)}$ as recorded by Grube (1892) unexplained. Regardless of the conditioning of the latter example, it is clear that there were likely no fricative-initial nouns in (Pre-)Proto-Amuric (excepting **h*, though phonetically a fricative, it does not alternate and is often analysed as a glide instead).

With fricative-initial nouns and verbs eliminated, some authors have also gone further to posit that Pre-Proto-Amuric originally lacked a plosive-fricative contrast all together. Austerlitz (1982; 1994: 229; 1990: 19-22) argues that all fricatives are the result of intervocalic spirantisation, although he has also posited that Pre-Proto-Amuric should be reconstructed with either a set of palatal fricatives (**s* and **z*) or a set of palatal plosives (**c*^{h} and **c*). The first assumption features elsewhere in his work as well (cf. Austerlitz 1977; 1983b; 1984a; 1985; 1994) and has also been suggested by Janhunen (2016: 10-11). Yet this hypothesis encounters an obstacle when medial plosives are considered. When it is assumed that all syllables in Pre-Proto-Amuric were CV, then every consonant would have found itself between two

[13] Grube (1892: 62, 76) lists several forms: <ktö, kθy> and also <ršy>. In the latter, initial †*k*$^{(h)}$ has been lost already.

[14] Similar to Halm (2017; 2019) and Halm and Slater (2018; 2020), I transcribe the fortis plosives as voiceless aspirated plosives and the lenis plosives as plain voiceless plosives in my reconstructions of Proto-Amuric and Pre-Proto-Amuric.

vowels in word medial position. Austerlitz (1983b; 1984a) resolves this issue by postulating original geminates, since at least in the South Sakhalin varieties, medial plosives are phonetically long. However, the exact phonetic value is unclear. Austerlitz (1983b: 84) states that he "writes these with stops from the fortis series". Whether this means they surface the same way as fortis plosives (i.e. as aspirated plosives) is ambiguous. Generally, Savel'jeva and Taksami (NRS) and Shiraishi (2006) transcribe only unaspirated medial plosives. Provisionally, I will render these geminates as doubled lenis plosives. According to this reconstruction then, there were three oppositions in plosives, which can be given as **t^h*, **t* and **tt* for the dental place of articulation.

6.2.2 Uvular obstruents in (Pre-)Proto-Amuric

In the previous section, the number of manner contrasts was reduced to only two. It has also been suggested that certain place contrasts are later innovations. The proposals that have been put forward pertain to the opposition between velars vs. uvulars and dentals vs. palatals. These arguments are largely based on postulated co-occurrence restrictions. This subsection focuses on the former, while the next addresses the latter.

It has often been noted that the velars and uvulars are in nearly complementary distribution in the Amuric languages, which had prompted Austerlitz (1990: 20) to not reconstruct this contrast for Pre-Proto-Amuric. Uvulars are clearly correlated with the low vowels /a, o/ and are only found in syllables headed by these vowels (Shiraishi 2006: 25-26; Hattori 1962b: 69; Jakobson 1971: 84-85). According to Jakobson (1971: 84-85) uvulars can also be found after /e/, which strongly suggests that uvularisation is conditioned by low vowels in general. Given that these vowels were possibly specified for RTR in earlier stages of Proto-Amuric (cf. §6.1), this process seems to be analogous to what is found in Tungusic. In these languages, uvular allophones of underlyingly velar consonants are only found before RTR vowels. Effectively, this can be seen as the vowel "sharing" its RTR specification with the preceding back consonant (in the words of Shiraishi & Botma 2017: 204), as [+RTR] is one of the features that differentiates uvulars from velars (Ko et al. 2014: 145-147).

A major obstacle for reconstructing Pre-Proto-Amuric without any uvulars is the fact that velars are not just found with high vowels but also with low vowels. Since such roots can even be reconstructed to Proto-Amuric, Fortescue (2011b: 1362-1363) questioned whether no uvular-velar contrast should be reconstructed. In the later Amuric languages, the velar-uvular opposition was varyingly created and eliminated by changing velars to uvulars and uvulars to velars while also changing low vowels to high vowels, notably **a* > /ə/ (cf. Austerlitz 1984a; Fortescue 2011b: 1362-1363; CND: 4-5; Halm 2017; Halm & Slater 2018). Nevertheless, it is not unthinkable that the contrast is indeed not old. According to Janhunen (2016: 13), the velar-uvular contrast is an indirect result of compounding. Some compounds would have consisted of roots with vowels of two different harmonic pairs. After these vowels were lost, they left only the effect they had on the preceding back consonant. Then words in which low vowels and velars co-occur are the result of compounding that is no longer transparent. Therefore, one could reconstruct Pre-Proto-Amuric without a contrast in dorsal consonants.

6.2.3 Dental and palatal obstruents in (Pre-)Proto-Amuric

Austerlitz (1982: 83) has suggested that the dental-palatal distinction in plosives is not old. In a later publication (cf. Austerlitz 1990: 19-21) he was more undecided but did regard this hypothesis as well-founded. There are some apparent examples of palatalisation. For instance, the correspondence AN *vəc* ‘iron’ = SSN *wat* ‘id.’ was explained as palatalisation by Austerlitz (1984: 42). He assumes that **t* (more specifically **tt*) was originally followed by **i*. A similar explanation could be given for the indicative/nominaliser suffix **-nt* > AN *-ɟ*, ESN *-nd* (cf. CND: 5).

Another case of palatalisation may be presented by AN /z/, which often corresponds to ESN /r/ (CND: 5). Note that, phonologically, /z/ is a palatal fricative and /r/ a dental fricative. Based on that, one could hypothesise that AN /z/ descends from **r* in a similar manner as AN /c/ in the previous examples. Yet there are also cases of AN /r/ corresponding to ESN /z/. Particularly complicated is the relationship of the verb *xiz-* ‘to use’ and the instrumental suffix *-kʰis* proposed by Austerlitz (1982: 86; 1984b: 232). His proposal was based on SSN but if one takes into account their cognates in AN and ESN, the derivation is less obvious. The final consonant in the

cognates of *xiz-* are reflected as /r/ in both AN and ESN (CND: 86). Confusingly, the instrumental suffix is attested as *-kir* in AN, but as *-kis* in ESN (CND: 168). This amounts to the following correspondences: AN /z/ = ESN /r/ = SSN /r/, AN /r/ = ESN /z/ = SSN /z/, AN /r/ = ESN /r/ =SSN /z/. These correspondences can hardly be called regular and as such are not counterevidence to an original dental-palatal contrast. Therefore, Austerlitz's (1982: 83) hypothesis cannot be accepted on the basis of current research.

The status of Proto-Amuric dental and palatal fricatives specifically has also been a matter of discussion. Recently, Janhunen (2016) has argued that Pre-Proto-Amuric both had **r* and **s* relying on Tungusic data even in a fricativeless reconstruction of Pre-Proto-Amuric, though both his examples are problematic. His example for the former is AN *oroo̯r* 'ginseng', from PTg. **o̤ro̤o̤-g-ta* 'grass, hay, plant'. In this example, PTg. **r* is preserved as /r/, while PTg. **t* was apparently changed to /r̥/. An issue with this example is the fact that AN /r/ is found in an environment where it could have descended from PPAm. **t* and could have been adapted as such. Therefore, **r* does not necessarily have to be reconstructed. His example for PPAm. **s* is AN *cʰam,* ESN *cʰamŋ* 'shaman', which he compares to PTg. **samaan* 'id.'. This is a word for which the direction of borrowing is unclear and therefore it is also unclear how to understand the correspondence Am. (*)*cʰ* = PTg. **s* in terms of PA. If it is originally Amuric, then it seems unlikely that Proto-Tungusic would have categorised PPAm. **cʰ* as **s* instead of **č*, while the opposite direction is completely regular (see Section 7 for PA as it applies to Tg. (*)*s*). Since both *oroo̯r* and *cʰam(ŋ)* are likely candidates for ancient borrowing, I discuss these examples further in Appendix D. At present, despite these issues, I tentatively accept Janhunen's hypothesis, since it could ultimately provide explanation to the irregular correspondences of /r/ to /z/ as well, similar to how PAm. **u̯* may explain irregular correspondences between AN and ESN as was discussed in Section 6.1.

6.2.4 The Pre-Proto-Amuric consonant inventory summarised

Although the most conservative reconstruction of the consonant inventory of Proto-Amuric is clearly the one closest to Table 7, there seem to be good reasons to reconstruct it for Pre-Proto-Amuric as in Table 8.

	Labial	Dental	Palatal	Velar	Glottal
Fortis plosives	*pʰ	*tʰ	*cʰ	*kʰ	
Lenis plosives	*p	*t	*c	*k	
Fricative		*s			*h[15]
Nasal	*m	*n	*ɲ	*ŋ	
Liquid		*r, *l			
Glide	*w		*j		

Table 8: Pre-Proto-Amuric consonant phonemes revised

This inventory is indeed quite close to the one reconstructed for Proto-Tungusic in Table 5. Yet for this comparison it has to be assumed that the fortis-lenis contrast is not distinct from the voicing contrast in Tungusic, even though these do not represent the same mechanism. Additionally, it has to be assumed that PPAm. **h* is comparable to PTg. **x*, even though the former is glottal, while the latter is velar. The relevance of these differences receives further attention in Section 7. What remains to be compared are the construction of larger units out of the segments that Pre-Proto-Amuric and Proto-Tungusic possessed.

6.3 (Pre-)Proto-Amuric and Proto-Tungusic syllable and root structure

All Tungusic languages follow a relatively simple syllable structure of (C)V(C). The nucleus of a syllable is a vowel, which may be long, short or a diphthong. There are generally no clusters in the coda or onset of syllables with the exception of some later developments in individual Tungusic languages. An initial syllable begins with either a consonant or a vowel, while non-initial syllables must always start with a consonant (Li 1996: 19-20). A survey of the roots reconstructed by Benzing (1956) reveals that roots are generally polysyllabic, having the maximal structure (C)V(C)CV(C)[16], while monosyllabic roots of the type (C)V(C) exist as well. In contrast, Amuric syllables and roots are much more complex. In all Amuric languages, the maximal possible structure of a syllable is (C)(C)V(C)(C)(C). Roots are usually monosyllabic, though disyllabic roots are attested as well. These have the

[15] As mentioned earlier, this segment is generally analysed as a glide in later Amuric languages but has been placed in the row of fricative to stress the parallels between the Pre-Proto-Amuric and Proto-Tungusic consonant inventories. For the same reason, geminate consonants have been left out.
[16] There are also examples of longer roots in Appendix D. These have to be considered individually.

maximal structure (C)V(C)CV(C) (Shiraishi 2009b: 86-87). A survey of the CND also shows that this structure can be reconstructed for Proto-Amuric on the basis of the Comparative Method but if Austerlitz's (1994; 1990: 21-22) hypothesis relying on internal reconstruction is to be accepted, syllables and roots may have been less complicated in the more distant past.

As I have discussed in Section 6.2.1, Austerlitz (1994; 1990: 21-22) derives medial fricatives from plosives by assuming that they originally occupied a position that was intervocalic. As a result, many roots would have to be reconstructed as polysyllabic for Pre-Proto-Amuric, constructed of CV-syllables. Yet as Shiraishi (2009a: 196) points out, even though this reconstruction is based on internal reconstruction, there is very little internal evidence for Austerlitz's hypothesis. The main example seems to be **ci-ka-tV* 'tree' (Austerlitz 1983: 78, 1986a: 188; 1994: 229). This root does exhibit some interdialectal variation (see also CND: 52), which could be indicative of an originally polysyllabic structure. Nevertheless, with few such examples, there is little evidence within Amuric for an original (C)V structure.

External comparison on the basis of Tungusic loans does however provide some evidence of polysyllabic roots and medial spirantisation (cf. Austerlitz 1985; Janhunen 2016). Janhunen's (2016) reconstructions also show that Pre-Proto-Amuric most likely had roots of the same (C)V(C)CV(C) structure that is still attested in modern Amuric languages, though it must have been more wide-spread than is currently the case. This is exactly the same structure that can also be reconstructed for Proto-Tungusic.

7 Phonological adaptation of Tungusic loanwords in Amuric

Section 6 has provided sufficient information on the phonological makeup of both (Proto-)Tungusic and (Pre-Proto-)Amuric to interpret and reinterpret instances of PA in these languages from a diachronic perspective. Since this overview relies on the findings of Krejnovič (1955a), Austerlitz (1956), Jakobson (1971), Janhunen (2016) and Shiraishi and Botma (2016), my treatment leans more to a discussion of the adaptation of Tungusic loans in the Amuric languages. These publications have contributed numerous examples that show how individual segments as well as longer

strings of vowels and consonants are accommodated to the phonological constraints shared by the languages in the Amuric clade.

Starting with larger units, there are the processes affecting syllable structure. Krejnovič (1955a: 141) and Janhunen (2016: 11-16) have shown that vowels in the initial, medial and final syllables of Tungusic loanwords in Amuric are often deleted. This frequently has as a result the formation of clusters and reduces the number of syllables that the word consists of to just one or two. Janhunen has argued that the loss of vowels is a diachronic process. Since polysyllabic roots are not illicit in Amuric, as discussed in Section 6.3, this is a more likely scenario than PA.

If vowels are still reflected in the initial syllable then they are retained without alteration according to Krejnovič's findings (1955a: 136-137). Table 9 lists his vowel correspondences, coupled with those discussed by Janhunen (2016).

Tungusic	Amuric
(*)*i*	(*)*i*
(*)*į*	(*)*e*
(*)*ə*	(*)*ə*
(*)*a*	(*)*a*
(*)*o*	(*)*o* or (*)*u*
(*)*ọ*	(*)*o*
(*)*u*	(*)*u*
(*)*ụ*	Either (*)*u* or (*)*o*
(*)*ịa* in the first syllable, (*)*a* preceded by a palatal consonant	(*)*e*

Table 9: Tungusic-Amuric vowel correspondences (adapted from Krejnovič 1955a: 136-137; Janhunen 2016)

Firstly, it should be noted that these correspondences are not regular (e.g. Tg. (*)*ụ* = Am. (*)*u*, (*)*o*) nor are the vowels identical (e.g. Tg. (*)*į* = Am. (*)*e*). In part, these irregularities may be explained by Shiraishi and Botma's (2016) findings. They found that the Amuric restriction on the co-occurrence of high and low vowels is also maintained in loanwords from Tungusic and Ainu. As a result, one of the vowels in a

disyllabic word changes to either a high or low vowel depending on the height of the other vowel in the word. An alternative explanation to these findings may be that Tungusic RTR vowels are simply categorised as low vowels in Amuric. This is contradicted by the correspondences of Tg. **ụ* and **o*. Yet this can possibly be resolved by postulating that Amuric had additional vowels that later merged with other vowels to become /u/ or /o/, as I have done in Section 6.1. Although from this same perspective, the correspondents of Tungusic velars are difficult to explain.

Both Austerlitz (1956: 262) and Jakobson (1971: 84) list a few instances of Tungusic loanwords in Amuric that have a velar followed or preceded by a low vowel. As these low vowels are the equivalents of Tungusic RTR vowels, this is contrary to expectation, since Tungusic velars are realised as uvulars in this environment (cf. §6.2.2). This is all the more confusing, since these uvular allophones are generally represented in the examples listed by Krejnovič (1955a). Possibly, this irregularity is reflective of the fact that uvulars and velars are non-contrastive in Tungusic. As a consequence, Amuric may have borrowed one allophonic variant.

A similarly irregular correspondence pertains to the reflexes of Tungusic fricatives in Amuric. As discussed by Krejnovič (1937: 52-54; 1955a: 138-141), initial fricatives of Tungusic loanwords are adapted varyingly as fricatives or plosives. This is possibly also reflective of a diachronic development albeit not necessarily the result of one. Krejnovič has suggested that these words are from different periods of borrowing. I agree with this analysis. As I have discussed in Section 6.2.1, there were likely no roots that started with a fricative in an earlier period of the lineage. Before this stage, initial **s* must have merged with **c^h* already, if this segment is to be reconstructed (cf. §6.2.3), but afterwards there certainly were no fricative-initial roots. When these were (re)introduced, the constraint against initial fricatives disappeared. Therefore these irregular correspondences suggest that they are loanwords from different eras: the plosive-initial ones were borrowed before this change, while the fricative-initial ones were borrowed after it.

A complicating factor is the treatment of Tg. (*)*x*. In those words where it is reflected as a fricative, it is either a dorsal fricative or (*)*h*, the segment that is

phonologically not a fricative. Krejnovič (1955a: 138) refrained from drawing any further conclusions and instead suggested that further research was required. It is indeed the case that such irregularities exist in his dataset, though those items that ended up in my dataset (see Section 8.2 for the criteria I have used) only exhibited the correspondence Tg. (*)*x*- = Am. (*)*h*-. Therefore, it is presently unnecessary to consider the issue further.

The least controversial correspondence pertains to the treatment of laryngeal features. Krejnovič (1955a: 137-138) has shown that Tungusic voiceless plosives are generally categorised as Amuric fortis plosives, while Tungusic voiced plosives are categorised as lenis plosives. This is however only established for the initial position and the correspondences in non-initial position are still understudied. An irregularity that can be pointed out (further examples are listed in Appendix D), is the fact that Am. (*)*keŋ* 'whale of intermediate size' corresponds to Ulc. *qeŋa* 'whale', where the lenis plosive (*)*k* corresponds to a voiceless plosive. This is however most likely a borrowing from Amuric, considering that terms for marine animals in the Tungusic languages are generally of Amuric origin (cf. Doerfer 1985b). Since this word is also attested in other Tungusic languages with a voiceless plosive (cf. SSTM1: 388), it would seem that Amuric lenis plosives were interpreted as voiceless plosives (probably on the basis of their phonetic realisations). Although more examples are needed, this provides reason to question whether Krejnovič's correspondence works both ways. Additionally, it puts Janhunen's (2016; 1981: 131) and Gruzdeva and Janhunen's (2020) assertions into question on the similarity of the phonologies of Proto-Tungusic and Pre-Proto-Amuric as pertains to laryngeal features. Yet only insofar as this word can be reconstructed to Proto-Tungusic, to which I return in Appendix D.

8 Data and methodology

8.1 Sources and transcription

The dataset I have used consists of a list of words that I extracted from all previously published research on Amuric-Tungusic lexical contact, i.e. Grube (1892), von Möllendorff (1894), Laufer (1902), Krejnovič (1937; 1955a), Austerlitz (1956; 1977;

1983b; 1983c; 1984a; 1984b; 1985; 1986b), Bouda (1960), Jakobson (1971), Panfilov (1973), Cincius (1975; 1977), Doerfer (1985a; 1958b), Janhunen (1986; 1998; 2005; 2008; 2011; 2014; 2016), Burykin (1987; 1988; 1989; 2005; 2012), Pevnov (1992; 2016; 2018), Blažek (1997), Shternberg (1999), Abondolo (2005), Burykin and Popov (2012), Robbeets (2015), Tangiku (2015; n. d.), Shiraishi and Botma (2016) and Fortescue (CND). In addition to these sources, I have added items from a list that I have compiled myself while going through the dictionaries of Savel'jeva and Taksami (NRS) and Grube (1892). Of this list, I have included only those items not included in any of my other sources. For Grube (1892), this means that I have included only those words for which he himself had not proposed a Tungusic parallel already. In the initial stage of this research, this list was compiled using Microsoft Excel, creating a rough database that eased searchability. Afterwards, I had started compiling this data in the format of a comparative dictionary with commentary, which can be found in Appendix D.

Since most sources do not list intra-Tungusic or intra-Amuric cognates exhaustively, I have tried to match every proposed item to their closest equivalents in lexicographic sources. For Amuric, Fortescue (CND) was my primary source. Whenever necessary, I have further completed his data by consulting Krejnovič (1955b), Hattori (1962a; 1962b; 1962c), Savel'jeva and Taksami (RNS; NRS), Shiraishi (2006), Tangiku et al. (2008) and Sangi and Gašilova (NRS-S), which taken altogether give an almost complete representation of both the Nivkh and the Nighvng branch. For "Old Amuric" (i.e. the earliest attested form of Amuric), I have consulted Grube (1892). The listed works employ various Cyrillic and Latin-based transcription schemes to represent Amuric, which I have converted into a unified scheme based on the International Phonetic Alphabet (hereafter IPA). Grube (1892) is however the exception, whose transcription I have left unaltered, since it is generally not very precise and therefore it would need too much interpretation to convert to IPA. Whenever I provide an interpretation however, it is preceded by a dagger (†).

For the Tungusic languages, I have relied on Cincius's (SSTM1; SSTM2) comparative dictionary, which is still the most complete compilation of lexicographic material on this family. Except for Jurchen, which is written in a Latin-based

transcription, she employs a Cyrillic-based transcription to represent the Tungusic languages. I have converted these characters into Latin characters following a scheme largely derived from the romanisation schemes found in Robbeets and Savelyev (2020: lv-lix) and the IPA. This scheme is given in Appendix A.

In sum, my primary means of transliteration is the IPA. Janhunen (2016: 8-9) has argued (in a footnote) against the use of a universal phonetic transcription like the IPA in Amuric-Tungusic comparisons, "since it would tend to obscure the phonemically relevant features behind a curtain of irrelevant phonetic details." However, as previous research has shown (cf. Kang 2011; Boersma & Hamann 2009), phonetic information is at least of influence in the PA of loanwords, to which I have also alluded in Section 7. Therefore, I argue that leaving out phonetic details would obscure this part of the borrowing process instead. For that reason, I have used an IPA-based transcription scheme in and throughout this thesis.

Since morphological segmentation plays a part in my analysis, as I discuss further in Section 8.3, I have also consulted reference grammars and historical grammars. I have relied on Benzing (1956), Sunik (1982) and Boldyrev (1987) for analyses of Tungusic morphology from a historical comparative perspective. For Amuric, Fortescue (CND) is the only source written from this perspective. His work employs only the Comparative Method, which suffices for Tungusic but as I have noted previously (cf. §6) this reaches only the relatively more recent Proto-Amuric than the more ancient Pre-Proto-Amuric. Consequently, his treatment of morphology can be supplemented by synchronic grammars on the Amuric languages. For this purpose, I have consulted the publications of Gruzdeva (1998), Mattissen (2003) and Nedjalkov and Otaina (2013). As the scope of my thesis does not permit an in-depth discussion of Amuric and Tungusic historical morphology, I have opted to give only a brief compilation of the affixes that I have identified in my dataset. These can be found in Appendix B and Appendix C.

8.2 Delimiting the dataset

This study only considers a selection of the data found in the sources listed above. After an initial review, it became clear to me that some of the items proposed in these

publications were not ancient loans or not borrowings at all. Such items are irrelevant to the questions at the centre of this research and therefore had to be eliminated from the dataset. In this section, I explain the criteria on which the selection is based as well as my reasoning behind them. I have considered the prerequisites of reconstruction and the specific requirement of my research: items must demonstrably be of common origin and should likely be old.

As I have argued in Section 1, the period when both Pre-Proto-Amuric and Proto-Tungusic existed, is the earliest possible point on the timeline of Amuric-Tungusic contact. Knowledge of these languages can only be arrived at through reconstruction. The standard method of reconstruction in historical linguistics is the Comparative Method. There are also other methods, which I have discussed in Section 6 but, as should also be clear from that same section, these rely on a more complex analysis, which would be overkill in the data selection stage. For that reason, I have only considered the prerequisites of the Comparative Method, though it should not be disregarded then that through this method only Proto-Amuric can be reconstructed, which is significantly more shallow than Pre-Proto-Amuric, as can be arrived at through other methods. A seven step outline of this method is given by Millar and Trask (2015: 196), which has been cited below in full.

1. We first decide by inspection that certain languages are probably genetically related and hence descended from a common ancestor.
2. We place side by side a number of words with similar meanings from the languages we have decided to compare.
3. We examine these for what appear to be systematic correspondences.
4. We draw up tables of the systematic correspondences we find.
5. For each correspondence we find, we posit a plausible-looking sound in the ancestral language, one that could reasonably have developed into the sounds that are found in the several daughter languages, bearing in mind what we know about phonological change.
6. For each word surviving in the various daughters, we look at the results of 5 and thus determine what the form of that word must have been in the ancestral language.

7. Finally, we look at the results of 5 and 6 to find out what system of sounds the ancestral language apparently had and what the rules were for combining these sounds.

The first four of these specifically apply to data selection and provide the prerequisites of reconstruction. Since I do not question the relatedness of the Tungusic languages nor the relatedness of the Amuric languages, the first requirement is to be regarded as fulfilled. The second requires semantic similarity between items and the third and fourth, essentially overlapping steps, require evidence of regular sound correspondences. Semantic similarity can be, admittedly, subjective. Yet if the meaning is neither identical nor close, the relation between two words is questionable. Comparatively, regular sound correspondences are more easily obtained through systematic analysis and are therefore less easily disputed. If there are no regular sound correspondences between words, then the relation is also easily dismissed. Yet both regular sound correspondences and semantic similarity are necessary to prove the relationship of two words. Therefore, I have focused on the following two criteria to come to an initial selection.

I. Proposed cognates of the item are semantically similar enough to be regarded as descendants from the same word.
II. Proposed cognates of an item exhibit regular sound correspondences upon comparison.

A shortcoming of these criteria and by extension the Comparative Method is that it provides no means of differentiating loanwords and cognates. As pointed out in Section 5, both inheritance and borrowing result in words that are formally (and semantically) similar across two or more languages. If all languages of a family have carried it with them from the time of their initial unity, then that is overwhelming evidence that it must be reconstructed to their common ancestor, even if it is a loan. This is precisely the shortcoming that was exploited in this phase of the selection process, since it allows for both the Tungusic member and Amuric member of a parallel to be reconstructed to their respective protolanguages and then allows for further comparison within the framework of this thesis.

The Comparative Method as described by Millar and Trask (2015: 196) does not necessarily require a word to be widely attested in a language family. However limited attestation may be an indicator of borrowing, which I discuss further in Section 8.3. If a word is only attested in a limited number of languages within the family and there is an obvious parallel to it in another related language, then it was likely not inherited from the protolanguage. These loanwords are then not the type of ancient borrowings that I have attempted to select. Thus, I have had to determine what words were likely not carried over from the families' respective forebears. For that, I have relied on the available literature.

Doerfer (1985a: 254-261) has shown that some words that are widely attested in Tungusic are not inherited from Proto-Tungusic but borrowed from another Tungusic language, specifically Jurchenic. The branches affected are Orochic and Nanaic but it also possible to find these borrowings in Ewenic languages, such as Solon and eastern and southern varieties of Ewenki (in Doerfer's classification of its varieties). Among those languages surveyed by Doerfer, only western dialects of Ewenki and Ewen are not affected by borrowing from Jurchenic. This provides a basis to argue that if a word is attested in Ewen and western dialects of Ewenki (in Doerfer's classification) as well as any of these Tungusic languages affected by Jurchenic or belonging to the Jurchenic branch, then it must have been inherited from Proto-Tungusic if it also fulfils criteria (I) and (II). A similar argument can also be made on the Amuric side. According to Gruzdeva (1998: 7), the Nivkh branch of Amuric came about largely due to contact with Tungusic languages. It is likely that this is indeed the case and therefore any word not attested in the Nighvng branch may be a representative of this later interaction with Tungusic instead. Therefore, these branches must be treated as if they constitute separate languages. Only words found in both the Nivkh and Nighvng branches are likely to be inherited from Proto-Amuric. From these considerations, I have arrived at a third criterion, which is given below.

III. Within Tungusic, the item is attested in western Ewenki and/or Ewen and any other Tungusic language. Within Amuric, it is attested in both the Nivkh and Nighvng branches.

If all these prerequisites are met, then it is possible to provide a reconstruction in accordance with the three remaining steps in Millar and Trask's (2015: 196) description. Their outcome is a convincing reconstruction of the particular item for both Proto-Tungusic and Proto-Amuric. However, this does not necessarily indicate that the proposed parallel was the result of borrowing. For that, it should also exhibit the structural closeness of a loanword to the word that is supposedly non-borrowed. This leads me to the fourth criterion, which is given below.

IV. The reconstructions of the item for (Pre-)Proto-Amuric and Proto-Tungusic have an identical or nearly identical phonological makeup.

In Section 5, I have relied on the work by Nichols (1996) to show how relatively long strings of segments shared between two languages are a strong indicator that both of these languages received it from a single source. This source may be their shared forebear, although one of the two could also have borrowed it from the other. In that case the ultimate source of the word is the ancestor of the donor language.

Methodologically, the question is what strings or sequences of segments should be compared to each other. Nichols (1996: 50-54) has shown how identical sequences are a stronger indicator of a single origin than sequences that are merely similar. This however depends on how this similarity is defined. A broad category of labials has a higher probability of occurrence than a more narrow category of dental plosives. For the latter, the difference between /t/ and /d/ is only in the feature [±voice] and can therefore be regarded as nearly identical. This is the reason why I have looked for only those words that are made up of either fully identical or nearly identical segments.

The segments that fall within the category of nearly identical can be grouped on the basis of differentiation for a single feature. As per the previous example, /t/ and /d/ are all but identical except for the feature [±voice]. It should be acknowledged however that all consonants fall within a broader phonological system. Cincius (SSTM2: 988) compares Amuric words with initial (*)*ŋ* to Tungusic words beginning with (*)*n* or (*)*ɲ* or (*)*ŋ*. All of these Tungusic nasals that are not (*)*ŋ* are only different from Am. (*)*ŋ* for one feature. Yet all these comparisons are

problematic: all taken together, Cincius basically posits that Am. (*)*ŋ* corresponds to any non-labial nasal in Tungusic, which makes the class more generic and thereby increases the probability of this being a chance occurrence. Additionally, these comparisons do not seem sensible from the perspective of PA.

To reiterate the point made in Section 5, in PA, foreign structures that violate the language's native constraints are repaired using a variety of strategies. In segmental adaptation, that often means that some features of a consonant are retained while others are sacrificed (Kang 2011: 2265). However, this necessarily requires that there are structures that violate the native phonology. In the previous example there are no such illicit structures. In neither Tungusic nor Amuric (*)*ŋ* is a consonant that violates any phonological constraints, since both must have had this consonant and did also allow it in initial position (cf. Benzing 1956; CND). That is not to say that it is impossible that there may not have been constraints on combining (*)*ŋ* with certain other segments, which would have triggered adaptation, yet this is already conjecture.

Closeness of segmental makeup is then best defined on the basis of a comparison of (Pre-)Proto-Amuric and Proto-Tungusic as well as the knowledge available on PA in cases of Amuric-Tungusic borrowing. In Section 6, I have shown that Pre-Proto-Amuric and Proto-Tungusic had nearly identical phoneme inventories. I have also shown in Section 7 that there is a great degree of featural preservation in borrowing as well as a clear systematicity to the adaptations. It seems therefore safe to assume that Pre-Proto-Amuric and Proto-Tungusic would have had little issue handling loanwords from each other, while this may have been more problematic at a later time. Taking into account those irregular sound correspondences based on height, RTR, voice and strength (i.e. fortis or lenis) that I have discussed, it seems safe to require of comparanda that their reconstructions for Pre-Proto-Amuric and Proto-Tungusic were all but identical to each other, accepting some leniency for those correspondences that have been found to be not fully regular.

As I have based my selection on Proto-Amuric, which is significantly younger than Proto-Tungusic due to the limitations of the Comparative Method when applied to

the Amuric language family, this requirement was more involved in practice. In its most strict application, I would have needed to eliminate nearly all items, since they were already quite different. Therefore, I had decided to consider those words that were "potentially identical" in structure. My shallow reconstructions potentially reflected those developments that followed the Pre-Proto-Amuric stage, those developments which have been retrieved using other methods (cf. §6, §7). If, after "reversal" of these developments (i.e. reconstruction), the Pre-Proto-Amuric etyma were identical or nearly identical in shape to the Proto-Tungusic etyma, then they were accepted.

8.3 Direction of borrowing

Usually, the process of borrowing is conceptualised as having a direction: a word is borrowed *from* one language *into* the other. The language that borrows is the "recipient language", while the language from which is borrowed is the "donor language". Establishing a direction of borrowing is essentially the preeminent proof of borrowing itself, since it shows how the word was transmitted. In this section, I lay out the methodology I have employed to complete this task.

I have formulated criteria to determine what language is the donor language and what language is the recipient language. These criteria have been derived from the work by Clauson (1959: 175-178), Rozycki (1994: 4-5), Haspelmath (2009: 45) and Campbell (2013: 62-66). Of their various criteria, some have shortcomings. Clauson's (1959: 175-178) first and fifth criterion (or method) rely heavily on an extensive literary tradition, which is not available for the Amuric and Tungusic languages (excepting the Jurchenic branch). Though all authors make use of phonology as a basis for their reasoning, they often do not make a distinction between diachronic phonology (sound change) and synchronic phonology (PA). Application of the latter as a criterion seems to be underdeveloped as well. Haspelmath (2009: 45) simply states that "[i]f a word shows signs of phonological integration in language A but not in language B, it must come from language B", but does not specify it further. Those sources that make mention of some repair strategies – such as those discussed by Kang (2011: 2265-2267, 2269-2273) – also do not make a clear distinction between segmental adaptation and phonotactic adaptation (i.e.

deletion and epenthesis) even though they entail fundamentally different repair strategies.[17] By adding such a distinction to the other available and useable criteria, I have arrived at nine criteria in total. I elaborate on each of these criteria individually in the next paragraphs. Afterwards follows a short evaluation.

1) *morphological segmentability.* Sometimes a borrowed word is a complete unsegmented block of morphemes that is inserted into the morphosyntactic frame of another language. If the morphological makeup of the lexeme is transparent in language A, but not in language B, then language A is the donor language and language B is the recipient language. This criterion is based on Haspelmath (2009: 45), Campbell (2013: 63-64), Rozycki (1994: 5) and Clauson (1959: 176-177).

2) *diachronic developments.* This criterion is based on Campbell (2013: 62). He describes how the distribution of phonemes in the synchronic phonology of a language is the outcome of different sound changes. To illustrate: segment X could occur in language A in an environment where it should not be found – based on what is known about this language's development. In that same environment, it is allowed in language B. This then suggests that B was the donor language and A was the recipient language.

3) *differential importation.* Differential importation refers to a kind of PA when there is, paradoxically no alteration at all. Often unexpectedly so, if the structure is disallowed in native words. In such a case, the phonology apparently allows exceptions for loanwords (Kang 2011: 2260). As a criterion, it was described by Campbell (2013: 62) and more or less the same is also described by Clauson (1959: 177). It can be applied as follows. A word in language A has a structure that would normally violate the phonology of language A. This same structure is not illegal in B. Then B must be the donor language, while A is the recipient language.

4) *Epenthesis.* Epenthesis is one of the possible repair strategies described by Kang (2011: 2269-2273). Epenthesis refers to the insertion of a segment to amend a violating structure. None of the sources I have consulted have used specifically

[17] Though it should be noted that Boersma and Hamann (2009) argue that epenthesis is actually not distinct from segmental adaptation in their adaptation-as-perception model.

epenthesis as a criterion, therefore this criterion is of my own design. It can be assumed that a language that allows more complex sequences of segments (e.g. consonant clusters) does not epenthesise easily. A may be such a language while B is not. Then, if language B seems to have inserted a segment where A, for example, has a cluster, the structure in A is probably original while in B it is not. That makes A the donor language and B the recipient language.

5) *Deletion.* Deletion is one of the possible repair strategies described by Kang (2011: 2269-2273). Deletion refers to the removal of a segment to amend a violating structure. As a criterion, it has been employed by Rozycki (1994: 5). In the case of deletion, it can be assumed that the language that is less restrictive in its distribution of consonants is less likely to delete any in loanwords. Therefore, if language B is this language and language A shows deletion, while B does not, then A is the recipient language and B is the donor language.

6) *Segmental adaptation.* The process of segmental adaptation has already been discussed in Section 5 and Section 8.2. Employed as a criterion, it is the process of comparing the sets of phonemes of two languages, identifying the contrasting features and establishing which are most likely to be lost when borrowed. If language A lacks a certain contrast that is present in language B and A apparently merged the two contrasts into one, while language B did not, then the word was borrowed from B into A. None of my sources have developed segmental adaptation as a criterion, although Janhunen (2016: 18) seems to follow a similar reasoning as presented here.

7) *Extent of attestation*. Both Campbell (2013: 64-65) and Haspelmath (2009: 45) describe this criterion. If a given word is attested in language A and not anywhere else in the family A belongs to, then it was borrowed from B if the word is found throughout the language family that language B belongs to. Assumed here is that an item that is pervasive in a family is likely inherited from its ancestor. If it has a more limited distribution, then it cannot be reconstructed for that ancestor. In that case, it is more likely to have been borrowed.

8) *Semantic narrowing.* This criterion originates from Rozycki (1994: 4-5). In his description, semantic narrowing is the phenomenon that a loanword has a more limited semantic range as compared to its original in the donor language. So, if in language A the word under study has a more limited meaning than it has in language B, it was borrowed from B into A.

9) *Extra-linguistic considerations.* Various authors suggest taking historical context (cf. Haspelmath 2009: 45), geographical and ecological associations (cf. Campbell 2013: 65) and social and economic background (cf. Clauson 1959: 178-187) into account. When a people comes into contact with another people that has a different society or has different economic activities than their own, they may require a set of new words in order to be able to refer to these foreign concepts. As such, the primary motivation for borrowing is mainly the acquisition of new knowledge for which words are necessary. In this case, the language of a people more strongly associated with a given context is most likely the donor language of any of the loanwords that can be found. I have presented a similar argument in Section 5 while discussing semantic domains.

These criteria present a general framework for studying direction of borrowing. The only criterion that appears dubious is (8), as it is debatable that borrowing necessarily requires semantic narrowing. The applicability of the others depends on the situation that is under study. Considering that Amuric syllable structure is generally more complex than Tungusic syllable structure (cf. §6.3), it is more likely that Tungusic would epenthesise in Amuric loanwords. Yet this complex syllable structure may be the result of the diachronic process of deletion that occurred in Amuric, a process that also affected loanwords from Tungusic (cf. §7). PA in Tungusic and sound change in Amuric working together completely obscure what segment is original or the product of epenthesis, rendering criteria (4) and (5) unusable. Criterion (6) runs into a similar issue. Only lenis plosives corresponding to voiceless plosives in Tungusic seem to suggest an Amuric origin, while their phonologies potentially were indistinguishable in their Proto-Tungusic and Pre-Proto-Amuric stages (cf. §6, §7). At the same time, comparatively little is known of Amuric linguistic history, which impedes the application of criterion (2). Criterion (7) is largely defeated by the methodological choice of including only widely attested items (cf. §8.2). In

conclusion, the least problematic seem to be criteria (1), (3) and (9) and for that reason they form the backbone of most of my argumentation.

8.4 Stratigraphy of borrowing

The term "stratigraphy" as applied to borrowing is adopted from Kallio (2015) and refers to the main task of this thesis. It describes the process of distinguishing different layers of loanwords. Both Kallio (2015) and Rozycki (1994: 227-230) employ similar reasoning to this end. As described in Section 5, loanwords participate only in those sound changes that occur *after* they enter into a certain lineage. This creates different sets of words that reflect different innovations. On the basis of these innovations, the sets can be assigned to different periods within a language's development. On this timeline, the oldest layer is the one that shows evidence of the innovations that must have occurred the earliest. Those that show later innovation then belong to subsequent stages of the languages involved.

The purpose of stratigraphy might seem to be defeated by the selection criteria outlined in Section 8.2. If all the items I have selected can be reconstructed to the respective ancestors of the two language families then there should be no innovations that would give away what stratum they belong to. However, Proto-Amuric and Proto-Tungusic as reconstructed by employing the Comparative Method cannot have existed concurrently, as I have stressed previously (cf. §6, §8.2). The more shallow of the two is Proto-Amuric. Therefore, my dataset needs to be examined for changes that must have occurred between Proto-Amuric and its ancestor Pre-Proto-Amuric. In previous research, such sound laws have been proposed through internal and external reconstruction (cf. §6, §7). As I have pointed out, some of these proposals still require supporting evidence, yet no other accounts of earlier stages of Amuric are currently available in the literature. Consequently, I could only rely on those that are available.

9 Results and analysis

9.1 Final dataset and direction of borrowing

In dictionary format, Appendix D lists all 50 items that passed the criteria that I established in Section 8.2. I have subjected each of them to a rigorous analysis, which I present in every item's respective entry. Based on my conclusions, some had to be discarded as they clearly did not belong to an ancient stratum of contact. Additionally, I have also attempted to establish a direction of borrowing for each item, using the criteria from 8.3. These findings are summarised in Table 10. The entry number is listed first, followed by reconstructions in Proto-Amuric and Proto-Tungusic (in that order). Next I provide the direction of borrowing (DoB), indicated by an arrow (→). The family that the donor language belonged to or was the ancestor to is given as either Tungusic (Tg.) or Amuric (Am.). If my argumentation incorporates a specific criterion from Section 8.3, I provide its number. If no number is given, I used other arguments. My criteria are repeated below.

1) morphological segmentability.
2) diachronic developments.
3) differential importation.
4) Epenthesis.
5) Deletion.
6) Segmental adaptation.
7) Extent of attestation
8) Semantic narrowing
9) Extra-linguistic considerations

Generally, the direction of borrowing I have posited is provisional and a more nuanced analysis is given in Appendix D. If the direction of borrowing is unclear, it is indicated with *U* and if the item turned out to be not a borrowing from either language, that is indicated with *N/A* or "non-applicable". The last column gives whether I accepted (A) the parallel as potentially old or dismissed (D) it. Those among the accepted are further classified into strata in the next section.

No.	Proto-Amuric	Proto-Tungusic	DoB	A/D
0001	**amᵩra-* ‘to taste’	**amta(-)* ‘taste’	Tg. → Am. (1)	A
0002	**aɲ(i)* ‘year’	**aɲŋa-* ‘id.’	Tg. → Am.	D
0003	**ara(-)* ‘almost’	**aran* ‘barely’	U	A
0004	**arak* ‘liquor’	**arakii* ‘id.’	Tg. → Am. (9)	D
0005	**aχma(-)l-* ‘to be a father-in-law’	**ama-kAA* ‘grandfather, father-in-law’	Tg. → Am. (1)	A
0006	**capŋ* ‘pincers’	**ʤapa-pOOn* ‘grab-NML:INS’	Tg. → Am. (1)	A
0007	**cʰai* ‘tea’	**čaị* ‘id.’	U	D
0008	**cʰamŋ* ‘shaman, eagle’	**samaan* ‘shaman’	Am. → Tg. (1)	A
0009	**cʰi* ‘you (sg.)’	**sii* ‘you (sg.)’	N/A	D
0010	**cʰoxt(ụ)-* ‘to get drunk’	**sọktọọ-* ‘id.’	U	A
0011	**aka-* ‘to be older’	**akaa* ‘elder brother’	Tg. → Am. (6)	A
0012	**əlv-* ‘to make a roof’	**əlbə-* ‘to cover’	U	A
0013	**amɣ~*amx* ‘mouth’	**amŋa* ‘id.’	U	A
0014	**ərŋ* ‘time’	**əruun* ‘id.’	U	A
0015	**ərɣŋ* ‘breath, life’	**ərii-gAAn* ‘breathe-NML:RES’	Tg. → Am. (1)	A
0016	**hal(ŋ)* ‘collar’	**xala* ‘tug’	Am. → Tg. (9)	A
0017	**hilɣ~*hilx* ‘tongue’	**xilŋu* ‘id.	U	A
0018	**I-a-* ‘how, to do what’	**jaa* ‘what’	N/A	D
0019	**joqo(n)* ‘Yakut’	**jaakọọ* ‘id.’	Tg. → Am. (9)	A
0020	**keŋ* ‘whale of intermediate size’	**kịaŋa* ‘whale’	Am. → Tg. (9)	A
0021	**la-mo* ‘sea’	**laamọ* ‘id.’	Am. → Tg. (9)	A
0022	**loci~*lụcj* ‘Russian’	**lụụča* ‘id.’	Tg. → Am. (9)	D
0023	**luk* ‘shaggy’	**lokko* ‘id.’	U	A
0024	**mam-* ‘to squeeze’	**nama-* ‘to hold’	U	A
0025	**maŋq-* ‘to be strong,	**ma(-)ŋ(g)a* ‘hard,	Tg. → Am. (1)	A

	difficult’	strong, difficult’		
0026	**mə-* ‘to hear’	**mə(-)də-* ‘to feel’	N/A	D
0027	**me(-)q* ‘musk deer’	**mi̯k(-)čaan* ‘id.’	Am. → Tg. (1)	A
0028	**murŋ* ‘horse’	**mu̯ri̯n* ‘id.’	Tg. → Am. (9)	A
0029	**ŋoŋ(-)j* ‘goose’	**ŋu̯ŋŋa(-kII)* ‘id.’	U	A
0030	**ŋarŋi* ‘fir’	**ŋu̯aŋta* ‘id.’	Tg. → Am.	D
0031	**oroo̯* ‘ginseng’	**o̯ro̯o̯-g-tA* ‘grass’	Tg. → Am. (1)	D
0032	**pʰisk* ‘kite’	**pigəəčuən* ‘id.’	U	D
0033	**pʰumt* ‘midge’	**pormii-g-tA* ‘id.’	U	A
0034	**pʰuv(-)ŋ* ‘saw’	**pu̯u̯-pOOn* ‘saw-NML:INS’	Tg. → Am. (1)	A
0035	**pur-* ‘to spread’	**boruu-* ‘to drop’	U	A
0036	**qʰalm* ‘whale of small size’	**kali̯mu̯* ‘whale’	Am. → Tg. (9)	A
0037	**qʰot(-)r* ‘bear’	**ku̯t(i̯)* ‘id.’	U	A
0038	**qoŋqŋ* ‘bell, jaw harp’	**ko̯o̯ŋ(g)aa(n)(-g-tA)* ‘bell (COL-COL)’	Am. → Tg. (6, 8)	A
0039	**(kʰ)o̯rə* ‘door’	**orkə* ‘id.’	N/A	D
0040	**tʰat-u-* ‘familiarise-TR-’	**tati̯-* ‘to get used to, learn’	U	A
0041	**tapsŋ* ‘salt’	**dabu̯su̯n* ‘id.’	Tg. → Am. (9)	A
0042	**tal(-)q* ‘stroganina’	**tala(kka)* ‘id.’	Am. → Tg. (1, 6 9)	A
0043	**taʁ* ‘epidemic’	**da(x)a-* ‘to be infected’	Tg. → Am. (6)	A
0044	**tʰac* ‘tin’	**tarči̯* ‘tin, lead’	Tg. → Am.	D
0045	**tʰə(-)l(-)ŋku-* ‘to tell a tale’	**təəloŋo(-)* ‘legend, tale’	Am. → Tg. (1)	A
0046	**tʰorq* ‘badger’	**do̯ro̯ko̯o̯n* ‘badger’	N/A	D
0047	**tʰoχ* ‘moose’	**to̯o̯-kII* ‘id.’	Tg. → Am. (1)	A
0048	**uɣŋ~*uɣŋ* ‘loon’	**ookəəŋi* ‘id.’	U	A
0049	**wa-* ‘to fight’	**waa-* ‘to kill’	N/A	D
0050	**χoŋ* ‘sheep’	**ko̯ni̯n* ‘id.’	Tg. → Am. (3, 9)	D

Table 10: Dataset and direction of borrowing

9.2 The stratigraphy of Amuric-Tungusic lexical contacts

Although all items in Table 10 can be reconstructed to Proto-Amuric and Proto-Tungusic, this does not necessarily make them a member of the oldest layer of borrowing. Internal developments within the lineage that a language is a part of are the main source of information on the strata that loanwords should be assigned to (cf. §8.4). Since my reconstruction of Proto-Amuric is based on the Comparative Method, it is more shallow than my reconstruction of Proto-Tungusic as the diversification within the former's lineage is much more limited. Although there are a few Tungusic items that may not be inherited from Proto-Tungusic (for a detailed discussion, I refer to Appendix D), most undoubtedly are. Determining the age of the items in Amuric is then left as my present objective.

Pre-Proto-Amuric, an older stage preceding Proto-Amuric, can be reached by reversing the developments that have been identified through internal and external reconstruction (cf. §6, §7). The presence of initial fricatives has already been put forward as one of the arguments to base a stratigraphy on (cf. §7). This has been applied to PAm. **χoŋ* 'sheep' (cf. 0050) to assign it to a recent stage of borrowing. Because this argumentation was not applicable to any other item, that only leaves considering vowel deletion and the emergence of fricatives. In the following subsections I rearrange my data based on the relevant sound changes they may represent. In this manner, I assess how certain items fit within the history of the Amuric language family.

9.2.1 Vowel deletion in non-initial syllables

Janhunen (2016) has posited that loss of non-initial vowels was a diachronic development in Amuric (cf. §7). My findings are in agreement with this hypothesis. Table 11 shows that word-final CV syllables in Proto-Tungusic regularly correspond to C in Proto-Amuric. Among these items are a few that are most likely of Amuric origin. For them, the Tungusic parallel confirms the erstwhile presence of a vowel that has since disappeared in Amuric.

No.	Proto-Amuric	Proto-Tungusic	DoB
0016	**hal(ŋ)*	**xal**a***	Am. → Tg. (9)
0020	**keŋ*	**kị̈aŋ**a***	Am. → Tg. (9)
0021	**la-mo*[18]	**laam**ọ***	Am. → Tg. (9)
0023	**luk*	**lokk**o***	U
0024	**mam-*	**nam**a**-*	U
0035	**pur-*	**bor**uu**-*	U
0037	**qʰot(-)r*	**kụt(**ị̈**)*	U
0040	**tʰat-u-*	**tat**ị̈**-*	U
0043	**taʁ*	**da(x)**a**-*	Tg. → Am. (6)
0047	**tʰoχ*	**tọọ-k**II***	Tg. → Am. (1)

Table 11: Vowel loss in word-final CV sequences in dissyllabic words

The only counterexamples of disyllabic words without word-final vowel deletion are PAm. **aka-* 'to be older' (cf. 0011), which corresponds to PTg. **akaa* 'elder brother', and PAm. **joqo(n)* 'Yakut' (cf. 0019), which corresponds to PTg. **jaakọọ* 'id.'. The former is a complex case, given that it must have been present in the Amuric lineage relatively early as one of its reflexes, AN *əkə-* underwent the change PAm. **a* > /ə/ (cf. §6.2.2). Both items are also attested with a final consonant, either in the bare stem or a derived stem. Yet this consonant did not block vowel deletion, as can be seen in Table 12.

No.	Proto-Amuric	Proto-Tungusic	DoB
0008	**cʰamŋ*	**sam**aa**n*	Am. → Tg. (1)
0014	**ərŋ*	**ər**uu**n*	U
0028	**murŋ*	**mụr**ị̈**n*	Tg. → Am. (9)
0034	**pʰuv(-)ŋ*	**pụụ-p**OO**n*	Tg. → Am. (1)

Table 12: Vowel loss in word-final CVC-syllables in disyllabic words

As the table shows, the vowel in the final syllable in Tungusic is absent in its Amuric counterpart. This strongly suggests that PAm. **joqo(n)* 'Yakut' is a recent loan and possibly that PAm. **aka-* is of recent introduction as well. If that is the case, then

[18] The reconstruction of this item is itself based on Tungusic. Its later reflex SSN *lam* 'small wave' lacks the final vowel (cf. 0021).

that could give reason to argue that **aka-* was borrowed from Tungusic. However, the issue may still be more complicated. PAm. **ara*(-) 'almost' (cf. 0003), which corresponds to PTg. **aran* 'barely', does not show deletion in the bare stem, but does in the derivation **ara-w-* 'to be equal, approach'. This possibly suggests that there was more specific conditioning to this development than can be retrieved on the basis of the lexical parallels in my dataset. However, since most of my data confirms this behaviour, I will not attempt to resolve this particular issue here.

That word-medial clusters did not prevent vowel deletion also appears clear, as is the pattern in Table 13.

No.	Proto-Amuric	Proto-Tungusic	DoB
0012	**əlv-*	**əlbə-*	U
0013	**amɣ~*amx*	**amŋ**a***	U
0017	**hilɣ~*hilx*	**xilŋ**u***	U
0025	**maŋq-*	**ma(-)ŋ(g)**a***	Tg. → Am. (1)
0029	**ɲoɲ(-)j*	**ɲuŋŋ**a**(-kII)*	U
0033	**pʰumt*	**porm**ii**-g-tA*	U
0038	**qoŋqŋ*	**kọọŋ(g)**aa**(n)(-g-tA)*	Am. → Tg. (6, 8)

Table 13: Vowel deletion in the final syllable of disyllabic words after a cluster

Exceptions are PAm. **amᵚra-* 'to taste' (cf. 0001), which corresponds to PTg. **amta(-)* 'taste' and PAm. **cʰoxt(ụ)-* 'to get drunk' (cf. 0010), which corresponds to PTg. **sọktọọ-* 'id.'. As most examples from Table 13 show that a non-initial vowel was lost after a cluster, it seems safe to assign **amᵚra-* to a later stratum. In the case of **cʰoxt(ụ)-*, the final vowel is only reflected in SSN (cf. CND: 35) and not in any other variety. Therefore, it may have been borrowed separately in SSN. Yet as it can be reconstructed with a velar instead of a uvular, it is probably of recent introduction in the Amuric lineage as a whole, since it is highly possible that this place contrast was introduced after the Pre-Proto-Amuric stage (cf. §6.2.2). This would place it then within the category of newer loans.

In addition to disyllabic words, there are also trisyllabic words in my dataset that seem to imply vowel deletion in Amuric. These are listed in Table 14.

No.	Proto-Amuric	Proto-Tungusic	DoB
0006	**capŋ*	**ʤap**a**-p**OO**n*	Tg. → Am. (1)
0015	**ərɣŋ*	**ər**ii**-g**AA**n*	Tg. → Am. (1)
0036	**qʰalm*	**kal**ị**m**ụ***	Am. → Tg. (9)
0041	**tapsŋ*	**dab**ụ**s**ụ**n*	Tg. → Am.
0042	**tal(-)q*	**tal**a**(kk**a**)*	Am. → Tg. (9)
0048	**uɣn~*uɣŋ*	**ook**əə**ŋ**i***	U

Table 14: Deletion of the second and third vowel in trisyllabic words

As this table shows, the trisyllabic Tungusic words have a monosyllabic parallel in Amuric. In my dataset there were two words that are disyllabic in Amuric but have a trisyllabic parallel in Tungusic. These are PAm. **aχma(-)l-* 'to be a father-in-law' (cf. 0005), corresponding to PTg. **amaa-kAA* 'grandfather, father-in-law' and PAm. **tʰə(-)l(-)ŋku-* 'to tell a tale' (cf. 0045), corresponding to PTg. **təəloŋ(g)o(-)* 'legend, tale'. In both cases, it would appear that the second vowel was deleted while the third stayed intact. Since it is not improbable that PAm. **tʰə(-)l(-)ŋku-* is actually not borrowed but inherited, that would greatly complicate the chronology of vowel loss. However, PAm. **aχma(-)l-* seems to be a loanword of later date. The sense 'father-in-law' is only attested in the Nanaic branch and Manchu (SSTM1: 34-35). In these languages, the reflex *amχa* of PTg. **amaa-kAA* already shows syncope. The observed change could then be external to Amuric. In the case of **tʰə(-)l(-)ŋku,* another explanation could be possible. The final segments may be identified as or was associated with the causative suffix **-ŋku*. Possibly, this suffix contains the transitiviser **-u*, which may also be reconstructed as **-w*. This yields **tʰə(-)l(-)ŋkVw-* and in that case, it would seem that **u* does not reflect an original vowel. If this reconstruction is correct, this lexeme did follow the regular pattern of vowel deletion.

The examples discussed above add further evidence to Janhunen's (2016) argument that vowel loss is a diachronic process. Such items as PAm. **aχma(-)l-* 'to be a father-in-law' , **amᵚra-* 'to taste' and **joqo(n)* 'Yakut' can only have retained their

vowels if they entered into the language after this sound law was not operational any longer. A notable example is **aka-* 'to be older'. If it is a loanword, it must have been borrowed into the unit that preceded the break off of the Nivkh branch, since it is attested in AN with a sound change that postdates its unity with the Nighvng branch (PAm. **a* > /ə/). Thus it would indeed seem that vowel loss preceded this stage in the history of the Amuric language family, as Janhunen suggests.

9.2.2 Equivalents of Tungusic voiceless plosives

According to Austerlitz (1983b; 1982; 1994; 1990: 19-22; 1984a), fricatives in Amuric are the result of intervocalic lenition of plosives. Those plosives that were not affected by this sound law, were most likely of a different nature, possibly geminate (cf. §6.2.1). Accepting this hypothesis creates a conundrum when attempting to fit loanwords into the chronology. As Table 15 and Table 16 show, Proto-Tungusic intervocalic voiceless plosives can correspond to fricatives as well as plosives in Proto-Amuric.

No.	Proto-Amuric	Proto-Tungusic	DoB
0005	**a**χ**ma(-)l-*	**amaa-**k**AA*	Tg. → Am. (1)
0034	**p^{h}u**v**ŋ*	**pụụ-**p**OOn*	Tg. → Am. (1)
0047	**t^{h}o**χ***	**tọọ-**k**II*	Tg. → Am. (1)
0048	**u**ɣ**ŋ~*u**ɣ**ŋ*	**oo**k**əəŋi*	U

Table 15: fricative-plosive correspondences

No.	Proto-Amuric	Proto-Tungusic	DoB
0006	**ca**p**ŋ*	**ʤa**p**a-**p**OOn*	Tg. → Am. (1)
0011	**a**k**a-*	**a**k**aa*	U
0019	**jo**q**o(n)*	**jaa**k**ọọ*	Tg. → Am. (9)
0023	**lu**k***	**lo**kk**o*	U
0037	**q^{h}o**t**(-)r*	**kụ**t**(ị)*	U
0040	**t^{h}a**t**-u-*	**ta**t**ị-*	U
0042	**tal(-)**q***	**tala(**kk**a)*	Am. → Tg. (1, 6, 9)

Table 16: plosive-plosive correspondences

PTg. **amaa-kAA* ‘grandfather, father-in-law’, **pụụ-pOOn* ‘saw-NML:INS’ and **tọọ-kII* ‘moose’ listed in Table 15 and PTg. **ʤapa-pOOn* ‘grab-NML:INS’ and **jaakọọ* ‘Yakut’ from Table 16 are all original to Tungusic. In their Amuric counterparts, the first set contains fricatives, while the latter contains plosives. Considering that PA is a systematic process (cf. §5), it is unlikely that Pre-Proto-Amuric would have settled on two different repair strategies for structures that are identical, i.e. intervocalic plosives. This irregularity is therefore best explained as the result of sound change that led to the Proto-Amuric stage.

For PAm. **capŋ* ‘pincers’, it may be possible to reconstruct **capp(ʰ)Uŋ* < **capʰVpʰUŋ* on the basis of PTg. **ʤapa-pOOn*. The medial plosive reflects then a secondary geminate that was the result of syncope. External reconstruction cannot explain the plosive in PAm. **joqo(n)* ‘Yakut’, however. Therefore, this item must belong to a different layer than those items that show spirantisation, which is also supported by the presence of a vowel in its final syllable (cf. §9.2.1).

Tentative reconstructions such as PTg. **lokko* ‘shaggy’ and **tala(kka)* ‘stroganina’ provide corroborating evidence to geminates in Pre-Proto-Amuric. Especially the latter example, because Pre-Proto-Amuric is its more plausible source. As non-geminate plosives became fricatives, the phonemic distinction between geminate and non-geminate plosives largely disappeared in later Amuric (cf. §6.2.1). Because degemination maintained the presence of plosives in medial position, it is unlikely that plosives in Tungusic loans of later date would be categorised as fricatives. Those loans that contain fricatives, should therefore have been borrowed in an earlier period, the one preceding spirantisation.

Considering that PAm. **joqo(n)* ‘Yakut’ is most likely one of these later loans on account of the vowel in its final syllable, it adds extra credence to the hypothesis that loanwords with medial plosives entered the Amuric lineage more recently. Yet **aχma(-)l-* ‘to be a father-in-law’ provides counterevidence. It is attested with a fricative but the vowel preserved in its final syllable suggests that it is a younger borrowing (cf. §9.2.1). Additionally, its source Ma., Na. or Ulc. *amχa* < PTg. **amaa-kAA* already exhibits spirantisation. Fricatives are also found in Ma. *fufun* ‘saw’ <

PTg. **pụụ-pOOn* ‘id.’, Ma. *toχo* ‘moose’ < PTg. **tọọ-kII* ‘id.’ and Neg. *uxəəŋi* ‘loon’ < PTg. **ookəəŋi* ‘id.’ (SSTM1: 34-35; SSTM2: 191-192, 257, 322-323). Thus it would seem that medial fricatives in PAm. **pʰuvŋ*, **tʰoχ*, **uɣɲ~*uɣŋ* could also be the result of developments external to Amuric.

If this is correct then none of the here discussed correspondences can be accepted as ancient. It is however to be considered that Proto-Tungusic intervocalic plosives were always categorised as geminate in Pre-Proto-Amuric. This kind of PA is sporadically found in ancient Germanic loans in Finnic as well (Kallio 2015: 30). If intervocalic plosives were indeed adapted in this way, then fricative-plosive correspondences as in Table 15 would be reflective of later borrowing. Yet currently this hypothesis remains unproven. Thus, spirantisation of intervocalic voiceless plosives is not a secure measure for gauging the age of an Amuric-Tungusic lexical parallel.

9.2.3 Equivalents of Tungusic voiced plosives

Based on the established pattern for initial obstruents (cf. §7), Amuric lenis obstruents should correspond to voiced plosives in Tungusic. In line with Austerlitz’s (1983b; 1982; 1994; 1990: 19-22; 1984a) hypothesis, it would be expected that, in Amuric, lenis fricatives would be the reflex of intervocalic lenis plosives. Therefore, intervocalic voiced plosives in Proto-Tungusic should correspond to lenis fricatives in Proto-Amuric. A single example in my dataset provides evidence for this hypothesis: PAm. **ərɣŋ* ‘life, breath’ (cf. 0015), which corresponds to PTg. **ərii-gAAn* ‘breathe-NML:RES’. Yet Tungusic non-intervocalic voiced plosives also seem to correspond to lenis fricatives. Notably in PAm. **əlv-* ‘to make a roof’ (cf. 0012). Here PAm. **v* corresponds to the plosive **b* in PTg. **əlbə-* ‘to cover’, which is found after the consonant **l*. One may be inclined to posit that there was originally a vowel after PAm./PTg. **l* were it not that this behaviour is more wide-spread in Amuric-Tungusic lexical parallels, such as those found by Krejnovič (1955a).

The explanation might be that medial lenis plosives always became fricatives no matter the environment. There are however counterexamples. For instance, PTg. **b* in **dabụsụn* ‘salt’ is reflected as PAm. **p* – a plosive – in **tapsŋ* ‘salt’ (cf. 0041).

However, **tavUsUŋ* may have been the intermediate stage. Krejnovič (1955a: 142) found that a Tungusic plosive in a medial cluster corresponds to a fricative in Amuric. His examples more specifically show that a plosive preceded by a nasal or a plosive followed by another plosive correspond to fricatives in Amuric. In my dataset, the former change can be found in PAm. **amᵑra-* 'to taste', which corresponds to PTg. **amta(-)* 'taste' (cf. 0001). An example of the latter is PAm. **cʰoxt(ụ)-* 'to get drunk', which corresponds to PTg. **sǫktǫǫ-* 'id.' (cf. 0010). Possibly, clusters of two fricatives became banned, which caused earlier **tavsŋ* to become OAm. *taps* (cf. Grube 1892: 89). In all modern Amuric languages, this cluster changed to **fc* (cf. CND: 41), the same order of fricative and plosive as in PAm. **cʰoxt(ụ)-*. This suggests that plosive-fricative clusters became preferred. Yet after a survey of the data in Fortescue (CND), it appears that fricative-fricative clusters were not disallowed nor was there a preference for fricative-plosive clusters. For instance, a fricative-fricative cluster is found in PAm. **pəzɣ-* 'to mix up', while a plosive-fricative cluster can be found in PAm. **otɣ* 'excrement' (CND: 29; 130).

Two indisputable counterexamples to positing that lenis plosives always became fricatives are found in PAm. *maŋq-* 'to be strong, difficult' (cf. 0025), corresponding to PTg. **ma(-)ŋ(g)a* 'hard, strong, difficult', and *qoŋqŋ* 'bell, jaw harp' (cf. 0038), which corresponds to PTg. **kǫǫŋ(g)aa(n)(-g-tA)* 'bell'. In both cases, PAm. **-ŋq-* corresponds to PTg. **-ŋ(g)-*. For PAm. **qoŋqŋ*, **-ŋq-* may be close to its original shape, given that Tungusic most likely borrowed this lexeme from Amuric. However, for PAm. **maŋq-* a Tungusic origin is more likely. Aside from contradicting the expectations of my earlier hypothesis, these examples also do not match the pattern observed by Krejnovič (1955a: 142). The expected parallel to PTg. **-ŋ(g)-* would be PAm. **-ŋʁ-*, not **-ŋq-*. Possibly, the cluster PAm. **-ŋʁ-* was banned specifically, a hypothesis I have so far failed to falsify.

No other patterns for the correspondences of Proto-Tungusic medial voiced plosives in (Proto-)Amuric can be observed in my dataset. Thus on the basis of my data, no sound laws can be established that better explain the correspondence of voiced plosives and lenis fricatives. Consequently, they are not suited for determining whether any item found in the Amuric lineage dates to a period where its ancestor interacted with Proto-Tungusic

9.3 Summary of findings

The accepted items displayed in Table 10 represent good candidates for ancient loans with either Pre-Proto-Amuric or Proto-Tungusic as their ultimate source. After subjecting them to a thorough analysis (cf. Appendix D), I was able to propose a direction of borrowing for many of them, fulfilling one of my research objectives.

After analysing how these items reflected the sound laws that have been proposed in earlier research (cf. §6, §7), I have concluded that **amɻa-* 'to taste' (cf. 0001), **amχa(-)l-* 'to be a father-in-law' (cf. 0005), **aka-* 'to be older' (cf. 0011) and **joqo(n)* 'Yakut' (cf. 0019) were not part of the oldest layer of borrowing. The complete absence of vowels in (Proto-)Amuric where Proto-Tungusic preserves them constitutes the only solid indicator that an Amuric-Tungusic lexical parallel is ancient. Given that deletion of vowels in non-initial syllables is found even in recent loans such as PAm. **χoŋ* 'sheep' (cf. 0050), other arguments should also be considered. The importation of initial fricatives is one of such arguments (cf. §7). Therefore, these two developments represent the main indicators of the stratum a word belongs to, pending further research.

10 Closing remarks

Extracting and selecting the data from the available literature on Amuric-Tungusic lexical parallels has been the most time-consuming part of my research. No exact number could be assigned to this dataset, as there were many doublets but I would estimate there to have been close to a thousand items in total. For every individual item of these, I have checked whether they were sufficiently attested in the Amuric and Tungusic language families and whether it was possible to reconstruct them to these families' respective protolanguages. The resulting list contained only fifty items, all of which I have discussed in detail. I have applied my criteria to establish a direction of borrowing and assessed whether they merited further analysis. These short treatments in and of itself provide a number of etymologies that can be of interest to other scholars. After this review of my data, the remaining parallels were to be assigned to layers, with finding the oldest among them as the primary objective. I have only partially succeeded in that endeavour.

The lack of knowledge on Amuric linguistic history beyond the stage retrievable by means of the Comparative Method has been the main challenge in my stratigraphy. While I was able to supply further evidence of vowel loss as a diachronic development, I could not do so for other developments. The correspondence of Tungusic intervocalic plosives to Amuric plosives was contrary to the expectations set by earlier research that employed internal reconstruction. The expected fricatives were found in only a few items and these are possibly later borrowings. Consequently, I can only claim to have found a layer of Amuric-Tungusic lexical parallels that are substantially old but not necessarily as ancient as Pre-Proto-Amuric and Proto-Tungusic must have been.

Thus it can be concluded that further research into the history of Amuric is necessary. The period after the unity of the Nivkh and Nighvng branches is understudied but even more so the period preceding. Some ancient loans may have avoided detection simply because sound changes made them too dissimilar from Tungusic to detect. Of course without descriptions of such sound laws, there is also little to base a stratigraphy of loans on. Therefore, further research into the more distant past of the Amuric language family is the next direction to take in the study of Amuric-Tungusic lexical contact.

To reach further into the past of the Amuric language family, the most promising method appears to be external reconstruction, as Janhunen (2016) has shown. This method connects the history of one linguistic lineage that is poorly understood (i.e. Amuric) to one that is much better understood (i.e. Tungusic). Through application of external reconstruction, a more detailed chronology of developments in the Amuric language family may be created. Accordingly, it may be possible in the future to clarify what layers existed in the contact between Amuric and Tungusic. Perhaps it might even be possible in the future to find solid proof of interaction between Proto-Tungusic and a more distant ancestor to the Amuric language family.

References

Abondolo, Daniel. 2005. “Invariance in Variance: Language, Prominence, and Silence in the Verbal Art of Ritual.” In *Shamanhood: An endangered language*, edited by Juha Pentikäinen and Péter Simoncsics, 49-63. Oslo: Novus Forlag.

Alonso de la Fuente, José Andrés. 2017. “Imperatives in Contact. Linguistic Interactions on the Sakhalin Island.” *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 134: 305-322.

Aralova, Natalia. 2015. *Vowel Harmony in Two Even Dialects: Production and Perception.* Doctoral dissertation, University of Amsterdam.

Austerlitz, Robert. 1956. “Gilyak Nursery Words.” *WORD* 12 (2): 260-279.

Austerlitz, Robert. 1959. “Gilyak Religious Terminology in the Light of Linguistic Analysis.” *The Transactions of the Asiatic Society of Japan* 3 (7): 207-233.

Austerlitz, Robert. 1976. “L’Appellation du Renne en Japonais, Aïnou et surtout en Ghiliak. In: *Tractata Altaic Denis Sinor Sexagenario Optime de Rebus Altaicis Merito Dedicata*, edited by W. Heissig, J. R. Krueger, F. Oinas, and E. Schütz, 45-49. Wiesbaden: Harrassowitz.

Austerlitz, Robert. 1977. “The Study of the Paleosiberian Languages.” In *Roman Jakobson: Echoes of his Scholarship*, edited by Daniel Armstrong and C. H. van Schooneveld, 13-20. Lisse: The Peter de Ridder Press.

Austerlitz, Robert. 1982. “Gilyak Internal Reconstruction, 1: Seven Etyma. *Folia Slavica* 5 (1-3): 81-88.

Austerlitz, Robert. 1983a. “Genetic Affiliation among Proto-Languages.” In *Symposium Saeculare Societatis Fenno-Ugricae*, edited by Juha Janhunen,

Anneli Peräniitty, and Seppo Suhonen, 51-57. Helsinki: Suomalais-ugrilainen seura.

Austerlitz, Robert. 1983b. “On the Penumbra of Questions Surrounding the Internal Reconstruction of Gilyak.” *Research on Language & Social Interaction* 16 (3-4): 75-87.

Austerlitz, Robert. 1983c. “Préhistoire des Guiliaks: Métaux. Gilyak Prehistory: Metals.” *Colloque international Sibérie 1582–1982* (Paris 24–27 mai 1983). Résumés/Abstracts/Rezjume: 1–2. Paris: C.R.N.S., Laboratoire de Slavistique.

Austerlitz, Robert. 1984a. “Gilyak Internal Reconstruction, 2: Iron and Questions Related to Metallurgy.” *Folia Slavica* 7 (1-2): 38-48.

Austerlitz, Robert. 1984b. “On the Vocabulary of Nivkh Shamanism: The Etymon of *qas* (‘drum’) and Related Questions.” In *Shamanism in Eurasia*, edited by Mihály Hoppál, 231-239. Göttingen: Herodot.

Austerlitz, Robert. 1985. “Etymological Frustrations (Gilyak).” *International Journal of American Linguistics* 51 (4): 336-339.

Austerlitz, Robert. 1986a. “Contrasting Fact with Fiction: The Common Denominator in Internal Reconstruction, with a Bibliography.” In *Linguistics across Historical and Geographical Boudaries. In Honour of Jacek Frisiak on the Occasion of his Fiftieth Birthday*, edited by Dieter Kastovsky and Aleksandr Szwedek, 183-192. Berlin: Mouton de Gruyter.

Austerlitz, Robert. 1986b. “Shaman.” *Ural-Altaische Jahrbücher* 58: 143-144.

Austerlitz, Robert. 1990. “Typology in Service of Internal Reconstruction: Saxalin Nivx.” In *Language Typology 1987*, edited by Winfred Lehmann. Amsterdam: John Benjamins Publishing Company.

Austerlitz, Robert. 1994. “Gilyak Internal Reconstruction, 3: Ligneous Matter.” In *Non-Slavic Languages of the USSR: Papers from the Fourth Conference [1985],* edited by Howard I. Anderson, 229-223. Columbus, Ohio: Slavica Publishers.

Barrere, Ian G., and Juha A. Janhunen. “Mongolian Vowel Harmony in a Eurasian Context.” *International Journal of Eurasian Linguistics* 1: 46-77.

Benzing, Johannes. 1956. *Die Tungusischen Sprachen: Versuch einer Vergleichenden Grammatik.* Wiesbaden: Franz Steiner Verlag.

Blažek, Václav. 1997. “Altaic Numerals.” *Studia Etymologica Cracoviensia* 2: 33-75.

Boersma, Paul and Silke Hamann. 2009. “Loanword Adaptation as First-Language Phonological Perception.” In *Loan Phonology,* edited by Andre Calabrese and W. Leo Wetzels, 11-58. Amsterdam: John Benjamins Publishing Company.

Boldyrev, B. V. 1987. *Slovoobrazovanie imen suščestvitel’nyx v tunguso-man’čžurskix jazykax v sravnitel’no-istoričeskom osveščenii.* Novosibirsk: Nauka.

Bouda, Karl. 1960. “Die Verwandtschaftsverhältnisse des Giljakischen.” *Anthropos* 55 (3/4): 355-415.

Bradley, David. 2007. “East and Southeast Asia.” In *Encyclopedia of the World’s Endangered Languages,* edited by Christopher Moseley, 349-424. London: Routledge.

Burykin, A. A. 1987. “Tunguso-man’čžursko-nivxskie leksičeskie paralleli i istorija nivxskogo vokalizma.” *Lingvističeskie Issledovanija* (1987): 46-54.

Burykin, A. A. 1988. “Tunguso-man’čžursko-nivxskie svjazi i problema genetičeskoj prinadležnosti nivxskogo jazyka.” In *Voprosy leksiki i sintaksisa jazykov narodov Krajnego Severa SSSR*, edited by G. A. Čemodanov, 136-150. Leningrad: Leningradskij gosudarstvennyj pedagogičeskij institut.

Burykin, A. A. 1989. “Tunguso-man’čžursko-nivxskie leksičeskie paralleli i istorija nivxskogo vokalizma II.” *Lingvističeskie issledovanija* (1989): 23-25.

Burykin, A. A. 2005. “Nazvanija metallov v nivxskom jazyke.” In *Kjunerovskie čtenija (2001-2004): Kratkoe soderžanie dokladov*, 177-185. Saint-Petersburg: MAÈ RAN.

Burykin, A. A. 2012. “Ètnogenez nivxov, mežètničeskie svjazi v Priamursko-Saxalinskom regione i problema genetičeskoj prinadležnosti nivxskogo jazyka.” In *Lev Šternberg – graždanin, učenyj, pedagog: K 150-letiju so dnja roždenija*, edited by E. A. Rezvan, 275-286. Saint-Petersburg: MAÈ RAN.

Burykin, A. A., and V.A. Popov. 2012. “Problemy izučenija sistemy terminov rodstva i svojstva nivxov.” In *Algebra rodstva,* edited by V. A. Popov. Saint-Petersburg: MAÈ RAN.

Campbell, Lyle. 2013. *Historical Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.

Cincius, V. I. 1975. *Sravnitel’nyj slovar’ tunguso-man’čžurskix jazykov: Tom I*. Leningrad: Nauka.

Cincius, V. I. 1977. *Sravnitel’nyj slovar’ tunguso-man’čžurskix jazykov: Tom II*. Leningrad: Nauka.

Clauson, Gerard. 1959. “The Earliest Turkish Loan Words in Mongolian.” *Central Asiatic Journal* 4 (3): 174-187.

Doerfer, Gerhard. 1973. “Das Kur-Urmiische und seine Verwandten.” *Zentralasiatische Studien* 7: 567-599.

Doerfer, Gerhard. 1978. “Urtungusisch ö.” *Tungusica* 1: 66-116.

Doerfer, G. 1985a. *Mongolo-Tungusica*. Wiesbaden: Harrassowitz.

Doerfer, Gerhard. 1985b. “Terms for Aquatic Animals in the Wu T‘i Ch‘ing Wên Chien.” In *Proceedings of the International Symposium on B. Piłsudski’s Phonographic Records and the Ainu Culture,* 190-202.

Fortescue, Michael. 2011a. *Orientation Systems of the North Pacific Rim.* Copenhagen: Museum Tusculanum Press.

Fortescue, Michael. 2011b. “The Relationship of Nivkh to Chukotko-Kamchatkan Revisited.” *Lingua* 121: 1359-1376.

Fortescue, M. 2016. *Comparative Nivkh Dictionnary*. München: LINCOM.

Greenberg, Joseph H. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press.

Grube, Wilhelm. 1892. *I. Giljakisches Wörterverzeichniss nach den Originalaufzeichnungen von L. v. Schrenck und P. v. Glehn, Zusammengestellt und mit Grammatischen Bemerkungen Versehen.* St. Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.

Gruzdeva, Ekaterina. 1998. *Nivkh.* München: LINCOM.

Gruzdeva, Ekaterina, and Juha Janhunen. 2020. “Notes on the Typological Prehistory of Ghilyak.” *International Journal of Eurasian Linguistics* 2: 1-28.

Gruzdeva, Ekaterina and Marina Temina. 2020. “Nivkh Toponyms in the Amur-Sakhalin Region of the Russian Far East.” *Altai Hakpo* 30: 167-204.

Gruzdeva, Ekaterina. 2021. “River as an Anchor of Nivkh Language Ecology.” In *Down by the Water,* edited by Emilia Mataix and Olesya Khanina. BAR Publishing.

Gusev, Valentin. 2015. “Some Parallels in Grammar between Nivkh and Tungusic Languages.” Journal of the Center for Northern Humanities 8: 63-75.

Halm, Robert. 2017. “Application of the Comparative Method to Vocoid Sequences in Nivkh.” Kansas Working Papers in Linguistics 38: 12-22.

Halm, Robert. 2019. “Initial Clusters in pre-Proto-Nivkh: Internal Reconstruction from Proto-Nivkh.” Kansas Working Papers in Linguistics 40: 1-3.

Halm, Robert and Jay Slater. 2018. “Application of the Comparative Method to Nivkh: Other Regular Sound Correspondences.” Kansas Working Papers in Linguistics 39: 21-36.

Halm, Robert and Jay Slater. 2020. “Application of the Comparative Method to Morpheme-Final Nasals in Nivkh.” Kansas Working Papers in Linguistics 20: 1-21.

Haspelmath, M. 2009. “Lexical Borrowing: Concepts and Issues.” In: Loanwords in the world’s languages: A comparative handbook, edited by M. Haspelmath and U. Tadmor. Berlin: De Gruyter Mouton.

Hattori, Takesi. 1962a. “Bemerkungen zur Phonologie des Südostgiljakischen.” *Studia Phonologica* 2: 1-26.

Hattori, Takesi. 1962b. “Versuch einer Phonologie des Südostgiljakischen (I): Phonembestand und Verteilung.” 北海道学芸大学紀要. 第一部. A, 人文科学編 13(1): 67-130.

Hattori, Takesi. 1962c. “Versuch einer Phonologie des Südostgiljakischen (II): Alternation.” 北海道学芸大学紀要. 第一部. A, 人文科学編 13(2): 29-96.

Haugen, Einar. “The Analysis of Linguistic Borrowing.” *Language* 26 (2): 210-231.

Ikegami, Jiro. 1980. *An Uilta Basic Vocabulary.* Sapporo: Hokkaido University.

Jakobson, Roman. 1971. “Notes on Gilyak.” In *Roman Jakobson: Selected Writings II,* 72- 97. The Hague: Mouton.

Janhunen, Juha. 1981. “Korean Vowel System in North Asian Perspective. *Hangeul* 172: 129-146.

Janhunen, Juha. 1986. “Siberian Shamanistic Terminology.” *Mémoires de la Société Finno-Ougrienne* 194: 97-177.

Janhunen, Juha. 1996. *Manchuria: an Ethnic History*. Helsinki: Mémoires de la Société Finno-Ougrienne.

Janhunen, Juha. 1997. “The Russian Monsters: On the Etymology of an Ethnonymic Complex.” *Studia Etymologica Cracoviensia* 2: 159-165).

Janhunen, Juha. 2003. “Proto-Mongolic.” In *The Mongolic Languages,* edited by Juha Janhunen. London: Routledge.

Janhunen, Juha. 2005. “The Eurasian Shaman: Linguistic Perspectives.” In: *Shamanhood: An Endangered Language,* edited by Juha Pentikäinen and Péter Simoncsics. Oslo: Novus.

Janhunen, Juha. 2008. “Liao: A Manchurian Hydronym and its Ethnohistorical Context.” *Studia Etymologica Cracoviensia* 13 (1): 89-102.

Janhunen, Juha. 2010. “Reconstructing the Language Map of Prehistorical Northeast Asia.” *Studia Orientalia 108: 281-303.*

Janhunen, Juha. 2011. “Unicorn, Mammoth, Whale: Mythological Connections of Zoonyms in North and East Asia.” *Linguistics, Archaeology and the Human Past* 12: 189-222.

Janhunen, Juha. 2012. “The Expansion of Tungusic as an Ethnic and Linguistic Process.” *Recent Advances in Tungusic Linguistics*, edited by Andrej L. Malchukov and Lindsay J. Whaley, 5-16. Wiesbaden: Harrassowitz.

Janhunen, Juha. 2013. “Personal Pronouns in Core Altaic.” In *Shared Grammaticalization with Special Focus on the Transeurasian Languages*, edited by Martine Robbeets and Hubert Cuyckens, 211-226. Amsterdam: John Benjamins Publishing Company.

Janhunen, Juha. 2014. “On the Ethnonyms Orok and Uryankhai.” *Studia Etymologica Cracoviensia* 19: 71-81

Janhunen, Juha. 2016. “Reconstructio Externa Linguae Ghiliacorum.” *Studia Orientalia* 117: 3- 27.

Janhunen, Juha. 2017. “Issues of Comparative Uralic and Altaic Studies (1): The Case of Proto-Mongolic *x.” 北方人文研究 10: 97-104.

Janhunen, Juha. 2019. “Ethnic Groups of Manchuria.” *Oxford Research Encyclopedia of Asian History.* *https://oxfordre.com/asianhistory/view/10.1093/acrefore/9780190277727.001.0001/acrefore-9780190277727-e-57.*

Joseph, Andrew, Seongyeon Ko, and John Whitman. 2020. “A Comparative Approach to Vowel Systems and Harmonies in the Transeurasian Languages and Beyond.” In *The Oxford Guide to the Transeurasian Languages*, edited by Martine Robbeets and Alexander Savelyev, 486-508. Oxford: Oxford University Press.

Kallio, Petri. 2015. “The Stratigraphy of the Germanic Loanwords in Finnic.” In *Early Germanic Languages in Contact*, edited by John Ole Askedal and Hans Frede Nielsen, 23-38. Amsterdam: John Benjamins Publishing Company.

Kang, Yoonjung. 2011. “Loanword Phonology.” In *Companion to Phonology*, edited by Marc van Oostendorp, Colin Erwen, Elizabeth Hume and Keren Rice, 2258-2282. Malden: Blackwell.

Khabtagaeva, Bayarma. 2019. *Language Contact in Siberia*. Leiden: Brill.

Ko, Seongyeon, Andrew Joseph, and John Whitman. 2014. “Comparative Consequences of the Tongue Root Harmony Analysis for Proto-Tungusic, Proto-Mongolic, and Proto-Korean.” In *Paradigm Change in the Transeurasian Languages and Beyond*, edited by Martine Robbeets and Walter Bisang, 141–176. Amsterdam: John Benjamins Publishing Company.

Krejnovič, E. A. 1937. *Fonetika nivxskogo (giljackogo) jazyka*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel’stvo.

Krejnovič, E. A. 1955a. “Giljacko-tunguso-man’čžurskie jazykovye paralleli.” *Doklady i soobščenija Instituta jazykoznanija AN SSSR* 8: 135-167.

Krejnovič, E. A. 1955b. *Slovar’ nivxskogo jazyka*. Unpublished manuscript.

Krejnovič, E. A. 1973. *Nivxgu*. Moscow: Nauka.

Krejnovič, E. A. 1979. “Nivxskij jazyk.” In *Jazyki Azii i Afriki III: Jazyki drevnej Perednej Azii nesemitskie*, edited by V. M. Solncev. Moscow: Nauka.

Laufer, Berthold. 1902. “The Decorative Art of the Amur Tribes.” *Memoirs of the American Museum of Natural History* VII: 1-86.

Li, Bing. 1996. *Tungusic Vowel Harmony: Description and Analysis.* Doctoral dissertation, University of Amsterdam.

Mattissen, Johanna. 2003. *Dependent-Head Synthesis in Nivkh: A Contribution to a Typology of Polysynthesis*. Philadelphia: John Benjamins Publishing Company.

Millar, Robert McColl, and Larry Trask. 2015. *Trask’s Historical Linguistics*. Abingdon-on-Thames: Routledge.

Nedjalkov, Vladimir P., and Galina A. Otaina. 2013. *A Syntax of the Nivkh Language.* Amsterdam: John Benjamins Publishing Company.

Nichols, Johanna. 1996. “The Comparative Method as Heuristic”. In: *The Comparative Method Reviewed : Regularity and Irregularity in Language Change*, edited by Mark Durie and Malcolm Ross. New York: Oxford University Press.

Norman, Jerry. 1978. *A Concise Manchu-English Lexicon*. Seattle: University of Washington Press.

Oskolskaya, Sofia and Lindsay J. Whaley. 2020. “The Classification of the Tungusic Languages.” In *The Oxford Guide to the Transeurasian Languages*, edited by Martine Robbeets and Alexander Savelyev, 81-91. Oxford: Oxford University Press.

Otaina, Galina A., and Vladimir P. Nedjalkov. 2007. “Reciprocal constructions in Nivkh (Gilyak).” In *Reciprocal Constructions,* edited by Vladimir P. Nedjalkov. Amsterdam: John Benjamins Publishing Company.

Panfilov, V. Z. 1973. “Nivxsko-altajskie jazykovye svjazi.” *Voprosy jazykoznanija* 6: 3- 12.

Pevnov, A. M. 1992. “Nivxskij i tunguso-man’čžurskie jazyki: Problemy kontaktov.” In: *B. O. Pilsudskij - issledovatel’ narodov Saxalina (materialy meždunarodnoj naučnoj konferencii. 31 oktjabrja – 2 nojabrja 1990g. g. Južno-Saxalinsk) – T. 2, 25-29.* Južno-Saxalinsk: Saxalinskij oblastnoj kraevedčeskij muzej.

Pevnov, A. M. 2009. “On Some Features of Nivkh and Uilta (in Connection with Prospects of Russian-Japanese Collaboration).” In サハリンの言語世界：北大文学研究科公開シンポジウム報告書, edited by Toshiro Tsumagari, 113-125.

Pevnov, Alexander M. 2012. “The Problem of Localization of the Manchu-Tungusic Homeland.” *Recent Advances in Tungusic Linguistics,* edited by Andrej L. Malchukov and Lindsay J. Whaley, 17-40. Wiesbaden: Harrassowitz.

Pevnov, A. M. 2016. *On the Specific Features of Orok as Compared to the Other Tungusic Languages.* Studia Orientalia 117: 47-63.

Pevnov, Aleksandr M. 2017. “On the Origin of Uilta (Orok) nōni ‘he or she’.” *Journal for the Center for Northern Humanities* (10): 71-77.

Pevnov, A. M. 2018. “Lingvističeskie svidetel’stva istoričeskix kontaktov orokov i oročej na Saxaline.” *Acta Linguistica Petropolitana* XIV (1): 441-463.

Pulleyblank, Edwin G. 1991. *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin.* Vancouver: UBC Press.

Robbeets, Martine. 2015. “Proto-Transeurasian: Where and When?”. *Man in India* 95 (4): 921-946.

Robbeets, Martine and Chuan-Chao Wang. 2020. “The Homeland of Proto-Tungusic Inferred from Contemporary Words and Ancient Genomes.” *Evolutionary Human Sciences* 2 (8): 1-12.

Robbeets, Martine, Juha Janhunen, Alexander Savelyev, and Evgeniya Korovina. 2020. “The Homelands of the Individual Transeurasian Proto-Languages.” In *The Oxford Guide to the Transeurasian Languages*, edited by Martine Robbeets and Alexander Savelyev, 753-771. Oxford: Oxford University Press.

Robbeets, Martine and Alexander Savelyev (eds.). 2020. *The Oxford Guide to the Transeurasian Languages.* Oxford: Oxford University Press.

Rozycki, W. 1994. *Mongol Elements in Manchu*. Bloomington, IN: Indiana University, Research Institute for Inner Asian Studies.

Ryzhkov-Shukumine, Alexandre. 2020. “Long Vowels in Proto-Tungusic.” *International Journal of Eurasian Linguistics* 2: 70-93.

Sangi, V. M., and L. B. Gašilova. 2016. *Slovar’ nivxsko-russkij (saxalinskij dialekt).* Saint Petersburg: Posveščenie.

Savel’jeva, V. N., and Č. M. Taksami. 1965. *Russko-nivxskij slovar’.* Moscow: Sovetskaja ènciklopedija.

Savel'jeva, V. N., and Č. M. Taksami. 1970. *Nivxsko-russkij slovar'*. Moscow: Sovetskaja ènciklopedija.

Smoljak, A. V. 1984. *Tradicionnoe xozjajstvo i material'naja kul'tura narodov Nižnego Amura i Saxalina.* Moscow: Nauka.

Shiraishi, Hidetoshi. 2006. "Topics in Nivkh Phonology." PhD dissertation, Rijksuniversiteit Groningen.

Shiraishi, Hidetoshi. 2009a. "Modelling Initial Weakenings." In *Strength Relations in Phonology,* edited by Kuniya Nasukawa and Phillip Backley. Berlin: De Gruyter.

Shiraishi, Hidetoshi. 2009b. "On the Alternation between CVCVC and CVCC Forms in Nivkh." In 北大文学研究科北方研究教育センター公開シンポジウム「サハリンの言語世界」, edited by Toshiro Tsumagari, 85-94.

Shiraishi, H., and B. Botma. 2016. "Asymmetric Distribution of Vowels in Nivkh." *Studia Orientalia* 117: 39-46.

Shiraishi, Hidetoshi, and Bert Botma. 2017. "On the Diachronic Origin of Nivkh Height Restrictions." In *Language Faculty and Beyond*, edited by Geoff Lindsey and Andrew Nevins, 14:202–14. Amsterdam: John Benjamins Publishing Company.

Shternberg, Lev Iakovlevich. 1999. *The Social Organization of the Gilyak*. American Museum of Natural History.

Sunik, O. P. 1982. *Suščestvitel'noe v tunguso-man'čžurskix jaxykax.* Leningrad: Nauka.

Takahashi, Moritaka. 1942. 樺太ギリヤク語. Osaka: Asahi Shinbun.

Tangiku, I., N. Ja. Tanzina, and N. V. Nitkuk. 2008. Basic Vocabulary of the Sakhalin Dialect of Nivkh Language (Nogliki Dialect). Tokyo: ILCAA.

Tangiku, Itsuji. 2015. “アイヌ語と北方言語の接触―アイヌ語樺太方言の交易品・動植物名における借用語彙”. Unpublished lecture handout.

Tangiku, Itsuji. No date. “Common Words for Instruments among Nivkh, Uilta and Ainu.” Unpublished lecture handout.

von Möllendorff, P. G. 1894. “The Ghilyak Language.” *China Review* 21 (3): 141-146.

von Schrenck, Leopold. 1891. *Reisen und Forschungen im Amur-Lande in den Jahren 1854-1856: Band III.* St. Petersburg: Eggers und Comp. und J. Glasunof.

Vovin, A. 1997. “Voiceless velars in Manchu.” *Journal de la Société Finno-Ougrienne*, 87, 263-280.

Vovin, Alexander. 2016. “On the Linguistic Prehistory of Hokkaidō.” *Studia Orientalia* 117: 29-38.

Zgusta, Richard. 2015. *The Peoples of Northeast Asia through Time: Precolonial Ethnic and Cultural Processes along the coast between Hokkaido and the Bering Strait*. Leiden: Brill.

Appendix A: Transliteration

Original transcription	Romanisation
а	a
ā	aa
ã	ã
ãā	ãã
‘a	’a
ъ̇	ə̣
б	b
б'	bʲ
в	w
в'	wʲ
г	g
ɣ	ɢ
γ	ɣ
г'	gʲ
ɣ'	ɢʲ
ƃ	ʁ
г̄	ḡ
д	d
д'	dʲ
ǯ	ʤ
ц	ʤ
ē	ee
е	e
ᵘе	ie
ę	ẽ
ε̄	εε
æ	ia
ä	ä
и	i
ӣ	ii

и̇	i̤
ӣ	i̤i̤
ӣ	ĩ
ӣ	ĩĩ
и̃	ĩ̤
и̃̄	ĩ̤ĩ̤
і	i
ı	ı
ӣ	ıı
ӣ́	ıı́
ы	y
j	j
к	k
қ	q
к'	kʲ
қ'	qʲ
к̄	k̄
х	x
ҳ	χ
х'	xʲ
ҳ'	χʲ
х̄	x̄
л	l
л'	lʲ
l	l
l’	lʲ
м	m
м'	mʲ
н	n
н'	ɲ
ң	ŋ
ң'	ŋʲ
о	o

ō	oo
o̰	õ
ō̰	õõ
‘o	’o
oo	uo
ӧ	ӧ
ӧ	ӧ
ɵ	ӧ
ɵ̄	ӧ
п	p
п'	pʲ
p	r
p'	rʲ
c	s
c'	sʲ
т	t
т'	tʲ
y	u
ȳ	uu
ẏ	ụ
ẏ̄	ụụ
ỹ	ũ
ỹ̄	ũũ
ẏ̃	ụ̃
ẏ̃̄	ụ̃ụ̃
ÿ	ü
ÿ̄	ü
ф	f
ф'	fʲ
h	h
h'	hʲ
ц	c
цз	ʤ

ч	č
ш	š
ш'	š^{j}
э	ə, e[19]
э̄	əə, ee[20]
ə̃	ə̃
ə̃̄	ə̃ə̃
‘э	’ə
ъ	ə

Table 17: Romanisation of Cincius (SSTM1; SSTM2)

Appendix B: Overview of Proto-Amuric morphology

Affix	Function	Gloss	
**I-*	third person undergoer	3SU	(Mattissen 2003: 55)
**U-*	reciprocal	RECI	(Mattissen 2003: 55)

Table 18: Proto-Amuric verbal prefixes

Affix	Function	Gloss	
**-v*	location nominaliser	NML:LOC	(Gruzdeva 1998: 22; CND: 175)
**-k*	subject nominaliser	NML:SBJ	(Nedjalkov & Otaina 2013: 69-71; CND: 170)
**-nt(I)*	indicative/nominaliser	IND/NML	(Mattissen 2003: 21-22; CND: 169)
**-ŋ*	participle	PTCL	(Mattissen 2003: 185-188; CND: 175)
**-r~*-z*	instrument nominaliser	NML:INS	(Gruzdeva 1998: 22; CND: 175)

Table 19: Proto-Amuric deverbal nominalising suffixes

[19] A romanisation specific to Ewen.
[20] A romanisation specific to Ewen.

Affix	Function	Gloss	
*-*l(a)*	permanent property	PERM	(Mattissen 2003: 229-232; CND: 170)
*-*ŋku*	causative	CAUS	(Gruzdeva 1998: 32; CND: 170)
*-*u*~*-*w*	transitiviser	TR	(Nedjalkov & Otaina 2013: 132-133; CND: 170)

Table 20: Proto-Amuric verbal valency changing and derivational suffixes

Appendix C: Overview of Proto-Tungusic morphology

Affix	Function	Gloss	
*-*dAA*	verbaliser	VBL	(Benzing 1956: 116)
*-*lAA*	verbaliser	VBL	(Benzing 1956: 116)
*-*mAA*	predilection verbaliser	VBL:PRL	(Benzing 1956: 116-117; Boldyrev 1987: 75-76)

Table 21: Proto-Tungusic denominal verbalising suffixes

Affix	Function	Gloss	
*-*gAAn*	result of action nominaliser	NML:RES	(Benzing 1956: 58, Sunik 1982: 85)
*-*n*	nominaliser	NML	(Boldyrev 1987: 149-153)
*-*pOOn*	instrument nominaliser	NML:INS	(Benzing 1956: 63; Ryzhkov-Shukumine 2020: 84; Sunik 1982: 91)

Table 22: Proto-Tungusic deverbal nominalising suffixes

Affix	Function	Gloss	
*-*čAAn*	diminutive	DIM	(Benzing 1956: 60)
*-*g*	collective	COL	(Benzing 1956: 68-69)
*-*g-sA*	class suffix for homogenous masses	COL-COL	(Benzing 1956: 70)
*-*g-tA*	class suffix for things that belong together	COL-COL	(Benzing 1956: 71-72)
*-*kAAn*	diminutive	DIM	(Benzing 1956: 58-59)
*-*ŋ(g)A*	denominal nominaliser of unclear function	NML	Boldyrev (1987: 147)
*-*kII*	suffix found in animal names		(Benzing 1956: 66-67; Sunik 1982: 45-47)
*-*sA*	collective	COL	(Benzing 1956: 69-71
*-*tA*	collective	COL	(Benzing 1956: 71-72)

Table 23: Proto-Tungusic nominal classification, collective and other nominal suffixes

Appendix D: Dictionary

0001. ***amra~amta*** ‘taste’, ***amra-ɟ*** ‘to taste’

PAm.: PAm. **amr̥a-ŋ* ‘taste-PTCL’, **amr̥(a)(-la)-* ‘taste(-PERM)-‘ (CND: 13; Grube 1892: 45)

PTg.: **amta(-)* ‘taste, to taste’ **amta-n* ‘taste-NML’, **amta-lAA-* ‘taste-VBL’ (SSTM1: 39).

Proposed by: von Möllendorff (1894: 144), Grube (1892: 45), Krejnovič (1955a: 158), Austerlitz (1983b: 86), Fortescue (CND: 13).

As the entry shows, both the Proto-Amuric and Proto-Tungusic comparandum are segmentable but also confoundingly similar in shape. The affixes PAm. **-ŋ* and PTg. **-n* are nearly identical in form and function. Furthermore, there is formal similarity

between PAm. *-*l(a)* ‘PERM’ and PTg. *-*lAA* ‘VBL’, yet no functional similarity. To add more complexity to the matter, there are also comparable etyma (**amtan* and **amta-la-*) in Mongolic. Although Doerfer (1985a: 19) has argued that Tungusic borrowed the item from Mongolic, they are morphologically too similar to regard this as proven.

That the donor is non-Amuric seems the most plausible. Although Austerlitz (1983b: 86) reconstructs an element **a*, he explicitly states it is not found in (*)*amu̯a-*, thus no morphological segmentation can be performed that would prove it is original to Amuric. Moreover, a Tungusic origin seems more likely. Evidence may be found in OAm. *ámθlač* ‘to try, to taste’ (Grube 1892: 45). Although PAm. **-l(a)* appears to be suffixed after it, **amu̯a-* is not a qualitative verb and therefore, it seems unexpected to find the suffix of permanent property here. That would favour an identification with PTg. **-lAA* instead, suggesting that Proto-Tungusic or one of its descendants was the donor language.

0002. ***aŋ*** ‘year’

PAm.: **aŋ(i)* ‘year’ (CND: 14)
PTg.: **aŋŋa-* ‘year’ (SSTM1: 43-44)
Proposed by: Grube (1892: 44), von Möllendorff (1894: 144), Krejnovič (1955a: 158), Bouda (1960: 405).

Although these etyma resembled each other enough to merit their inclusion in this list, there is reason to dismiss it. The cluster **ɲŋ* was not illicit in Proto-Amuric, since it is attested in other words, e.g. **ɲŋaʁ* ‘eye’ and **ŋəɲŋər* ‘gum’ (CND: 116, 126), therefore it is unlikely that it was borrowed from Proto-Tungusic. Within Tungusic, only Manchu and Ulcha have *aŋa(n)* as the reflex of PTg. **aŋŋa*, which makes either of them a more plausible source than Proto-Tungusic. An Amuric origin is also unlikely, as there is no explanation for the extension of **aŋ* to **aŋŋa*. In sum, no evidence here seems to point to these words being involved in a very early period of contact between Amuric and Tungusic. Nevertheless, the word might have been borrowed from Tungusic at a later date, which remains to be determined.

0003. ***ara*** ‘almost’, ***v(-)ara-ɟ*** ‘to be equal in quantity’

PAm.: **ara* ‘almost’, **ara-* ‘something almost happened’, **U-ara-* ‘same’, **U-ara-w-*, **I-ara-w-* ‘to be equal, approach’ (Grube 1892: 44; Krejnovič 1955b: 23, 48; CND: 15).

PTg.: **aran* ‘barely’ (SSTM1: 48).

Proposed by: von Möllendorff (1894: 144)

Von Möllendorff (1894: 144) compared OAm. *ārā́* to Ewk. *araj* ‘barely’. This word is listed by Cincius (SSTM1: 48), with such cognates as Ewk. *aran* ‘barely’, Ew. *arạ̄n* ‘id.’ and Ma. *arqan* ‘id.’. According to Doerfer (1985a: 44) Ewk. *araj* and these cognates are all from Mo. *arai*, which is ultimately of Turkic origin. However, while Ewk. *araj* is indistinguishable from Mo. *arai*, Ewk. *aran*, Ew. *arạ̄n* and Ma. *arqan* structurally diverge from it, which suggest that they descend from different etyma.

This root may be reconstructed as **aran*, which are reflected transparently in Ewenki and Ewen. Ma. *arqan* also seems to derive from this root, although this may not be immediately apparent. Its structure suggests that it descends from **aran-kAAn*, a diminutive of PTg. **aran*. Firstly, PTg. **rk* normally yields Ma. /č/ and Ma. /q/ is the regular result of PTg. **k* before RTR vowels (Benzing 1956: 47-48; Li 1996: 155-157). Therefore /rq/ is most likely the result of a syncopated vowel. Reconstructing this vowel places PTg. **k* in intervocalic position. This environment regularly triggers spirantisation of **k*, yet is blocked if there is a nasal preceding (cf. Vovin 1997). To explain the absence of spirantisation, a nasal should be reconstructed, yielding **aran-kAAn*.

OAm. *ārā́*, the Amuric comparandum in von Möllendorff’s (1892: 144) proposal, reflects PAm. **ara*, which is attested as *ara* in AN and ESN, while SSN has *aṛu* (CND: 15). Though syntactically an adverb, it can also be found as the verb root (*)*ara-* ‘something almost happened’ in AN and ESN (Krejnovič 1955b: 23). Additionally, Fortescue (CND: 15) has suggested that a reciprocal form **U-ara-* as well as a derivation **aru-* can also be reconstructed. The reflex of PAm. **U-ara-* is attested as *vara-* ‘to be equal in quantity’ in AN, which has *bara-* as its alternant. On that basis, PAm. **para-* should be reconstructed. However, this could also be

explained as reanalysis of *v*- that is the reflex of **w*- < PAm. **U*-. As for **aru*-, it should be derived from **ara*- by means of the transitiviser suffix **-u*~**-w*. The stem can be combined with the third person undergoer prefix *I*-, which is emblematic of a transitive verb (CND: 15). As an intermediate development from **ara-w*- to **aru*-, syncope should be assumed.

It is possible that from PAm. **ara*- 'something almost happened' the participle **ara-ŋ* was derived, which corresponds to PTg. **aran* 'barely'. PTg. **aran* appears to have only adverbial use, which impedes deriving it from a nominal like **ara-ŋ*. However, **ara-ŋ* could have occurred in a position that would be reminiscent of an adverb. As Mattissen (2003: 185) describes, such participles may occur directly before verbs. Example (5) shows that this is the same position taken by the adverb (*)*ara*.

(5)	*if*	*ara*	*q^{h}o-ɟ*	
	3SG	almost	sleep-IND/NML	
	'He is almost asleep.'			(NRS: 34)

**ara-ŋ* may have been reinterpreted as an adverb by speakers of Proto-Tungusic. Since this segmentation is impossible in Proto-Tungusic, its source must belong to the Amuric lineage. Alternatively, (Pre-)Proto-Amuric could have borrowed PTg. **aran* as **araŋ*, which was then reanalysed as **ara-ŋ*. As both scenarios seem equally plausible, I would regard the direction of borrowing unestablished.

0004. ***arak*** 'vodka'

PAm.: **arak* 'liquor' (NRS: 34; Tangiku et al. 2008: 29; Krejnovič 1955b: 23; Grube 1892: 44).

PTg.: **arakii* 'liquor' (SSTM1: 48)

Proposed by: von Möllendorff (1894: 143), Krejnovič (1955a: 158), Tangiku (2015: 2; n.d.: 1).

This word seems to have spread all over Asia, with Arabic as its ultimate source (Doerfer 1985a: 37). PTg. **arakii* may be reconstructed, since there are ample

cognates available (cf. SSTM1: 48). Yet given its parallels in non-Tungusic languages and its relationship to certain technological innovations, it could also have been borrowed in every Tungusic language separately, merely creating the impression that it is inherited.

Given that liquor was one of the trade goods the Nivkh people bought from the Chinese and Manchus, as reported by Schrenck (1891: 608), Tangiku (2015: 2; n.d.: 1) is most likely correct that PAm. **arak* was borrowed from Manchu. However, in Manchu it is attested as *arki*, with no medial vowel. Therefore, it would seem any of the other Tungusic languages of the Amur Basin may have been the intermediary. Irrespective of its exact origin, it appears to be borrowed from a descendant of Proto-Tungusic rather than Proto-Tungusic itself. Therefore it is a younger loan.

0005. ***axmalk*** ‘father-in-law’

PAm.: **aχma(-)l-* ‘to be a father-in-law’ (CND: 8; Grube 1892: 44-45).
PTg.: **amaa-kAA* ‘grandfather, father-in-law’ (SSTM1: 34-35)
Proposed by: von Möllendorff (1894: 144), Krejnovič (1955a: 158)

While von Möllendorff (1894: 144) and Krejnovič (1955a: 158) both suggest a Tungusic source, neither suggest the same comparandum. Von Möllendorff connects OAm. *ámal* to Tg. (*)*anda* ‘friend’ (SSTM1: 42-43). The phonological differences are too great and therefore the comparison should be rejected. Krejnovič is most likely correct in connecting it to Ma., Na., Ulc. *amχa* ‘father-in-law, which exhibits greater phonological similarity. In Ewenki, there also seems to be the cognate *amaakaa* ‘grandfather’. For all these languages, it seems to be a derivation of **amaa* ‘father’ (SSTM1: 34-35). Although the function of **-kAA* is unclear to me, this evidence is sufficient to establish Proto-Tungusic as the ultimate source.

The correspondence of the final consonants in AN *axmalk* and ESN *aʁmalŋ* are irregular, which requires explanation. It would seem that AN *-k* reflects the subject nominaliser **-k* and ESN *-ŋ* the participle suffix **-ŋ*. The latter could also be attested in OAm. *ámal, ármal* ‘friend, relative of the wife’s side [of the family]’ (cf. Grube 1892: 44-45), with loss of the final nasal as is typical for AN. In addition, its parallel

PTg. **amaa-kAA* would further suggest that **l* in PAm. **aχma(-)l-* is a suffix, perhaps **-l(a)* 'PERM'. Thus it would seem that PAm. **aχma(-)l-* behaves like a verbal stem.

0006. ***cap*** 'pincers'

PAm.: **capŋ* 'pincers' (NRS: 375; Grube 1892: 77).
PTg.: **ʤapa-pOOn* 'grab-NML:INS' (SSTM1: 241)
Proposed by: Krejnovič (1955a: 162)

As comparanda, Krejnovič (1955a: 162) provides only verbal roots descending from PTg. **ʤapa-* 'grab', possibly because there is no transparent equivalent in Tungusic. The closest semantic match is Ewk. *ʤawa-wun* 'tongs, pincers', where *-wun* reflects PTg. **-pOOn* 'NML:INS', a derivation apparently absent from its Tungusic relatives (SSTM1: 241). No PAm. **cap-* exists, while **capŋ* may derive from **ʤapa-pOOn* through **capʰapʰUŋ* > **capp(ʰ)Uŋ* > **capŋ*. Thus, a Tungusic source seems to be the most plausible. This does however require the assumption that the reflexes of PTg. **ʤapa-pOOn* in other Tungusic languages were lost or are unrecorded.

0007. ***cʰaj*** 'tea'

PAm.: **cʰai* 'tea' (NRS: 440; Tangiku et al. 2008: 69).
PTg.: **čaị* 'tea' (SSTM2: 377).
Proposed by: Tangiku (2015: 2; n.d.: 1)

As *cʰaj* 'tea' is both attested in AN and ESN (NgN), the reconstruction should be PAm. **cʰai,* as **cʰaj* would have yielded AN ***cʰəj* (NRS: 440; Tangiku et al. 2008: 69; Halm 2017). Similarly, the word (*)*čaị~čaj* is widely attested in Tungusic (cf. SSTM2: 377) and thus can be reconstructed for Proto-Tungusic. Since these two words are formally identical, it is uncertain what the direction of borrowing would have been. Tangiku (2015: 2; n.d.: 1) suggests Manchu as its origin, yet it could also have been borrowed directly from Ch. 茶 *chá*. Moreover, individual Amuric or Tungusic languages may have borrowed it separately from Chinese (or possibly through Russian or Mongolic) (SSTM2: 377). Considering this, it is most likely not

an old word in neither lineage even though it can technically be reconstructed for both protolanguages. Therefore, it cannot form the basis of any further analysis and, for that purpose, must be rejected.

0008. ***cʰam*** ‘shaman, eagle’

PAm.: **cʰamŋ* ‘eagle, shaman’ (CND: 30, 143; Grube 1892: 78, 81)
PTg.: **samaan* ‘shaman’ (SSTM2: 59)
Proposed by: Grube (1892: 78, 81), von Möllendorff (1894: 145), Krejnovič (1937: 53; 1955a: 163), Bouda (1960: 402), Jakobson (1971: 93), Austerlitz (1984b: 233; 1986b: 143-144), Janhunen (1986: 110-111; 2005: 25-26; 2016: 18-19), Abondolo (2005: 58-59), Fortescue (CND: 143).

Fortescue (CND: 30, 143) reconstructs two distinct etyma: PAm. **sam* ‘shaman’ and PAm. **cʰamŋ* ‘eagle’, even though these two words are homophonous in all Amuric languages. The deciding piece of evidence for PAm. **sam* appears to be SSN *ʂam*, which is an interpretation of <*ʃam*> listed by Takahashi (1942: 214), Fortescue’s main source on SSN. Takahashi also provides <*tʃam*> as a variant of the same word. SSN *cʰamŋ* is also attested (Hattori 1962c: 84). Therefore, Takahashi probably recorded the two alternants of *cʰam(ŋ)*. Thus PAm. **cʰamŋ* should be reconstructed instead, which is homophonous with PAm. **cʰamŋ* ‘eagle’.

Whether this is a true case of homophony or polysemy instead has been a matter of debate. Austerlitz (1984b: 233; 1986b: 143-144) has suggested that there is a mythological link between the eagle and the shaman. Janhunen (2005: 25-26), in contrast, found no mythological connection in the available descriptions of Amuric mythology. These explanations also inform the views of the authors on the source of this word. According to Austerlitz, the connection between ‘eagle’ and ‘shaman’ suggest that PTg. **samaan* was borrowed from Amuric. Janhunen has suggested that only **cʰamŋ* ‘eagle’ belongs to Amuric, while **cʰamŋ* ‘shaman’ was borrowed from Tungusic. Earlier (cf. Janhunen 1986: 110-111), he has also proposed that Tg. **samaan* was borrowed in Amuric with the meaning ‘shaman’, which then later gave rise to the meaning ‘eagle’ motivated by a taboo.

As these extralinguistic arguments have failed to provide conclusive evidence, the morphology of the words should be considered. One may argue that **samaan* is original to Tungusic on the basis that it could ultimately derive from **saa-* 'to know' (SSTM2: 49-51). As Austerlitz (1986b: 144) points out, that leaves me to identify **-maan*. Boldyrev (1987: 75-76) describes the deverbal derivational suffix (*)-*mAA-n* that forms noun denoting agents that are "inclined towards the action" that the root refers to. The suffix consists of **-mAA* 'VBL:PRL' and **-n* 'NML'. Deriving PTg. **samaan* from **saa-mAA-n* would require an explanation for the shortening of the vowel in the first syllable, for which there is none. Therefore, this etymology cannot be correct.

It might be possible to give a plausible morphological analysis for Amuric. Abondolo (2005: 59) suggest that **cʰamŋ* 'shaman, eagle' is derived from **tʰmə-* 'to go across'. A shaman is then 'the one who crosses over'. While it is certainly possible *cʰamŋ* is a deverbal derivation using the participial suffix **-ŋ, *tʰmə-* is most likely not the root that it is derived from. The initial consonant is not palatal nor is the vowel low. The most likely candidate would have the shape **cʰam(V)-* or possibly **cʰ(V)ma-*. As it turns out, there exists an etymon of the latter structure: **cʰma-* 'to visit' (CND: 33). Then PAm. **cʰama-ŋ* would have been 'visitor', which gave rise to the meaning 'shaman' (I leave the semantic extension to 'eagle' for later research). The different vocalisms of PAm. **cʰma-* and **cʰamŋ* may be the result of different stress placements.

Krejnovič (1979: 298-299) reports no clear pattern in assigning stress to syllables within any Amuric language as well as variation between different Amuric languages. Austerlitz (1983b: 78, 1986a: 188; 1994: 229) has explained the varied retention of vowels in different Amuric languages by positing moveable stress. Based on these finding, I postulate that these two words continue variants of the same roots with different syllables stressed. PAm. **cʰma-* 'to visit' continues **cʰamá-*, with stress on the last syllable, and PAm. **cʰamŋ* 'shaman' continues **cʰáma-ŋ,* with stress on the first syllable.

This would of course imply that Proto-Tungusic borrowed **samaan* from an ancestor to the Amuric languages. This would also give reason to reconstruct **s* for Pre-Proto-Amuric as initial **s* in PTg. **samaan* could not be explained otherwise. Although the added vowel length still requires explanation – as Janhunen (2016: 18-19) has pointed out – a greater obstacle for this reconstruction would be the quality of the vowel in PAm. **cʰVma-* > **cʰma-* 'to visit'. I have reconstructed **a* through association with PAm. **cʰamŋ*. However, **V* could also have been a different vowel. Bouda (1960: 402) compares Am. (*)*cʰma-* 'visit' to Na. *ʤiːma-* 'to visit' (cognates are listed under SSMT1: 257). With identical meaning, this counterproposal cannot be ignored. However, the correspondence between the initial plosives is irregular, as Amuric fortis plosives are unlikely to have been categorised as voiced plosives in Tungusic (cf. §7). Thus, I tentatively accept PAm. **cʰma-* < **cʰamá-* < **sama-* until further research proves otherwise.

0009. ***cʰi*** 'you (sg.)'

PAm.: **cʰi* 'you (sg.)' (CND: 32; Grube 1892: 78)
PTg.: **sii* 'you (sg.)' (SSTM2: 72-73)
Proposed by: Krejnovič (1955a: 163), Burykin (1988: 148)

PTg. **sii* was reconstructed as **si* by Benzing (1956: 108), while Pevnov (2017: 72) reconstructs **sii*. Neghidal and Ewen provide the deciding evidence for a long vowel, which is reflected as short in the other Tungusic languages. It should be noted that this long vowel is only found in the nominative stem, while for the other cases the stem was clearly **sin-*, with a short vowel (SSTM: 72). In the Amuric languages, PAm. **cʰi* is attested with the allomorphs (*)*cʰi,* (*)*cʰe* and (*)*cʰ*, depending on vowel harmony and other aspects of the phonological structure of the following case suffix. Similarly, the allomorphs of its bound form (*)*cʰi-*, (*)*cʰe-* and (*)*cʰ-* are conditioned by the phonological structure of the following stem they are prefixed to (Mattissen 2003: 12-13, 55-58).

Krejnovič (1955a: 138-139) has shown that the adaptation of fricatives as plosives in Amuric is motivated by CA, as fricatives and plosives alternate in this system. The most likely environment that would trigger CA of PAm. **cʰi*, would be in its bound

form. However, it is never preceded by an input sequence that would yield a fricative as output (E. Gruzdeva, pers. comm., September 27, 2021). Only if fortition of initial fricatives affected even non-alternating words, could PAm. **cʰi* have been similar enough to PTg. **sii* in an earlier period to be accepted in my dataset. I would rather regard this as a chance match, given that pronouns of a similar shape to these two have been found all over Eurasia; even in unrelated languages (Janhunen 2013: 212-214).

0010. ***cʰoxt-c*** 'to get drunk'

PAm.: **cʰoxt(ụ)-* 'to get drunk' (CND: 35; Grube 1892: 79)
PTg.: **sọktọọ-* 'to get drunk' (SSTM2: 106)
Proposed by: Grube (1892: 79), von Möllendorff (1894: 145), Krejnovič (1955a: 163), Jakobson (1971: 90), Panfilov (1973: 9), Fortescue (CND: 35)

This word also has a parallel in Mongolic (Doerfer 1985a: 39). Since it is not attested with a long vowel in Mongolic, the final long vowel of the root in Tungusic is more difficult to explain. Whether to reconstruct a final vowel for this root in Proto-Amuric is also unclear. A final vowel is only attested in SSN *cʰoxtu-nt* (CND: 35). Given that it is an intransitive verb, SSN /u/ cannot be interpreted as the transitivising suffix **-u*. There are root-final vowels in two of the three variants reported by Grube (1892: 79): *čóxnič, čoxnt* and *čoxtĕnt*. <nič> should probably be read as †*ɲc*, which seems to have been the structure of the indicative/nominaliser at that time (= AN *-ɟ*), which indicates that the root is †*cʰox-*, as is also reflected in *čoxnt*. The second vowel in *čoxtĕnt* may be epenthetic or a transcriptional error. With SSN *cʰoxtu-nt* as the only basis, one could reconstruct **ụ*, to fit the expected pattern of vowel harmony. Its cognates in the other Amuric languages may however be more archaic, since they suggest PAm. **cʰoxt-*. This form would indicate that this word is relatively old, as the correspondent of PAm. **cʰ* is PTg. **s*.

That leaves the direction of borrowing. It cannot be derived from anything transparently in Tungusic. Panfilov (1973: 9) seems to suggest that it bears relation to PAm. **cʰoχ* 'juice, blood' (cf. CND: 35). This word has also been variously

connected to PTg. **səə-g-sə* ‘blood’ and **čuu-g-sə* (Krejnovič (1955a: 163; Bouda 1960: 403; Panfilov 1973: 9; Burykin 1989: 47-48; SSTM2: 107, 411). Neither proposal is a close phonological match to PAm. **cʰoχ* and will therefore not be considered further. Deriving PAm. **cʰoxt(u)-* from PAm. **cʰoχ* would require a suffix of the shape **-tV*. I have been unable to find such a suffix although this may explain why its root is attested as OAm. †*cʰox-*. Yet **cʰoxt-* seems to have contained a velar fricative, while **cʰoχ* contains a uvular fricative. This analysis may then be rejected, although that does not aid in further clarifying the direction of borrowing.

0011. ***əkən*** ‘elder brother’, ***əkə-ɟ*** ‘to be older’

PAm.: **aka-* ‘to be older’, **aka-n*, **aki* ‘elder brother’ (CND: 10; Grube 1892: 43, 53)

PTg.: **akaa*, **akịịn* ‘elder brother’ (SSTM1: 23-24)

Proposed by: Grube (1892: 53), von Möllendroff (1894: 144), Krejnovič (1955a: 158), Panfilov (1973: 10), Bouda (1960: 403), Shternberg (1999: 34), Fortescue (CND: 10).

Various forms of PAm. **aka-* are attested. There is the clear verbal form AN **əkə-ɟ* ‘to be older’ < PAm. **aka-nt(I)*, which also seems to be the origin of ESN *aka-nd* ‘elder brother’. Its counterpart in the Nivkh branch is also attested as *ýkyńč* in Grube (1892: 53). Nominalisations of this root are built using the suffix **-n* instead of the more common **-ŋ*, e.g. AN *əkə-n* ‘elder brother’, SSN *aka-n* ‘elder brother’. Additionally, there is the form ESN *aki* ‘elder brother’ < **aka(-)j* (?), a form of address (Krejnovič 1973: 469; CND: 10).

Although translated here as ‘elder brother’, it is probably better understood as ‘elder male relative’, as it may refer to an elder brother, but also an elder male cousin (Krejnovič 1973: 469). Shternberg (1999: 24) also states that it may refer to an elder female relative but this relation seems to be more generally expressed by (*)*nanak* ‘elder sister’ (Krejnovič 1973: 478-479; CND: 111). It is however not unthinkable that **aka-n* may have referred to female relatives as well, given that the root **aka-* seems to have the more generic meaning of ‘to be older’.

Given that PTg. **akịịn* has a nasal, it may suggest that PAm. **aka-n* is its source, since the nasal seems to be a morpheme. However, the correspondence between the vowels of the final syllables is irregular. A Tungusic origin is not improbable, given that this word has low vowels but needs to be reconstructed with a velar. This is typical of a Tungusic loan (cf. §7). It could also be a chance resemblance, however. As Shternberg (1999: 34) has pointed out already, there are many similar looking words for 'elder brother' across Eurasia, for instance Mo. (*)*aka* 'elder brother', to which Panfilov (1973: 10) AN *əkə-n* and ESN *aka-(n)d* also compares. Yet this does not necessarily mean that it could not have been borrowed since other kinship terms were apparently borrowed as well (cf. 0005).

0012. ***əlv-ɟ*** 'to make a roof', ***əlv-s*** 'roof'

PAm.: **əlv-* 'to make a roof' (CND: 166; NRS: 463; NRS-S: 103).
PTg.: **əlbə-* 'to cover' (SSTM2: 445).
Proposed by: Krejnovič (1955a: 159), Bouda (1960: 405), Burykin (1987: 47).

Fortescue (CND: 166) reconstructs PAm. **əlvr* 'roof'. The final element he compares to the numeral classifier **-vr*, while the root he tentatively assumes to be related to **ul-* 'hight'. This analysis is probably incorrect, considering that a verbal root **əlv(ə)-* is attested in AN *əlv-ɟ* 'to make a roof' and ESN *əlvə-d, əvl-d* 'id.' (NRS: 463; NRS-S: 103). Therefore, PAm. **əlvr* should be analysed as **əlv-r* instead as Bouda (1960: 405) has proposed. The suffix **-r*~**-z* can be identified as the deverbal instrument suffix. This analysis is further supported by the fact that this suffix became AN *-s* and ESN, SSN *-r̥*, a correspondence also reflected in AN *əlv-*s 'roof' and ESN, SSN *əlv-r̥* 'id.'. In contrast, the numeral classifier for places **-vr* is reflected as AN *-vr* and ESN, SSN *-vr̥* (CND: 166, 175, 179). Therefore, I can safely associate PAm. **əlv-r* 'roof' with PAm. **əlv-nt(I)* 'to make a roof'. Although the reflex of PAm. **əlv-* has a final vowel in ESN, it is non-etymological. Savel'jeva and Taksami (NRS: 463) have recorded ESN *əlv-r̥*, while Sangi and Gašilova (NRS-S: 103) report ESN *əlvə-r̥*. This second vowel /ə/ can also be found in the verbal root *əlvə-d* 'to make a roof' listed by them, although they also list *əlv-d* as a variant. Therefore, /ə/ is most likely epenthetic. Neither PAm. **əlv-* nor PTg. *əlbə-* is a

complex stem and as no other criteria seem to apply, the direction of borrowing will remain an indeterminate.

0013. ***əŋg*** ‘mouth’

PAm.: **amɣ~*amx* < **am(V)kV* (?) < **amŋV* (?) (CND: 12; Grube 1892: 45, 46, 53)
PTg.: **amŋa* (SSTM1: 38-39)
Proposed by: von Möllendorff (1894: 144), Krejnovič (1955a: 158), Panfilov (1973: 10), Burykin (1987: 52; 1988: 147)

Although the correspondence PAm. **ɣ~*x* = PTg. **ŋ* is contrary to expectation, given that both languages had **ŋ* as a phoneme, it seems to be somewhat regular. It is also found in PAm. **hilɣ~*hilx* (cf. 0017), which may be compared to PTg. **xilŋu*. This may be sufficient reason to compare these words. Possibly, the consonant was originally **ŋ*, which became **g* through dissimilation caused by **m*. After that it became **ɣ*. The occurrence of a velar after a low vowel is unexpected (cf. §6.2.2), which could be explained by positing **ŋ* as the ancestor to **ɣ~*x,* since there is no evidence of a uvular nasal in the history of Amuric (cf. §6.2, §6.2.4). These sound changes would have preceded the Proto-Amuric stage.

Since the development of PAm. **a* > AN, WSN, NSN /ə/ is also attested in this word, it must be substantially old. Its formal and semantic closeness to PTg. **amŋa* should also be obvious and the irregularities may be explained as I have done above. For the direction of borrowing, currently the evidence is in favour for an Amuric origin. Austerlitz (1983b: 86) reconstructs an element **a*, which can be found in a number of words relating to the mouth, e.g. *af* ‘moustache’. Yet the other elements in his posited derivations of **a* remain unexplained. Therefore, I find it difficult to accept this analysis, which leaves the direction of borrowing unestablished.

0014. ***ər*** ‘time’

PAm.: **ərŋ* ‘time’ (CND: 167)
PTg.: **əruun* ‘time’ (SSTM2: 463-464)

Proposed by: Krejnovič (1955a: 159), Bouda (1960: 402), Cincius (SSTM2: 464), Burykin (1987: 47)

The vocalism for Proto-Tungusic cannot be reconstructed accurately. Ew. *ərin*, Sol. *ərĩĩ* and Uil., Ulc. *ərun* suggest PTg. **əruun*. However, this is contradicted by Ma. *ərin*, Na. *ərĩ*, which both should have had ***ərun* as the reflex of PTg. **əruun*. Additionally, there are also Or. *əruun* and Neg. *əjun*, which are not regular reflexes of PTg.**əruun*. For both languages **əji(i)n* is the expected outcome (Benzing 1956: 24-25, 47). The only solution seems to be supposing intra-Tungusic borrowing, where those languages that have the regular reflex were the donor to those languages with the closest irregular reflex. Whether it was borrowed from Tungusic to Amuric or the other way around, is unclear. There is no evidence that would point to either direction.

0015. ***ərx*** 'breath'

PAm.: **ərɣŋ* 'breath, life' (Takahashi 1942: 227; Krejnovič 1955a: 159; NRS: 466)
PTg.: **ərii-gAAn* 'breathe-NML:RES' (SSTM2: 464)
Proposed by: Krejnovič (1955a: 159, Cincius (SSTM2: 464)

In the Nivghng branch, ESN *ərɣəŋ* 'life' and SSN *yřgan* 'id.' are found, which suggest there may have been a final vowel. However, this is not attested in AN *ərx* 'breath' (Takahashi 1942: 227; Krejnovič 1955a: 159; NRS: 466). Therefore, in ESN and SSN, the vowel must be epenthetic. Proto-Tungusic is the unambiguous donor of this word, since it is transparently a derivation of **ərii-* 'to breathe', using the suffix **-gAAn* 'NML:RES'.

0016. ***hal*** 'collar'

PAm.: **hal(ŋ)* 'collar' (CND: 70)
PTg.: **xala* 'tug' (SSTM1: 28-29)
Proposed by: Krejnovič (1955a: 165), Bouda (1960: 405), Tangiku (2015: 4; n.d.: 1)

Cincius (SSTM1: 29) reconstructs PTg. **xalag*, however, final **g* is only reflected in the reflex Ewk. *alaɣ*. As suggested by Khabtagaeva (2019: 227), this has to be identified as the collective suffix (*)-*g*. Therefore, PTg. **xala* should be reconstructed. PAm. **halŋ* can be reconstructed on the basis of ESN *halŋ* and AN *hal* (CND: 70). The comparison with PTg. **xala* requires that the nasal in ESN *halŋ* is secondary, which is a possibility given that a secondary nasal also seems to be found in ESN *χasaŋ* 'scissors', which corresponds to Ulc. *xaʤa* 'scissors' (CND: 165).

In Ewenki, *alaɣ* 'tug' is part of a reindeer harness. Its cognates in Ewen refers to a parts of a reindeer sleigh but also a sleigh pulled by dogs. In the Tungusic languages of the Amur and Sakhalin belonging to the Ewenic, Orochic and Nanaic branches, the *(x)ala* is a part of a dog harness, as it also is in the Amuric languages (SSTM1: 28-29; Krejnovič 1955a: 165). As there are no clear formal arguments to base a direction of borrowing on, the semantics of PTg. **xala* and PAm. **halŋ* provide the main clues for basing a direction of borrowing on. Most evidence would point to a connection for the harness of a dog sleigh, which could be the original meaning.

Pevnov (1992: 26-27) and Tangiku (2015: 3-4; n.d.: 1-2) have pointed out that the Amuric and Tungusic languages share a number of terms connected with dog sleighs. Pevnov expresses surprise by the fact that lexemes from a semantic domain connected with native Nivkh culture are of Amur Tungusic origin. Zgusta (2015: 156-157) classifies this cultural practice as part of a Neolithic stratum in the cultures of the indigenous peoples of the Lower Amur, i.e. those speaking Amuric and Tungusic languages. If Janhunen (2012: 14) is correct, then the Amuric languages would have entered this area first before the Tungusic languages did, which would support that this word for 'collar' or 'tug' ended up in the Amuric languages before it did in Tungusic, if it is not an inherited word in Amuric. In this scenario, an Amuric source is the most likely for PTg. **xala*.

0017. ***hilx*** 'tongue'

PAm.: **hilx*~**hilɣ* 'tongue' (CND: 74)

PTg.: **xilŋu* ‘tongue’ (SSTM1: 316-317)
Proposed by: von Möllendorff (1894: 145), Krejnovič (1955a: 165), Burykin (1987: 50; 1988: 149; 2012: 281)

As with **amx~*amɣ* ‘mouth’ (cf. 0013), PAm. **x~*ɣ* appears to correspond to PTg. **ŋ*. As with the aforementioned example, earlier **ŋ* could have changed to **g* and then to **ɣ*. However, the conditioning is not the same, as **l* is not a nasal. Possibly the change was unconditioned aside from being in the middle of the word and in a cluster. Neither PAm. **hilx~*hilɣ* nor PTg. **xilŋu* seem to be derivations nor do any of the other criteria I could use apply here. Consequently, no direction of borrowing can be established.

0018. ***j-a-ɟ*** ‘to do what’

PAm.: **I-a-* ‘how, to do what’ (CND: 81)
PTg.: **jaa* ‘what’ (SSTM1: 286-288)
Proposed by: Cincius (SSTM1: 288)

PAm. **I-a-* ‘how, to do what’ and PTg. **jaa* ‘what’ may not be the same word. Savel’jeva and Taksami (NRS: 495) list AN *jaɟ* and *aɟ* which means that *j-* is here the third person undergoer prefix (<**I-*) and *a-* is the root. This may be reason to dismiss the comparison, but alternatively may be reason to pronounce Pre-Proto-Amuric as the donor language, since the Tungusic form cannot be segmented. Given these uncertainties, I will not consider this item further.

0019. ***joqo, jeχo*** ‘Yakut’

PAm.: **joqo(n)* ‘Yakut’ (NRS: 82; NRS-S: 22)
PTg.: **jaakoo* ‘Yakut’ (SSTM1: 338-339)
Proposed by: Bouda (1960: 405), Cincius (SSTM1: 339)

Considering that Tungusic speaking peoples have lived in closer geographical proximity to the Yakut than Amuric speaking peoples have, PAm. **joqo(n)* is most likely borrowed from (Proto-)Tungusic. A final nasal is only attested in ESN *joqon*, which also has *joqo*. This segment may therefore be non-etymological.

0020. ***keŋ*** ‘whale of intermediate size’

PAm.: **keŋ* ‘whale of intermediate size’ (CND: 58)
PTg.: **kịaŋa* ‘whale’ (SSTM1: 388)
Proposed by: Krejnovič (1955a: 164), Bouda (1960: 403), Cincius (SSTM1: 388), Doerfer (1985a: 261; 185b: 198), Burykin (1989: 45)

PTg. **kịaŋa* can be reconstructed as there are sufficient cognates to draw from. It is attested as EEwk. *keeŋas* ‘dolphin leather’, Ew. *kịaŋas* ‘thin (of leather or paper)’, which both contain an additional element (*)-*s*, which may be identified as the collective suffix *-s*A*. In Nanaic it is found in Uil., Ulc. *keŋa* ‘whale’ (SSTM1: 388). Doerfer’s (1985a: 261; 1985b: 198) observation that many Tungusic terms related to aquatic animals have parallels in Amuric, may be explained by the fact that Amuric speaking peoples have practiced fishing and maritime hunting for a long time (cf. Zgusta 2015: 99-100). Since Amuric speaking peoples likely arrived at the coast before Tungusic speaking peoples did (cf. Janhunen 2012: 14), the environment may also be a considered as evidence that the Tungusic peoples borrowed the word for ‘whale’ from Amuric.

Am. (*)*keŋ* was most likely borrowed recently. In Ewenki and Ewen it is found in only a single variety of both languages. Nevertheless, the information provided does point to **kịaŋa* as the original shape, which may also aid in the reconstruction of an earlier stage of PAm. **keŋ*. SSN *keŋ* apparently triggers spirantisation instead of fortition in CA. In this variety, there is a split between ‘strong’ /m, n, ɲ, ŋ/ and ‘weak’ /n, ŋ/. After the strong nasals, the output of CA is a fricative (Gruzdeva 1998: 19; Hattori 1962c). Halm and Slater (2020) reconstruct this behaviour to Proto-Amuric and, in order to explain it, they tentatively assume contrasts of both strength and length in the nasals. The Tungusic data however points to a final vowel, which regularly trigger spirantisation. Thus, positing PAm. **keŋa* could solve this issue.

There have been two alternative etymologies. Doerfer (1985a: 261; 185b: 198) and Cincius (SSTM1: 450) have suggested a relationship with (*)*kəŋgin* ‘walrus, sea lion, cuttlefish’, which may be dismissed. It is attested only in Oroch, Ulcha, Nanai and Manchu and therefore is not necessarily ancient. Additionally, it is phonologically

too distinct from PAm. **keŋ*. Von Möllendorff (1894: 143) has suggested that OAm. *keṅ* is related to Ch. 鯨 'whale', which can be romanised as *jīng* for modern Mandarin Chinese and its pronunciation can be reconstructed as †*giajng* for Early Middle Chinese (Grube 1894: 56; Pulleyblank 1991: 158). Although phonologically similar to PAm. **keŋ*, it leaves the final vowel in the Tungusic comparanda unexplained and does not explain the 'strong' behaviour of the nasal. Thus it remains more likely that Proto-Amuric did not borrow this word and inherited it from Pre-Proto-Amuric.

0021. ****la-mo*** 'sea'

PAm.: **la-mo* 'sea'
PTg.: **laamọ* 'sea' (SSTM1: 490-491)
Proposed by: Janhunen (2016: 12-13; 2008: 97-100)

The only reflex of PAm. **la-mo* attested in Amuric seems to be SSN *lam* 'small wave' (Yamaguchi & Izutsu 2004: 30, as cited by Pevnov 2016: 57). It cannot be found in any of the lexicographic sources I have used. Although the evidence is clearly lacking, Janhunen's (2016: 12-13; 2008: 97-100) argumentation is quite convincing. In short, he identifies it in PAm. **lams* 'east wind', which was borrowed by the Ewenic languages, cf. Ewk. *laamus* '(east) wind, east, snow, bad weather' and Ew. *namạs* 'deep snow' (SSTM1: 491). PAm. **la*, from which it is derived, means 'wind' and is also provide the AN name for the Amur river, which could indicate that it also referred to a water basin. Janhunen identifies this root also in other words, such as **la-ɣ̣* 'wave', **la-sq* 'bullhead', **la-ŋr* 'seal' (see CND: 92-95 for comparanda). Easterly wind would come from the direction of the Sea of Okhotsk, which would connect the meaning 'east' and 'sea'.

However, only AN *lams* means 'east wind', its cognates in ESN and SSN mean 'north' and 'north wind' respectively (CND: 92). The geographical location of the sea, relative to where ESN is spoken, is to the east, while it was to the east or south of SSN. So the cardinal direction where the wind is coming from and the location of the sea relative to the settlements where the languages are or were spoken are not as strong as AN *lams* would suggest.

Yet spatial reference in the Amuric languages is not based on the cardinal directions, but rather the local environment, the flow of the river and position relative to the coast or riverbank. This also affects how the direction of the wind is expressed. For example, *tʰaŋgi* ‘northwest wind’ is based on the root *tʰa-* ‘from the shore or other side of the river’. Alternatively put, it is wind that goes in the same direction as the Lower Amur flows (Gruzdeva 2021: 7-9; Fortescue 2011a: 88-92). Therefore, **lams* could be based on a variety of other points of reference and not just the sea.

The solution may lie in a reinterpretation of **la* as ‘current’, which is a feature that unifies wind and bodies of water. This could be supported by the aforementioned PAm. **laᵚr* ‘wave’ but also its homonym PAm. **laᵚr* ‘seaweed’. Fortescue (CND: 93) connects the latter with PAm. **larlar-* ‘to sway’, but it may just as well explain the meaning of the former. Given that PAm. **la* most commonly refers to ‘wind’, that would imply that this current could go in any direction. Yet PAm. **laᵚr* would imply that its direction is towards or along the coast, as does PAm. **lav-* ‘to moor’, which may contain the element **la*, and that Fortescue (CND: 95) connects to PAm. **lav-* ‘to go side by side’. Possibly, this then means that **lams* is wind that follows a sea current along the coast. I am however not aware of any fieldwork that has checked with modern speakers how the direction of **lams* is perceived.

Nevertheless, these considerations as well as the fact that wind and water are connected in the Amuric orientation system provide sufficient reason to interpret PAm. **la-mo* as something associated with a large body of water such as the sea. An issue with any application of internal reconstruction is what the individual elements mean. Although PAm. **la* is based on a root that is preserved in all Amuric languages (cf. CND: 92), **-mo* is not. Perhaps it is to be identified with the suffix (*)-*m* that forms verb stems of spatial reference. For instance, the aforementioned root *tʰa-* ‘from the shore or other side of the river’ can be found in **tʰam-* ‘to be far away from the shore’ (Gruzdeva 2021: 8).

Given that PAm. **la-mo* is only attested directly in SSN *lam* ‘wave’, the comparison with PTg. **laamọ* remains tentative. If the comparison is to be accepted, then the direction of borrowing was from Pre-Proto-Amuric to Proto-Tungusic as Janhunen

(2016: 12-13; 2008: 97-100) has argued. It clearly fits within a morphological paradigm in Pre-Proto-Amuric but not in (Proto-)Tungusic. This further supports Doerfer's (1985a: 261; 1985b: 198) observation that maritime and riverine terminology in Tungusic is of Amuric origin. Since Pevnov (2012: 31) localises the Proto-Tungusic homeland near the sea on the basis of PTg. **laamọ*, an Amuric origin would also be an argument against his hypothesis.

0022. ***loci, locʰa*** 'Russian'

PAm.: **loci~*lụcj* 'Russian' (Grube 1892: 74; NRS: 164; NRS-S: 43)
PTg.: **lụụča* 'Russian' (SSTM1: 513)
Proposed by: Bouda (1960: 405), Janhunen (2016: 13-14), Shiraishi and Botma (2016: 44)

In both Amuric and Tungusic, there are sufficient cognates to reconstruct a word meaning 'Russian' to their protolanguages. However, due to historical considerations, it seems more likely that it entered into both languages at a later date and was most likely borrowed into Amuric from a Tungusic source, as Janhunen (1997; 2016: 13-14) suggests. Since it is not an old loan, it will not be considered further.

0023. ***luk*** 'shaggy'

PAm.: **luk* 'shaggy' (NRS: 165; NRS-S: 43)
PTg.: **lokko* 'shaggy' (SSTM1: 508-509)
Proposed by: Krejnovič (1955a: 160), Bouda (1960: 405), Cincius (SSTM1: 509)

The reconstruction PTg. **lokko* is adapted from Doerfer (1978: 79). He reconstructs a geminate for Tungusic on the basis of Ma. *luku*, since intervocalic velar plosives usually become fricatives and therefore certain conditioning should be assumed that would have blocked this development. In neither (Pre-)Proto-Amuric nor Proto-Tungusic, the forms are morphologically complex and no other criteria seem to apply, which leaves the direction of borrowing undetermined.

0024. ***mam-ɟ*** ‘to squeeze’

PAm.: **mam-* ‘to squeeze’ (CND: 101)
PTg.: **nama-* ‘to hold’ (SSTM1: 581)
Proposed by: Cincius (SSTM1: 581)

To relate the two to each other, it has to be assumed that the initial **n* changed to **m* in Proto-Amuric by way of assimilation. For Proto-Tungusic, dissimilation would have to be assumed. PTg. **nama-* ‘to hold’ closely resembles PTg. **nama* ‘pack’ (SSTM1: 580-581). The semantic extension of ‘pack’ to ‘to hold’ or the other way around is not farfetched and therefore one may be a derivation of the other. Yet there is no morphological material used for this derivation, which provides little basis to establish a direction of borrowing on. Either direction could have been possible.

0025. *maŋɢ-ɟ* ‘to be strong, difficult’

PAm.: **maŋq-* ‘to be strong, difficult’ (CND: 101-102)
PTg.: **ma(-)ŋ(g)a* ‘hard, strong, difficult’ (SSTM1: 529-530)
Proposed by: Grube (1892: 108), von Möllendorff (1894: 146), Krejnovič (1955a: 160), Bouda (1960: 401), Panfilov (1973: 9), Cincius (SSTM1: 530), Fortescue (CND: 101-102)

Krejnovič (1955a: 160) also compares Am. (*)*maŋq-* to Ewk. *mannii* ‘strong’. In Cincius’s (cf. SSTM1: 528) entry where this word can be found, most words seem to continue PTg. **masị* ‘strong’. There is a possibility that PTg. **maŋ(g)a* contains the suffix **-ŋ(g)A* of unclear function. Although **ma-* is not attested as a root separately, it may be the same root found in **masị*. If that is the case, Proto-Tungusic must be the donor language.

0026. ***mə-ɟ*** ‘to hear’

PAm.: **mə-* ‘to hear’ (CND: 109-110)
PTg.: **mə(-)də-* ‘to feel’ (SSTM1: 563-564)
Proposed by: Krejnovič (1955a: 160), Bouda (1960: 401), Burykin (1987: 47),

Burykin (1988: 147)

PTg. **mədə-* ‘to feel’ also has a parallel in Mo. (*)*mede-* ‘to know’. Doerfer (1985a: 24) is undecided which is the recipient and which is the donor language. Connecting it with PAm. **mə-* ‘to know’ required Burykin (1987: 47) to posit a suffix **-də*. There is a denominal verbaliser **-dAA* in Proto-Tungusic, but this suffix has to be reconstructed with vowel length, which is not reflected in PTg. **mədə-*. This leaves little phonological evidence to draw upon for comparison. Therefore, it should be dismissed.

0027. ***meq*** ‘musk deer’

PAm.: **me(-)q* ‘musk deer’ (CND: 103-104)
PTg.: **mi̯k(-)čaan* (SSTM1: 536)
Proposed by: Burykin (1987: 49-50)

Reflexes of PTg. **mi̯kčaan* ‘musk deer’ are only attested in Ewenki and Manchu (SSTM1: 536). Given the linguistic and geographical distance between the two, this evidence would seem sufficient to reconstruct PTg. **mi̯kčaan*. The first syllable **mi̯k* is a solid correspondences to PAm. **meq*, yet that leaves etymologising the second syllable. This would be the suffix **-čAAn*, a diminutive suffix. PAm. **meq* may be segmentable. At least one reliable example of a Pre-Proto-Amuric suffix **-q* that creates animals is available: PAm. **pʰe-q* ‘chicken’ from **pʰe-* ‘to pick, to gather’ (Austerlitz 1983b: 76-77). There is however no verb from which **meq* could have been derived. Nevertheless, the identification of **-q* as a possible suffix does favour Pre-Proto-Amuric as the donor language.

0028. ***mur*** ‘horse’

PAm.: **murŋ* (CND: 109)
PTg.: **mu̯ri̯n* (SSTM1: 558-559)
Proposed by: Grube (1892: 111), von Möllendorff (1894: 146), Krejnovič (1955a: 160), Bouda (1960: 401), Panfilov (1973: 8), Tangiku (2015: 24; n.d.: 7), Robbeets (2015: 936-938), Janhunen (2016: 15-

16), Fortescue (CND: 109)

It is generally assumed that this word spread from Tungusic to Amuric. As the first evidence of horse riding in Northeast Asia goes back to the first millennium BC and is it is most likely of recent introduction in both languages. This can also be tracked by considering the spread of the word for 'horse'. The ultimate source of PTg. **mụrịn* 'horse' is PMo. **morï-n* 'id.', which is segmentable and therefore a Mongolic language should be the donor language (Robbeets 2015: 936-938). Yet Proto-Mongolic is not the immediate source of PTg. **mụrịn* since the vowel **ụ* is not the expected equivalent of PMo. **o*, therefore it should be borrowed from a Mongolic language that underwent a change PMo. **o* > **u~*ụ* (Janhunen 2016: 15-16). Given the phonological similarity of PAm. **murŋ* to PTg. **mụrịn* as well as the southern origin of horse riding, it would indeed appear that Amuric borrowed this word from Tungusic.

0029. *ɲoɲi* 'goose'

PAm.: **ɲoɲ(-)j* 'goose' (Grube 1892: 87; NRS: 209; NRS-S: 52)
PTg.: **ɲụŋɲa(-kII)* 'goose' (SSTM1: 646-647)
Proposed by: Grube (1892: 87), von Möllendorff (1894: 145), Krejnovič (1955a: 166), Bouda (1960: 405), Cincius (SSTM1: 647), (Burykin 1987: 49)

In Amuric, *ɲoɲ(-)i* seems to contain the suffix **Vj* found by Janhunen (2016: 14-15) and Austerlitz (1976). PTg. **ɲụŋɲa* appears to have been integrated into the Tungusic system of animal naming at a late date, since not all cognates are suffixed with **-kII*, which is commonly found in animal names. This may point to borrowing but not necessarily. The direction of borrowing is therefore unclear.

0030. ***ŋarŋi*** 'fir'

PAm.: **ŋarŋi* 'fir' (CND: 122)
PTg.: **ŋụaŋta* 'fir' (SSTM1: 657-658)
Proposed by: von Möllendorff (1894: 145), Bouda (1960: 404), Cincius

(SSTM1: 654)

Bouda (1960: 404) compares ESN *ŋań* ‘fir’ to Na. *waŋta* ‘fir’, which has to little correspondence to be accepted. Cincius compares AN *ŋarŋi* ‘fir’ to Ewk. *ɲərmi* ‘larch’. The most probable comparison was proposed by von Möllendorff (1894: 145), who compared OAm. *ṅarni* ‘pine tree’ to Tg. *ngankta* = Ewk. *ŋaaŋtə* ‘fir’ (Grube 1892: 68). Possibly, PAm. **ŋarŋi*, corresponds to earlier **ŋaŋ̯rV-j*, with metathesis. The regular Proto-Amuric equivalent of PTg. **ŋt* would be **ŋ̯r*, as spirantisation occurs after nasals, which was observed by Krejnovič (1955a: 142). The same process is also found in PAm. **amɲ̯ra-* (cf. 0001). While still a possible borrowing, the donor language should be one belonging to the Ewenic branch. In this branch, PTg. **u̯a* > **aa* occurred, which may be observed in Neg. *ŋaaŋta* ‘fir’, while it is generally preserved in Nanaic, e.g. Na. *waŋta* ‘id.’, although with loss of initial **ŋ* (SSTM1: 657-658; Ryzhkov-Shukumine 2020: 73). Since some remnant of **u̯* should have been retained in Amuric if this were a borrowing from or to Proto-Tungusic, it is of post-Proto-Tungusic age and should therefore not be considered further given the framework of my thesis.

0031. *oroɲ̯* ‘ginseng’

PAm.: **oroɲ̯* ‘ginseng’ (Grube 1892: 50; NRS: 247; NRS-S: 61)
PTg.: **o̜ro̜o̜-g-ta* ‘grass-COL-COL’ (SSTM2: 24)
Proposed by: Janhunen (2016: 17-18)

Janhunen (2016: 17-18) proposes an intermediate stage of **orota* from **oroo-g-ta* ‘grass’, which would be the form that lead to AN *oroɲ̯* ‘ginseng’. Its meaning is instead taken over from Ma. *orχoda* ‘ginseng’, which is a compound of *orχo* ‘grass’ (< PTg. **oroo-g-ta*) and *da* ‘root’. The development **gt* > **t* > **ɲ̯* is unexpected, given that Tungusic clusters of two plosives systematically yield fricative-plosive clusters in Amuric (Krejnovič 1955a: 142).

The reflex *oroɲ̯* ‘ginseng’ of PAm. **oroɲ̯* ‘id.’ is attested in both AN and ESN (NRS: 247; NRS-S: 61). A plausible cognate may also be identified in Grube’s (1892: 50) data: OAm. *orohóxt* ‘burnet’. This may be analysed as *oroh-óxt*, where <oroh>

probably is to be read as †*oroχ*, which closely resembles **oroχo*, the form that should have preceded Ma. *orχo*. <óxt>, in turn, can be identified as AN *oxt* ‘medicine, powder’, which corresponds to Ma. *okto* ‘id.’ and similar words in other Tungusic languages (Krejnovič 1955a: 159). It would seem that this earlier form †*oroχ* is the ancestor of what later became (*)*oroᵲ*. The development **χ* > *ᵲ* is not regular, therefore another solution should be proposed. Morphological reanalysis is plausible but there seems to be no derivational morpheme that **χ* could have been associated with.

The irregularity of the development posited by Janhunen (2016: 17-18) casts doubt on the age of PAm. **oroᵲ* ‘ginseng’. Additionally, the semantic parallel with Manchu may be too specific for a later development. This is further reinforced by the possibility of **oroχ* as an earlier shape of the word, which closely corresponds to Ma. *orχo*. Thus it would have been part of a later stage of contact. Consequently, I have too little grounds to regard this item as belonging to an ancient stage of contact and therefore cannot accept it.

0032. ***pʰisk*** ‘kite’

PAm.: **pʰisk* ‘kite’ (Grube 1892: 101-102; NRS: 287; NRS-S: 69)
PTg.: **pigəəčuən* ‘kite’ (SSTM2: 322-323)
Proposed by: Grube (1892: 101-102), von Möllendorff (1894: 146), Krejnovič (1955a: 161), Bouda (1960: 399)

Bouda (1960: 399) views the final **-k* as a suffix. While certainly a possibility, given that dorsals seem to show up more commonly in animal names (cf. 0027). In that case, Na. *piičuən* ‘kite’ is the most similar to PAm. **pʰisk* ‘id.’. However, there is also a second possibility that would suggest this word belongs to an ancient stage of contact. PAm. **pʰisk* may have been closer to PTg. **pigəəčuən* by assuming the following intermediate stages of development: **pʰikVcʰV(ŋ)* > **pʰiks(ŋ)* > **pʰisk(ŋ)*. That does raise the question what happened with the nasal however. Deletion of PAm. **ŋ* is a regular process in AN, which does not occur in ESN. Yet for ESN *pʰisk* has been recorded as well (NRS-S: 69).

For Ewenki, both *hiɣən* ‘kite’ and *hiɣəəčəən* ‘id.’ are attested (cf. SSTM2: 322-323), which suggests that *hiɣəəčəən* is derived from *hiɣən* by way of the diminutive suffix (*)-*čAAn*. This would suggest that PAm. **pʰisk* is borrowed from Proto-Tungusic. However, Ewk. /əə/ is not a regular reflex of PTg. **uə,* which should have become /uu/ instead (Ryzhkov-Shukumine 2020: 83-86). Nevertheless, Na. *piičuən* does suggest this to be the original vocalism. Given that there are such indeterminates, I will not consider this item further.

0033. ***pʰudm*** ‘midge’

PAm.: **pʰumt* ‘midge’ (CND: 137)
PTg.: **pormii-g-tA* ‘midge-COL-COL’ (SSTM2: 348)
Proposed by: Krejnovič (1955a: 161), Burykin (1987: 48)

The vocalism of this item can be retrieved by considering Ewk. *hunmii-ktə* ‘midge’, Ew. *hum-tə* ‘id.’ and Na. *purmi-ktə* ‘id.’. As Doerfer (1978: 89-90) has shown, this points to the vowel of the first syllable to be non-RTR /o/. The reflexes of the cluster are not fully regular but based on the same three examples, **rm* may be retrieved, in accordance with Benzing (1956: 49).

PAm. **pʰumt* ‘midge’ is based on AN *pʰudm* ‘id.’ and ESN *pʰədm* ‘id.’. In both of these languages /d/ usually reflects **Nt*. The nasal can be identified as **m* in this case, so I have assumed metathesis. PTg. **r* corresponds to PAm. **t* probably through post-nasal fortition with **pʰumr* as an earlier form of the same word. Based on my reconstruction for Proto-Tungusic, **pʰurmV* must be assumed to be the most ancient form of the word and therefore there seem to have been two instances of metathesis in the history of this lexeme. The similarities are great enough to assume borrowing, although the direction is unclear.

0034. ***pʰuf*** ‘saw’, ***fuv-ɟ*** ‘to saw’

PAm.: **pʰuv(-)ŋ* ‘saw’, **I-pʰuv-* ‘to saw’ (CND: 138)
PTg.: **pụ̈ụ̈-pOOn* ‘saw-NML:INS’, **pụ̈ụ̈-* ‘to saw’ (SSTM1: 336)
Proposed by: Grube (1892: 103), von Möllendorff (1894: 146), Krejnovič

(1955a: 161), Bouda (1960: 400), Burykin (1987: 49), Tangiku (2015: 14, 21; n.d.: 3, 5)

The reconstruction PTg. **pụụ-* ‘to saw’ is adapted from Ryzhkov-Shukumine (2020: 84). As he has shown, all later Tungusic language that appear to reflect **pụpụ-* instead, must have derived it from **pụụ-pOOn* ‘saw’, itself an earlier derivation from PTg. **pụụ-*. As for Proto-Amuric, the nominal form would be **pʰuv-ŋ* a participial derivation of **I-pʰuv-* ‘to saw’, according to Mattissen (2003: 28). Yet considering that the common instrument nominaliser of the Amuric language family is **-r~*-z* and not **-ŋ,* this seems improbable. Instead, I propose that **pʰuvŋ* was reanalysed as a deverbal nominalisation, which gave rise to the verbal root **pʰuv-*. For that same reason, **pʰuvŋ* cannot be regarded as segmentable, which makes Proto-Tungusic the more likely candidate to be the donor language, given that **pụụ-pOOn* is a transparent nominalisation on the basis of the instrument nominaliser **-pOOn*.

0035. ***vur-*** ‘to spread’, ***pus*** ‘rubbish’

PAm.: **pur-* ‘to spread’, **pur-r~-z* ‘rubbish’ (CND: 27)
PTg.: **boruu-* ‘to drop, lose (SSTM1: 113)
Proposed by:

Fortescue (CND: 27) connects PAm. **pur-r~-z* ‘rubbish’ (or **pur* in his reconstruction) to PAm. **pur-* ‘to spread’. This must be correct, since **pur-r~-z* is reflected as AN *pus*, ESN *puṛ*. The correspondence AN /s/ = ESN /ṛ/ is reminiscent of the deverbal instrument nominaliser, cf. AN *-s*, ESN *-ṛ* (CND: 175). On that basis, I have interpreted the correspondence as resulting from the nominalisation, which leads me to the segmentation **pur-r~-z*. On the Tungusic side, I have reconstructed PTg. **boruu-* on the basis of Ewk. *burii-* ‘to drop, throw out’ and Ma. *buru-bu-* ‘to lose oneself’ (SSTM1: 113). For the same vowel correspondences, Doerfer (1978: 85-86) also reconstructs PTg. **CoCu(u)*.

PAm. **pur-r~-z* has been compared to PTg. **bor(ə)ki* ‘dust, powder’ (cf. Bouda 1960: 400), while PTg. **boruu-* has been compared to PAm. **por-~*poz-* ‘to lie down’ (cf. Burykin 1988: 145; 1989: 47) (SSTM1: 113; CND: 27, 25). The former

etymology could be dismissed due to too much absent phonological material in the Amuric comparandum. Additionally, there is the plausible connection with **pur-* 'to spread'. The latter comparison is to be dismissed, since the low vowel in PAm. **por-~*poz-* should correspond to an RTR vowel in Proto-Tungusic, yet PTg. **boruu-* contains only non-RTR vowels. Additionally, they are semantically not a good match, nor is their valency the same: PAm. **por-~*poz-* is intransitive, while PTg. **boruu-* is transitive. As a counterproposal, I suggest that PTg. **boruu-* corresponds to PAm. **pur-*, which are semantically close and have the same valency. The correspondence of Proto-Tungusic non-RTR vowels to a Proto-Amuric high vowels is more regular. As neither PTg. **boruu-* nor PAm. **pur-* can be segmented further, there is little to found a direction of borrowing on, so it remains unclear.

0036. ***qʰalm*** 'whale of small size'

PAm.: **qʰalm* 'whale of small size' (Grube 1892: 54; NRS: 147; NRS-S: 36)
PTg.: **kalịmụ* 'whale' (SSTM1: 366-367)
Proposed by: Krejnovič (1955a: 164), Doerfer (1985a: 261; 1985b: 197-198), Janhunen (2011: 206-207; 2016: 13)

Krejnovič (1955a: 164) lists *qʰalm*, while Savel'jeva and Taksami (NRS: 147) cite AN *qalm*, although within the same part of their dictionary as words beginning with /qʰ/. Sangi and Gašilova (NRS-S: 36) provide ESN *qalm*. Grube's (1892) OAm. *kalm* is also not particularly helpful in determining the specification for laryngeal features of the initial consonant. Krejnovič's and Savel'jeva and Taksami's data may be sufficient to reconstruct **qʰ*.

In Tungusic, the correspondences are not regular, cf. Ewk. *kalim*, Ew. *qalịm*, Neg. *kalịm*, Or. *kaalima~kaalma*, Ud. *kalima*, Ulc. *qalma*, Or. *qalịma*, Na. *qaalịma*, Ma. *qalimu*, all meaning 'whale' (SSTM1: 366-367). Per Doerfer (1985b: 197-198) and Janhunen (2011: 206-207; 2016: 13) the final vowel was most likely **ụ*, while those words with different vocalisms would be part of different layers of borrowing from Amuric. It has subsequently been borrowed into the Mongolic languages as well as

Yakut. As part of the semantic domain of marine terminology, it is highly plausible that it was borrowed from Pre-Proto-Amuric into Proto-Tungusic.

0037. ***qʰotr*** ‘bear’

PAm.: **qʰot(-)r* (Grube 1892: 58-59; NRS: 159)
PTg.: **kụt(i)* (SSTM1: 440)
Proposed by: von Möllendorff (1894: 144), Burykin (1987: 49), Pevnov (1992: 28)

PAm. **qʰotr* is only attested in AN *qʰotr* (NRS: 159). Grube (1892: 58-59) also has OAm. *kotr* for WSN, but this variety is closely related to AN. There is a possibility that this is the original term for ‘bear’ in Amuric but has since been replaced by the taboo form (*)*cʰxəf.* This terms is a derivation of the verb root **cʰxə-* ‘to be eerie’ (normally found reduplicated), with the location nominaliser **-v* (Austerlitz 1959: 212-213). However, Fortescue (CND: 141) suggests that even **qʰotr* may be a taboo word: a derivation from **qʰo-* ‘to sleep’. What the function of **-tr* would be, is unclear.

Reflexes of PTg. **kụtị* are only attested in Ewenki (both in an eastern and western variety) and Manchu, cf. Ewk. *kuti* ‘bear’, Ma. *quatiki, qutqa* ‘bear cub’, *quatiri* ‘animal resembling a bear’ (SSTM1: 440). Von Möllendorff (1894: 144) proposed a comparison to Ma. *qutqa*, Burykin (1987: 49) suggested Ewk. *kuti* and Ma. *quatiri*, the latter of which was also suggested by Pevnov (1992: 28). While Ma. *quatiri* bears the greatest resemblance to PAm. **qʰotr*, the final element *-ri* is most likely not part of the root. There are numerous nouns with this suffix, which all meaning ‘something resembling X’. For instance, Ma. *dobiri* ‘an animal resembling a fox that can climb trees’ from *dobi* ‘fox’ (Norman 1978: 60). The root is most likely *qut-* as attested in *qutqa*, where *-qa* is most likely the diminutive suffix descending from PTg. **-kAAn*. Although this root is attested with a final /i/ in both *quatiki* and *quatiri*, this cannot be part of the root originally, as PTg. **t* > Ma. /č/ before **i*, **ị* (Benzing 1956: 31). This would then leave identifying the function of *-r* in PAm. **qʰotr.* Possibly, this had a similar function to PTg. **-kII*. Cincius (SSTM1: 504) has proposed to compare Ewk. *loŋkii* ‘salmon’, Ew. *nọŋqị* ‘coho salmon’ to AN *loŋr* ‘trout’. In this comparison, Tg.

(*)-*kII* corresponds to Am. *-*r*. Only if Fortescue (CND: 141) is correct that PAm. *q^h*otr* derives from PAm. *q^h*o-* is there morphological evidence that the word is of Amuric origin. Otherwise, there seems little basis to establish a direction of borrowing on.

0038. ***qoŋɢ*** ‘bell’

PAm.: **qoŋqŋ* ‘bell, jaw harp’ (CND: 67-68)
PTg.: **kọọŋ(g)aa(n)(-g-tA)* ‘bell(-COL-COL)’ (SSTM1: 410)
Proposed by: Krejnovič (1955a: 164); Tangiku (2015: 21, 22; n.d.: 5); Fortescue (CND: 67)

AN *qoŋɢ* ‘jaw harp’ and ESN *qoŋɢoŋ* ‘bell’ are attested as reflexes of **qoŋqŋ* (CND: 67-68). Since the vowel in its final syllable is only attested in ESN and not preserved in AN, it is most likely epenthetic. Fortescue (CND: 67) postulates a Chinese source for PAm. **qoŋqŋ* ‘bell, jaw harp’ with a Tungusic intermediary but provides no Chinese comparandum. Reflexes of PTg. **kọọŋ(g)aa(n)-g-tA* ‘bell’ are found throughout the Tungusic language family, so if it were borrowed from an unknown Chinese word, this borrowing should be far in the past. Most of its descendants are suffixed with **-g-tA*, while Manchu has *χoŋgon* ‘bell’ (SSTM1: 410). This suggests that a nasal was originally part of the root, which makes PTg. **kọọŋ(g)aa(n)* a close match to PAm. **qoŋqŋ*. In both protolanguages, the word could also be of onomatopoeic origin and not be related at all. Certainly, PAm. **qoŋqŋ* refers to two dissimilar musical instruments, which would suggest that their name is based on the sound they produce rather than their shape. This may rule out Proto-Tungusic as the donor language but not necessarily Pre-Proto-Amuric. In all Amuric languages, reflexes of PAm. **qoŋqŋ* have an initial lenis consonant (cf. CND: 67-68), which is not the expected outcome of PA if the word had a Tungusic source (cf. §7). Proto-Tungusic would have borrowed this word only in a single sense of the word and this semantic narrowing may be an argument to further support an Amuric origin.

0039. ***ɻə*** ‘door’

PAm.: **(k^h)ɻə* ‘door’ (Grube 1892: 62, 76; CND: 153)

PTg.: **orkə* ‘door’ (SSTM2: 286)
Proposed by: Austerlitz (1977: 19)

Austerlitz (1977: 19) compares this item to Uil. *utə* ‘door’, which descends from PTg. **orkə* ‘door’, the reconstruction of which is adapted from Doerfer (1978: 85). As Pevnov (2009: 119) shows, Uil. *utə* is most likely an example of depalatalization. The earlier form *učə* ‘door’ is still preserved in Ulcha, where /č/ is the regular outcome of PTg. **rk* (Bezing 1956: 47-48; SSTM2: 286). As addressed in Section 6.2.1, PAm. **(kʰ)ɻ̥rə* is attested with an initial †*k(ʰ)* in OAm. *ktö* ‘door, cape’ (cf. Grube 1892: 62), which is not found in *utə* and may therefore be a mismatch. The initial velar found in OAm. *ktö* is not retained in any other Amuric language and in fact Grube (1892: 76) also lists OAm. *ršy* ‘door’, which clearly lacks †*k(ʰ).* Still, this could be indicative of variation and does not necessarily have to mean that *ktö* is a transcriptional error. Reconstructing initial **kʰ* makes PAm. **(kʰ)ɻ̥rə* ‘door’ much more similar to PTg. **orkə* and a comparison with it may provide a solution to why PAm. **kʰ* was lost. However, no other similar parallels are presently available to me, which leaves me with too little to found an analysis on. I will therefore not consider this comparison further and instead leave it to later research.

0040. ***ratu-ɟ*** ‘to teach’

PAm.: **tʰat-u-* ‘familiarise-TR-’ (Grube 1892: 97; Krejnovič 1955a: 162; 1955b: 590; NRS: 304, 345)
PTg.: **tatị-* ‘to get used to, learn’ (SSTM2: 170-171)
Proposed by: Grube (1892: 97), von Möllendorff (1894: 146), Krejnovič (1955a: 162)

For the reconstruction of this root, there is some disagreement among the data I have used. Grube (1892: 97) lists OAm. *θátuč,* which is most likely a transcription of †*ɻ̥ratu(n)c,* with a fortis fricative. The palatalisation of the final segment is typical of a variety belonging to the Nivkh group. However, Savel’jeva and Taksami (NRS: 304) list AN *ratuɟ*, where the initial fricative is lenis. Krejnovič (1955a: 162; 1955b: 590) reports ESN *tʰatund* and *ɻ̥atunt*, where the alternants /tʰ/ and /ɻ̥/ are clearly reflective of an underlying fortis obstruent. Considering that its Old Amuric cognate contains a

fortis fricative, which is still reflected in ESN, I reconstruct PAm. **I-tʰatu-*. The final **u* is unexpected in this root, as it contains a low vowel. This would suggest it is a suffix, most likely the transitiviser suffix **-w~*-u*. Further support for this analysis may be found in AN *tatc* 'to make oneself familiar with'. For this word, it also has to be assumed that AN /t/ is a later development or transcriptional error.

A clear parallel to AN *tatc* 'to make oneself familiar with' is PTg. **tatị-* 'to get used to, learn'. The causative of this verb generally means 'to teach' in the Tungusic languages (cf. SSTM2: 170-171), as does PAm. **I-tʰat-u-*. The comparison is therefore a solid semantic and formal match. PAm. **tʰat-* and PTg. **tatị-* cannot be analysed any further and any other criteria are seemingly not applicable. Therefore, either language could have been the donor or recipient.

0041. ***tafc*** 'salt'

PAm.: **tapsŋ* 'salt' (CND: 41)
PTg.: **dabụsụn* 'salt' (SSTM1: 186)
Proposed by: Grube (1892: 89), von Möllendorff (1894: 143), Krejnovič (1955a: 162), Bouda (1960: 401), Panfilov (1973: 8-9), Abondolo (2005: 58), Tangiku (2015: 14, 21; n.d.: 3, 5), Fortescue (CND: 41)

In ESN *tafciŋ* 'salt' and SSN *dafciŋ* 'id.' there is a vowel before the final nasal that is not preserved in AN *tafc* (CND: 41). This means that it is most likely epenthetic. In Old Amuric, aside from *tafč* 'salt' also *taps* 'id.' is attested (Grube 1892: 89). After this period a kind of "featural metathesis" seems to have occurred that affected all varieties, with the plosive **p* becoming a fricative **v* or **f* and the fricative **s* became a plosive **c*. That PAm. **tapsŋ* must be the original shape of the word is further suggested by the fact that it bears a clear parallel to PTg. **dabụsụn* 'salt'.

PTg. **dabụsụn* is transparently a Mongolic borrowing, as it contains the suffix **-sU/n* denoting liquids or liquifiable masses (Janhunen 2003: 13). This same analysis cannot be done in Tungusic. Similar to entry 0028, a transmission from Mongolic through Tungusic could be assumed for PAm. **tapsŋ*.

The Nanaic and Orochic languages generally reflect **daụsụn* (SSTM1: 186). According to Doerfer (1985a: 38), its source is Ma. *dabsun* ‘salt’. Yet in Nanaic and Orochic, elision of intervocalic **b* is a regular process affecting words inherited from Proto-Tungusic (Benzing 1956: 34). Therefore, **daụsụn* cannot be differentiated from any inherited items. Certainly, a later period of borrowing cannot be ruled out, so this leaves the age of the parallel undetermined, although there is sufficient data to reconstruct PTg. **dabụsụn* as it is also attested in Western Ewenki (cf. SSTM1: 186). Therefore, according to my criteria (cf. §8.2), I have to accept the parallel.

0042. ***talk*** ‘stroganina’

PAm.: **tal(-)q* ‘stroganina’, **tal(-)qa-~*tal(-)va-* ‘to be raw’ (Grube 1892: 88; Krejnovič 1955a: 162; 1955b: 618; CND: 38)
PTg.: **tala(kka)* ‘stroganina’ (SSTM2: 156)
Proposed by: Krejnovič (1955a: 162)

As a noun, *talk~talq* ‘stroganina’ (a dish made of raw fish) is apparently only attested in AN. However, a similar root (*)*talqa-* ‘to be raw’ is attested for ESN and SSN. Its equivalent in AN is *talva-*, which is also found in ESN (Grube 1892: 88; Krejnovič 1955a: 162; 1955b: 618; CND: 38). The resemblance of AN *talk~talq* to ESN, SSN *talqa-* would suggest that it is much older. Either it is derived directly from **talqa-* or from a root **tal-*. This may be the root in AN, ESN *talva-* and ESN, SSN *talqa-*. The irregular correspondence between **v* and **q* in these comparanda would suggest that **-va* and **-qa* are to be analysed as separate morphemes. PAm. **talq* ‘stroganina’ would then be a derivation from PAm. **tal-*, using the suffix **-q*. Yet as long as the function of these suffixes remains unidentified, this analysis remains within the realm of speculation. Reconstructing a separate etymon **tal-* would explain why Tungusic reflects two different roots, one with and one without a dorsal consonant: **tala* and **talakka* (SSTM2: 156). The geminate in the latter is based on Ma. *talqa-*, defined by Norman (1978: 271) ‘to cook fish half-done’. Other Tungusic languages show that the velar was in intervocalic position, which is the conditioning environment for spirantisation in Manchu (Vovin 1997: 278). The absence of spirantisation could be explained by positing gemination, which the plosive in PAm. **tal(-)q* could point to as well (cf. §6.2.1).

Krejnovič (1955a: 162) views the similarity of Am. (*)*tal(-)q* to Tg. (*)*talakka* as an irregularity, since an Amuric lenis plosive corresponds to a Tungusic voiceless plosive. However, given that it is part of the semantic domain of marine terminology, it is highly possible that this is a borrowing from Pre-Proto-Amuric in Proto-Tungusic. The categorisation of PAm. **t* to PTg. **t* is then to be expected (cf. §7). This provides further evidence of an Amuric origin.

0043. ***taχ*** ‘epidemic’

PAm.: **taʁ* ‘epidemic, cold, influenza, smallpox’ (CND: 40; Grube 1892: 88)
PTg.: **da(x)a-* ‘to be infected’ (SSTM1: 184)
Proposed by: Krejnovič (1955a: 162)

Krejnovič (1955a: 162) compares AN *taχ* ‘epidemic disease’ to Na. *daktụ* ‘infected’. Phonologically, this comparison is not sound, since only a limited number of segments are corresponding. Nevertheless, if one considers that Na. *daktụ* is derived from PTg. **daa-* ‘to be infected’, the comparison could be accepted. This root is attested as *daa-* in both Ewenki and Nanai. Usually, a long vowel in Ewenki corresponds to a short vowel in Nanai, yet in a few cases it does not. To explain this irregular correspondence as well as why PTg. **x* is only attested word-initially, Doerfer (1973: 584) has posited that Ewk. /aa/ and Na. /aa/ go back to PTg. **axa*.[21] PTg. **daa-* is then to be reconstructed as **daxa-*, which makes it a perfect match of PAm. **taʁ*. Provisionally, a Tungusic origin can be posited on the basis of the correspondence PAm. **t* = PTg. **d.* If it had been borrowed from Pre-Proto-Amuric, current evidence would suggest that PAm. **t* would have been categorised as PTg. **t* instead (cf. §7).

0044. ***tʰac*** ‘tin’

[21] Recently, Ryzhkov-Shukumine (2020: 72-75) has shown that Nanai regularly retains PTg. **aa* at least in the initial syllable. Thus, it would be unnecessary to posit PTg. **daxa-*. Nevertheless, this reconstruction remains a possibility, given that the general lack of word-medial **x* in Proto-Tungusic is still unexplained. Because it seems to be reflected as **χ~*ʁ* in Proto-Amuric, I tentatively accept the reconstruction **daxa-*.

PAm.: **tʰac* ‘tin’ (CND: 145)

PTg.: **tarčị* ‘tin, lead’ (SSTM2: 169)

Proposed by: Grube (1892: 88), von Möllendorff (1894: 88), Krejnovič (1955a: 162), Austerlitz (1983c: 2; 1984a: 45), Burykin (2005: 4), Fortescue (CND: 145)

Grube (1892: 88), von Möllendorff (1894: 88), Krejnovič (1955a: 162) compare Am. (*)*tʰac* ‘tin’ to PTg. **tụụḍʒa* ‘lead’ and its descendants (cf. SSTM2: 205). Austerlitz (1983c: 2) is likely correct in doubting this connection, as the correspondences are irregular. Burykin (2005: 4) and Fortescue (CND: 145) proposed different reflexes of PTg. **tarčị*, which is a much closer match. Even more so Ulc. *tači* and Neg. *tatči* (SSTM2: 169). As both of these languages exhibit an innovation from Proto-Tungusic, an innovation also present in Proto-Amuric, it would appear that PAm. *tʰac* was borrowed from one of these descendants of Proto-Tungusic and thus must be excluded from further analysis.

0045. ***tʰəlgu*** ‘legend’

PAm.: **tʰə(-)l(-)ŋku-* ‘to tell a tale’ (CND: 154)

PTg.: **tʰəəloŋ(g)o(-)* ‘legend, tale’ (SSTM2: 233)

Proposed by: Krejnovič (1955a: 162), Bouda (1960: 401), Panfilov (1973: 11), Cincius (SSTM2: 233), Janhunen (2016: 22)

Sangi and Gašilova (NRS-S: 88) list a complete paradigm which shows that PAm. **tʰəlŋku-* ‘to tell a tale’ is a verb, cf. *tʰəlgu-d~tʰəlgu-nd* ‘to tell’, *tʰəlgu-k* ‘storyteller’, *tʰəlgu-ɻ* ‘legend, myth’, with the indicative/nominaliser suffix *-nd*, subject nominaliser *-k* and the instrument nominaliser *-ɻ* respectively. In Tungusic, **tʰəəloŋ(g)o(-)* acts as both a noun and a verb (cf. SSTM2: 233). The vocalism of this etymon is unclear. Benzing (1956: 37) reconstructs **tālüŋü* (= **təəluŋu*) but Or. *təəluŋu,* Na. *təəluŋgu* suggest that the vowels of the second and third syllable could not have been **u*, as Or. ***təəliŋi* would have been the expected outcome. The correspondence between Oroch and Nanai can best be explained as reflective of PTg. **o*, which Ew. *teeləŋ~teeloŋ* also implies (SSTM2: 233).

According to Bouda (1960: 401), Panfilov (1973: 11) and Fortescue (CND: 154), PAm. **tʰəl* may be derived from **tʰə-* 'to be far'. According to Fortescue, the individual elements that make up PAm. **tʰə-l-ŋku* should be identified as **-l(a)*, suffix of permanent property, and **-ŋku,* the causative. Proto-Tungusic **təəloŋ(g)o(-)* resists segmentation and therefore it was most likely borrowed from Pre-Proto-Amuric contra Janhunen (2016: 22) and Krejnovič (1955a: 162). Although increased valency is usually also coupled with spirantisation (caused by the third person undergoer prefix **I-*) this does not happen in causatives derived from intransitive verbs (cf. Nedjalkov & Otaina 2013: 234). Thus there appears to be no obstacle to accepting an Amuric origin for this word.

0046. ***tʰorq*** 'badger'

PAm.: **tʰorq* 'badger' (Grube 1892: 91; RNS: 60; NRS: 383; RNS-S: 60)
PTg.: **dǫrǫkǫǫn* (SSTM1: 217)
Proposed by: Grube (1892: 91), von Möllendorff (1894: 145)

PTg. **dǫrǫkǫǫn* has a close parallel in PMo. **dorogon*. In Mongolic, there is a historic change of **k >*g* at the boundary of the third and second syllable (Janhunen 2017: 100-101). PMo. **dorogon* also seems to be a representative of this change, with the original voiceless velar still attested in Tungusic. This may be an example of very early contact between Mongolic and Tungusic.

As for PAm. **tʰorq,* it is clear that it did not begin with a lenis stop nor end in a nasal, as would be expected when it bore relation to PTg. **dǫrǫkǫǫn*. Savel'jeva & Taksami (RNS: 60) list both AN *tʰorq* 'badger' and *ɣorq* 'id.', which shows that the initial obstruent was undoubtedly fortis. This is further corroborated by Sangi and Gašilova (RNS-S: 60), who list ESN *tʰoɣq* and which also lacks a final nasal that would normally be retained in ESN. Consequently, the evidence clearly points to PAm. **tʰorq*, which differs from PTg. **dǫrǫkǫǫn* in the two aforementioned respects. They are nevertheless quite similar, yet positing that they would be identical at an earlier point in time is unfounded. It should therefore be removed from the list of plausible parallels.

0047. ***tʰoχ*** ‘moose’[22]

PAm.: **tʰoχ* ‘moose’ (Grube 1892: 90; CND: 150)
PTg.: **tọọ-kII* ‘moose’ (SSTM2: 191-192)
Proposed by: Grube (1892: 90), von Möllendorff (1894: 145), Krejnovič (1955a: 162), Bouda (1960: 402), Cincius (SSTM2: 192)

This item is clearly original to Tungusic as it is transparently a stem based on the nominal classification suffix *-kII* used to describe animals.

0048. ***uɣŋ*** ‘loon’

PAm.: **uɣŋ~*uɣŋ* ‘loon’ (CND: 155)
PTg.: **ookəəŋi* ‘loon’ (SSTM2: 257)
Proposed by: Burykin (1987: 48)

Both PAm. **uɣŋ~*uɣŋ* and PTg. **ookəəŋi* are likely to have been identical historically. Neither is segmentable, which leaves little information to base a direction of borrowing on.

0049. ***va-ɟ*** ‘to fight’

PAm.: **wa-* (CND: 158)
PTg.: **waa-* ‘to kill’ (SSTM1: 127-129)
Proposed by: Krejnovič (1955a: 165), Bouda (1960: 400)

Although phonologically similar, these may not be the same word. Firstly, it is highly possible that PAm. **wa-* ‘to fight’ was originally **U-a-*, with the reciprocal prefix. This is to be expected since this verb is lexically reciprocal as well – it requires a plural subject or a patient in the allative or comitative case. It is therefore functionally intransitive (Otaina & Nedjalkov 2007: 1737-1738). In contrast, PTg. **waa-* ‘to kill’ is transitive. On that basis, I find it doubtful that these words share a common origin and therefore have elected to dismiss it.

[22] I have elected to use the translation ‘moose’, as it is less ambiguous than ‘elk’.

0050. ***χoŋ*** ‘sheep’

PAm.: **χoŋ* ‘sheep’ (CND: 165)

PTg.: **kọnịn* ‘sheep’ (SSTM1: 409-410)

Proposed by: Krejnovič (1955a: 165), Bouda (1960: 404), Panfilov (1973: 8), Fortescue (CND: 165)

PTg. **kọnịn* ‘sheep’ is reflected as *χonị(n)* ‘id.’ in the Nanaic languages, which corresponds to Ma. *χonin* ‘id.’ and Ewk. *konin* ‘id.’ These correspondences are irregular, since Na., Uil., Ulc. /x/ should correspond to Ewk., Ma. ∅ (Benzing 1956: 42-43). Therefore, it is more likely that it is a loanword in the Nanaic languages and Manchu is the most likely donor, as suggested by Doerfer (1985a: 37) as well. He has also connected it to Mo. (*)*konin*, which has a Turkic origin. Animal husbandry relying on sheep, just as horse-riding, is a relatively recent development in the part of Asia where these languages are spoken (Robbeets 2015: 21, 32). Therefore, it is highly possible that Amuric borrowed (*)*χoŋ* from a Tungusic language as they did with their word for ‘horse’ (cf. 0028). Additionally, as fricative-initial nouns were illicit earlier in the history of Amuric, it is almost certain that this item is of Tungusic origin and moreover a recent borrowing as earlier loans were adapted to this older constraint (cf. §7). Since no reflex of **ŋ* is attested in neither AN nor SSN (cf. CND: 165), it would seem this word was also subject to borrowing within Amuric, which even more so suggests that this item is of recent age. Even in Tungusic the word for ‘sheep’ may be of recent introduction as there is also a possibility that Ewenki and Manchu borrowed it separately from Mongolic. In sum, it cannot be regarded as part of the oldest layer of borrowing.