

# RECONSTRUCTIO EXTERNA LINGVAE GHILIACORUM

*Juha Janhunen*

*University of Helsinki*

Following the line of research initiated by Robert Austerlitz, this paper takes up the question concerning the prehistory of the Ghilyak or Nivkh language. Since Ghilyak is a genetic isolate, or a small family of closely related idioms, information on the earlier stages of the language can be obtained only from two kinds of sources, internal and external. Firstly, Ghilyak exhibits a large number of systematic morphophonological alternations and derivational correlations, which offer good opportunities for the application of the method of internal reconstruction. At the same time, due to external contacts, Ghilyak shares a number of diachronic developments, grammatical features, and lexical elements with the neighbouring languages. This paper, then, discusses some of the information that the traces of external contacts can yield about the prehistory of the Ghilyak language.

В продолжение исследований, начатых Робертом Аустерлицем, в настоящей статье рассматривается вопрос о предьстории нивхского языка. Поскольку нивхский язык является генетически изолированным, или представляет собой небольшую семью близкородственных языков, информацию о более ранних стадиях его развития можно получить только из двух типов источников – внутренних и внешних. С одной стороны, для нивхского языка характерны морфофонологические чередования и деривационные корреляции, которые предоставляют возможности для применения метода внутренней реконструкции. С другой стороны, в результате внешних контактов нивхский язык разделяет ряд исторических процессов, грамматических черт и лексических элементов с соседними языками. В статье обобщена информация о том, каким образом эти явления отражают предьсторию нивхского языка.

## 1. INTRODUCTION

In a lecture given in Helsinki, April 1972, under the title “Reconstructio interna linguae Ghiliacorum”, Robert Austerlitz presented an outline of the possibilities of internal reconstruction as applied to the Ghilyak or Nivkh language (Austerlitz 1972). Although essentially a genetic isolate with relatively little internal variation, Ghilyak exhibits synchronically a large number of systematic morphophonological alternations and derivational correlations, which allow conclusions to be made about the earlier diachronic stages of the language. Austerlitz subsequently

*Studia Orientalia* 117 (2016), pp. 3–27

Licensed under Creative Commons  
Attribution 3.0 License.  
ISSN: 0039-3282

continued this line of research in a number of more detailed papers (Austerlitz 1980; 1982; 1984a; 1984b; 1990; 1994). As a result, it is today possible to reconstruct the sound system and several details of root structure and grammar for a prehistorical stage of the language that may be termed Pre-Proto-Ghilyak. Chronologically Pre-Proto-Ghilyak represents a level that once preceded Proto-Ghilyak, which, by definition, may be understood as the relatively shallow proto-form of the modern and historically attested regional varieties of Ghilyak.<sup>1</sup>

The present paper takes up the issue concerning the reconstruction of Pre-Proto-Ghilyak from the point of view of the external relations of the language. As a starting point it may be taken for certain that Ghilyak, in spite of several attempts, has so far not been proven to be related to any other known language or language family, a situation suggesting that it has no surviving relatives. It does, however, share both lexical elements and structural features with the neighbouring languages, which represent several different language families, including, in particular, Ainuic (Ainu) and Tungusic, but more distantly also Mongolic, Koreanic, Japonic, and Kamchukotic (Chukchee-Kamchadal). The largest number of lexical parallels links Ghilyak with Tungusic, and these parallels exhibit regular phonetic correspondences, which in many respects confirm the conclusions made on the basis of internal reconstruction. The external evidence also allows more definitive conclusions to be drawn about the absolute chronology of some of the developments between Pre-Proto-Ghilyak and Proto-Ghilyak.

Since no comprehensive study has ever been carried out concerning the dialectal variation of Ghilyak, the internal taxonomy of the language is still open to alternative interpretations. However, the basic division into the Amur and Sakhalin varieties is probably taxonomically relevant, though the terminology is geographically misleading in that the so-called Amur dialect extends from the Amur Delta region on the continent to northwestern Sakhalin, covering both sides

---

1 In the present paper the “obsolete” term Ghilyak (Ghiliak, Gilyak) is used for the language which in the Soviet Union was renamed “Nivkh”. Although today considered to be the politically correct appellation of the language and its speakers, the term Nivkh involves several problems. For one thing, it is based on only one of the two major dialectal forms of the language, the Amur variety, while its counterpart in the other major dialect, the (East) Sakhalin variety, would be better rendered as *Nighbng*. As the mutual comprehensibility of the two varieties is limited, the speakers do not always consider themselves as belonging to a single homogeneous ethnic group, and there is occasionally some reluctance to use the term “Nivkh” among the speakers of the Sakhalin dialect. Another problem is that the term “Nivkh” is not an unambiguous ethnonym, but, rather, the term for ‘human being’ in the Ghilyak language. These problems are avoided when we use the “obsolete” term Ghilyak (Ghiliak, Gilyak), which is a true generic ethnonym covering all Ghilyak speakers. This term is ultimately derived via Russian from the Tungusic languages spoken to the west and south of Ghilyak, and it was never used by the Ghilyak speakers themselves. Moreover, it has no inherent negative or derogatory connotation.

of the Amur Liman. The so-called Sakhalin dialect, on the other hand, comprises two or three local varieties, including the dominant East Sakhalin dialect and the marginal (and today virtually extinct) South Sakhalin and North Sakhalin (Schmidt Peninsula) varieties. It is generally assumed that the South Sakhalin dialect represents in some respects the most “archaic” variety of Ghilyak, though this remains to be demonstrated with factual evidence. The North Sakhalin dialect, on the other hand, is often considered to represent a transition between the Amur and Sakhalin major varieties, though recent information suggests that its position may be more complicated and perhaps not easily describable in terms of a binary family tree (Gruzdeva, pers. comm. 2014–2015).

Depending on what point of view is taken concerning the degree and relevance of internal differentiation within Ghilyak, we are, consequently, dealing either with a single isolate language with relatively significant dialectal variation or with a small and shallow family with two or more closely related languages. The relationships between the varieties can be approached with the standard comparative method. However, although no systematic comparative analysis of the Ghilyak regional varieties has been carried out, it is obvious that the conclusions that can be drawn from this approach are relatively trivial compared with what can be said on the basis of internal reconstruction, on the one hand, and external evidence, on the other. To place Ghilyak in a wider chronological and geographical context it is therefore particularly important to take a look at its relationships with the neighbouring languages. This is not to say that a thorough comparative analysis of the synchronic interdialectal correspondences within Ghilyak could not give valuable and interesting information.

In a wider ethnohistorical context, Ghilyak may be classified as a language of the Greater Manchurian region. Although it today occupies only a marginal position in this region, there are many reasons to assume that it represents the last surviving trace of a language family that may have had more members in the past, and that, in any case, occupied a geographically more central position in pre- and protohistorical Manchuria. Since the trajectory of movement of this language family must have been the Sungari-Amur basin, it may conveniently be referred to by the term “Amuric” (Janhunen 1996: 73–79). In the present paper, Ghilyak is, therefore classified as a member of the Amuric language family. Although no other Amuric languages are documented in the extant historical sources, it may be speculated that some of the protohistorical ethnic groups recorded from the region were actually Amuric speaking.

Assuming that the Amuric language family originated in central Manchuria, the modern Ghilyak speakers may be seen as the physical descendants of several local aboriginal populations that became Amuric speaking due to the expansion

of the Amuric family northwards along the Amur and ultimately to Sakhalin. There is also a possibility that the Ghilyak language area once extended down to northern Hokkaido in connection with the expansion of the mediaeval Okhotsk Culture (Vovin 2016, in the present volume). The historically documented Ghilyak speakers were culturally relatively “primitive” semi-settled fishers and sea mammal hunters, but the possibility that an ancestral form of the Ghilyak language was once also used by some population with a “higher” culture is suggested, among other things, by the fact that the language has a layer of cultural vocabulary of its own, not shared with the neighbouring languages. Perhaps most importantly, Ghilyak has a number of apparently indigenous lexical items related to metallurgy, such as, for instance, *doto* ‘silver’, *dews* ‘copper’, *tac* ‘tin’ (NRS 360, 370, 380), suggesting that its speakers once had some independent knowledge concerning the production and handling of metals.<sup>2</sup> Such activity is likely to have taken place in some more favourable natural and cultural environment than that offered by the present-day habitat of the Ghilyak speakers.

No exhaustive study has been made of the lexical connections of Ghilyak with the neighbouring languages. Important preliminary work on the contacts with Tungusic has, however, been carried out by E.A. Kreinovich (1955) and V.Z. Panfilov (1973), as well as by Gerhard Doerfer (1985 *passim*) and Alexandr Pevnov (2016, in the present volume). The history of individual etymons has also been dealt with in a number of specialized papers (e.g. Austerlitz 1976; 1989; Janhunen 2008; 2014). Although far from complete, this material is sufficient to illustrate some of the basic aspects of the external reconstruction of Ghilyak, as discussed below in the light of selected lexical parallels. As for terminology, I prefer to make a rather clear distinction between (1) the *comparative method* (based on synchronic data from two or more related idioms), (2) *internal reconstruction* (based on synchronic data from a single idiom), and (3) *external reconstruction* (based on data from two or more unrelated idioms).<sup>3</sup> In the present paper, the focus is on external reconstruction.

---

2 Note that on this issue, Austerlitz (1983; 1984) had a somewhat different opinion. Basically, he tried to identify either external sources or internal explanations for most of the Ghilyak metal names. However, although Ghilyak has borrowed names for some “secondary” metals, like *ka* ‘steel’ (ultimately from Chinese), most of the other external and internal explanations proposed by Austerlitz in this connection are problematic. Most importantly, even in cases when Ghilyak unambiguously shares a metal name with its continental neighbours, especially Jurchen-Manchu, it is not at all clear which language was the donor. In many cases, we may actually be dealing with Amuric loanwords in Pre-Proto-Jurchenic.

3 Note again that Austerlitz had a different view on how these terms should be defined. Although his work on the internal reconstruction of Ghilyak is generally based on the internal information provided by the synchronic data of the language, he felt free to include information from both

## 2. GHILYAK PHONOLOGY: INTERNAL INFORMATION

All varieties of modern Ghilyak have a relatively simple symmetric vowel system comprising 6 distinct segments. The vowels are divided into rounded back, unrounded front, and unrounded back segments with two levels of opening (Table 1). It may be immediately noted that the vowel system has an exact parallel in the neighbouring Tungusic languages of the Nanai group, notably Nanai proper and Ulcha (cf. e.g. Avrorin 1959: 18–21). In these languages, the vowels are arranged in vertically organized harmonic pairs, and the same may be assumed to be true of Ghilyak, although vowel harmony in Ghilyak is not as transparent as in Tungusic. Even so, the high unrounded back vowel [ə] ~ [ɨ] (in linguistic literature on Ghilyak often rendered as <y>, Cyrillic <Ы>, but here written *e* < \*e) may be viewed as the harmonic counterpart of the low vowel *a* (< \*a), while the low unrounded front vowel [e] (here written *é* < \*i) is the harmonic counterpart of the high vowel *i* (< \*i). There are small phonetic differences between the individual languages of the region, but the system is, in principle, the same.

Table 1 The vowel system of Ghilyak

<i>u</i>	<i>e</i>	<i>i</i>
<i>o</i>	<i>a</i>	<i>é</i>

From the areal point of view it is interesting to note that Orok, or Uilta, the Tungusic language that is historically most closely associated with Ghilyak on Sakhalin, has an additional rounded central vowel [ə], a trace of Proto-Tungusic \**ö* (Ikegami 1953), which is otherwise preserved only in Ewen (Doerfer 1978). This suggests that the Ghilyak vowel system developed in closer contact with the Tungusic languages spoken on the continent, while Orok is likely to represent a secondary intrusion into the Ghilyak linguistic environment. Ultimately, it is unclear how Orok, in all other respects a close but innovative relative of Ulcha

---

dialects and loanwords. According to him: “Internal reconstruction [...] means employing all possible resources — the meaning of a word, its phonological and grammatical shape, the culture of its speakers, and that of the speakers of surrounding languages — to permit us to draw conclusions about earlier stages of the language, that is, about the shape and the meaning of its words and grammatical constructions. Information from other dialects of the language is very useful. I make free use of it. (In some definitions of internal reconstruction, dialect information is not freely used. This is a matter of definition and since the position chosen for my work is clear, need not detain us.) In the practice of internal reconstruction loanwords are of supreme importance, when they can be identified on the formal and on the contentual side. The benefits which accrue from a clearly identified loanword and the direction in which it travelled are many: chronology, earlier stages of the two cultures and of the two languages in question, historical phonology, and progress in methodology.” (Austerlitz 1989: 3–6)

and Nanai, can alone have preserved a Proto-Tungusic vowel that is absent in the other Tungusic languages of the region.

By contrast, the synchronic consonant system of Ghilyak is very different from that of the neighbouring Tungusic languages. While all the Tungusic languages have typically rather few consonant phonemes, organized in a loose system with many gaps, Ghilyak has an exceptionally rich and compact system of some 23 to 34 distinct consonant phonemes, with the number depending on how the actual contrasts work in each given variety of the language. The maximal system comprises 3 sets of stop obstruents, 2 sets of continuant obstruents, 1 set of nasals, and 1 set of non-nasal non-obstruent (sonorant) continuants, divided between 4 or 5 different places of articulation. It may be noted that only the obstruents can exhibit the maximum number of 5 places of articulation, while the nasals and non-obstruents have only 4 (Table 2).

Table 2 The maximal synchronic consonant system of Ghilyak

<i>m</i>	<i>n</i>	<i>ny</i>	<i>ng</i>	
<i>bb</i>	<i>dd</i>	<i>jj</i>	<i>gg</i>	<i>qq</i>
<i>b</i>	<i>d</i>	<i>j</i>	<i>g</i>	<i>qb</i>
<i>p</i>	<i>t</i>	<i>c</i>	<i>k</i>	<i>q</i>
<i>f</i>	<i>rh</i>	<i>s</i>	<i>x</i>	<i>xh</i>
<i>v</i>	<i>r</i>	<i>z</i>	<i>gh</i>	<i>hh</i>
<i>w</i>	<i>l</i>	<i>y</i>	<i>h</i>	

The only language in the region that comes close to Ghilyak in the complexity of the consonant system is Korean, but even Korean has less phonemes and its system is not as compactly arranged as that of Ghilyak. The closest parallel between Ghilyak and Korean involves the stop systems, which in both languages have three sets. The phonetic realizations of these sets are, however, different in the two languages. In Ghilyak, the stops may be characterized as weak or basic (*b d j g qb*), strong or aspirated (*p t c k q*), and distinctively voiced (here written *bb dd jj gg qq*), while in Korean the weak or basic and the strong or aspirated sets contrast with the so-called reinforced or glottalized set, which is unvoiced and unaspirated. Even so, it cannot be immediately ruled out that the similarity of the Ghilyak and Korean stop systems could have an areal background.<sup>4</sup>

<sup>4</sup> Note that the transcription used in the present paper is in some respects idiosyncratic and differs from the transcriptions used, for instance, by Austerlitz, and also from the conventions current in international Nivkh studies today (cf. e.g. Gruzdeva 1998: 10; Mattissen & Drossard 1998: 6–7). One reason for using a “new” system of transcription for Ghilyak is to make the no-

Even a simple synchronic analysis allows, however, the Ghilyak consonant system to be reduced somewhat, in that the voiced stops (*bb dd jj gg qq*), though synchronically phonemic, tend to occur after nasals (including deleted nasals) and may be analysed as morphophonologically secondary. Even more importantly, the back velars or uvulars (here written *qq qh q xh hh*) are in a virtually complementary distribution with the corresponding regular or front velars (*gg g kx gh*). The back velars generally occur in combination with the low vowels *o é a*, while the front velars occur in combination with the high vowels *u i e* (cf. e.g. Austerlitz 1982: 82–83, Table IV), suggesting that the vertical harmonic pairs are actually verticalized back-front pairs. Minimal pairs between the front and back velars are synchronically possible in syllable-final position, as in *ask* ‘spider’ vs. *asq* ‘younger sibling’ (Austerlitz 1982: 82), but the functional load is minimal, and even phonetically the two series are often confused by the speakers. The main reason why the back velars are today normally listed as full members of the consonant paradigm may be the fact that they are incorporated into the official Cyrillic orthographies of the language. It has to be noted, however, that a diachronically secondary set of back velars is also present in some other languages of the region, notably Manchu (cf. Kiyose 1996).

If we ignore the voiced and back velar sets, we get a minimal synchronic consonant system, which in fact may be valid for some modern speakers of Ghilyak and which is certainly closer to the diachronic origin of the system (Table 3). This system is even more compact, in that it has 6 sets of segments organized according to four distinct places of articulation: labial, dental, palatal and velar. The 6 sets (horizontal rows in the table) may be defined as nasals (*m n ny ng*), weak or basic stops (*b d j g*), strong or aspirated stops (*p t c k*), strong or voiceless fricatives (*frh s x*), weak or voiced fricatives (*v r z gh*), and non-obstruent continuants (*w l y h*). The last set is phonetically somewhat heterogeneous and comprises glides (*w y* and possibly *h*), a lateral (*l*), and a laryngeal (*h*), all of which (with the possible exception of the laryngeal) could also be defined as sonorants. The phonemic status of the glides (*w y*) with regard to the high vowels (*u i*) is open to alternative analyses, but from the systemic point of view it is reasonable to classify the glides as distinct members of the consonant paradigm.

---

tation more compatible with the traditions used for the neighbouring language families, notably Ainu(ic) and Tungusic. This is especially relevant when we are dealing with lexical parallels, in which Ghilyak sounds systematically correspond to similar sounds in other languages. Using a universal phonetic transcription system (like the IPA) would not serve the purpose, since it would tend to obscure the phonemically relevant features behind a curtain of irrelevant phonetic details.

Table 3 The minimal synchronic consonant system of Ghilyak

<i>m</i>	<i>n</i>	<i>ny</i>	<i>ng</i>
<i>b</i>	<i>d</i>	<i>j</i>	<i>g</i>
<i>p</i>	<i>t</i>	<i>c</i>	<i>q</i>
<i>f</i>	<i>rh</i>	<i>s</i>	<i>x</i>
<i>v</i>	<i>r</i>	<i>z</i>	<i>gb</i>
<i>w</i>	<i>l</i>	<i>y</i>	<i>h</i>

The core part of the Ghilyak consonant system is formed by the obstruents, which even in the reduced system comprise 16 distinct segments. These are also the main source of information for the internal reconstruction of Ghilyak, in that the stops and fricatives stand in a systematic morphophonological relationship, in which the strong stops (*p t c k*) alternate paradigmatically with the strong fricatives (*frh s x*), while the weak stops (*b d j g*) alternate with the weak fricatives (*v r z gb*) (Kreinovich 1937: 36–50). It may be noted that the dental fricatives are phonetically manifested as rhotics (*rh r*), of which the strong member (*rh*) is realized as the universally rare sound of voiceless fricative trill, which positionally, dialectally and/or idiolectally can be reduced to a voiceless palatoalveolar or retroflex sibilant of the type [ʃ] ~ [ʂ]. The palatal fricatives, on the other hand, are manifested as more or less regular dental sibilants (voiceless vs. voiced).

The usage of the different sets of obstruents is regulated by lexical and morphological factors which are no longer synchronically transparent at the phonological surface level. The stops and fricatives can therefore synchronically contrast in different positions. However, due to the regularity of the alternations it is possible to postulate a system of morphophonemes in which Ghilyak has only two series of obstruents (strong vs. weak), which, then, are realized as either stops or fricatives depending on the rules. In this system of morphophonemes Ghilyak has only 16 distinct consonant segments. With the method of internal reconstruction we might therefore assume that Pre-Proto-Ghilyak had a system with these 16 consonants (Table 4).

Table 4 The consonant morphophonemes of Ghilyak

<i>m</i>	<i>n</i>	<i>ny</i>	<i>ng</i>
<i>p</i>	<i>t</i>	<i>c</i>	<i>q</i>
<i>b</i>	<i>d</i>	<i>j</i>	<i>g</i>
<i>w</i>	<i>l</i>	<i>y</i>	<i>h</i>

It is another matter how likely and “natural” such a system with 16 consonants would have been. At first glance, it would appear particularly “unnatural” that there is no distinctive set of strong continuant obstruents. Even so, we can immediately see that at least some parts of this system have parallels in the neighbouring languages. For instance, the subsystem of stop obstruents with two sets, strong (*p t c k*) vs. weak (*b d j g*), is well attested in Tungusic, and from the systemic point of view it is irrelevant whether their phonetic distinction was originally based on aspiration (in the strong set) or voice (in the weak set), as both types are synchronically attested in the region. It is therefore not necessary to reduce the system further by eliminating one of the stop series, though such a possibility has been proposed with reference to other neighbouring languages, notably Ainu and Japanese, which have an even less complicated original system of consonants (Austerlitz 1990: 22–25). To gain more light on this and other issues we may now take a look at the available external information.

### 3. GHILYAK PHONOLOGY: EXTERNAL INFORMATION

The diachronically relevant lexical parallels between Ghilyak and the neighbouring languages are the result of loan contacts that took place at the Pre-Proto-Ghilyak stage. From the point of view of the external reconstruction of Ghilyak it is irrelevant in which direction the loans were transmitted, but there are reasons to assume that lexical borrowing took place both from Pre-Proto-Ghilyak into the other languages of the region and vice versa. In many cases, the direction of borrowing is impossible to establish with certainty. It is also relevant to note that the contacts underlying the lexical parallels did not necessarily, and in many cases certainly did not, take place in the current and historically observed habitat of Ghilyak speakers. Rather, they took place closer to the original area of Amuric further to the south in central Manchuria.

However, although the circumstances of borrowing remain in many cases obscure, the lexical parallels show a number of regular sound correspondences that make it possible to verify several developments in the Ghilyak sound system for which otherwise only internal information would be available. Such developments include, among others, those listed below.

#### *Vowel loss in non-initial syllables*

There are ample examples of lexical items in which Ghilyak shows a root-final consonant or consonant cluster, while an areal cognate in another language has a final vowel. Tungusic, for instance, has predominantly bisyllabic stems ending

in a vowel and with only intervocalic consonant clusters (CVC(C)V), while the Ghilyak counterparts of such items have no final vowel (1) and can have clusters in final position (2). In such cases, it has to be assumed that the Pre-Proto-Ghilyak form had a vowel which was lost before the stage of Proto-Ghilyak. Depending on the rules of consonant phonotactics some items have even lost two vowels in consecutive syllables (3).

(1) Manchu *hala* ‘clan, lineage’ < Proto-Tungusic *\*kala* = Pre-Proto-Ghilyak *\*kala-ng* > modern Ghilyak *kal* (*qal*), Sakhalin Ghilyak *kalng* (*qalng*) (NRS 147) id. This is typically an item for which the direction of borrowing is difficult to determine, though there is a possibility that the Ghilyak item is actually based on the more simple Ghilyak root *ka* (*qa*) < *\*ka* ‘name’ (NRS 146), in which case the word would have to be of Amuric origin. On the Tungusic side the word is attested in all branches of the family (SSTM I: 459–460). Importantly, the word is not of Mongolic origin, although it is present as a recent borrowing in a few Manchurian Mongolic idioms (Dagur and Khamnigan). The modern Ghilyak item seems to be an independent secondary derivative in *-ng*, though this type of final nasal has a potential counterpart in Tungusic and Mongolic (as discussed further below). It may be noted that an etymologically identical element *-kal* also occurs in Ghilyak in the ethnonym *kékal* (*kéqal*) ‘Kyakala’ = ‘Udeghe’ (NRS 130). The latter is, however, a secondary borrowing from Tungusic, where it derives from the compound *\*kīra+kala* ‘border clan’ (Janhunen et al. 1999: 249).

(2) Ewenic (*\*laamus* ‘(warm) wind (from the sea)’ < Proto-Tungusic *\*laamos* ← Pre-Proto-Ghilyak *\*lamos* > modern Ghilyak *lams* ‘eastern wind’ (NRS 155) This is verifiably an Amuric loanword in Tungusic, for it is based on the Amuric simple root *\*la* ‘wind’ > modern Ghilyak *la* id., also: ‘the river Amur’. Moreover, it seems that a large part of the maritime terminology of Tungusic derives from Amuric, suggesting that Amuric speakers once had a more immediate access to the sea than Tungusic speakers, or that they were more maritime-oriented (Doerfer 1985: 261). The form *\*la-mo-s* is a second derivative based on *\*la-mo*, a form no longer synchronically attested in Ghilyak but transmitted into Tungusic in the meaning ‘sea, ocean’ and well attested in all Tungusic languages (SSTM I: 490–491; Janhunen 2008: 97–100).<sup>5</sup> It may be concluded that the original bisyllabic structure of the word is preserved on the Tungusic side, while in Ghilyak the vowel of the non-initial syllable was lost rendering the word

5 Pevnov (2016, in the present volume) quotes Orok *lamu* ‘small wave’, which he derives from Ghilyak (Poronaisk) *lam* id. This might be another reflex of Pre-Proto-Ghilyak *\*la-mo*, and it may be noted that here, also, the original final vowel of the word is preserved on the Tungusic side but lost in Ghilyak. Orok *lamu* must, however, be a separate, local and relatively recent loanword from Ghilyak, since Proto-Tungusic *\*laamo* ‘sea’ is represented in Orok as *namu*. On the other hand, Pevnov (2016, in the present volume) suggests that Ghilyak *la* in reference to ‘the river Amur’ could be a Tungusic loanword, cf. Manchu *ula* ‘(large) river’. However, although it is possible that the two meanings of Ghilyak *la* represent separate etymons, this is not particularly likely in view of the overall semantic coherence of the derivational set *la* : *la-mo* : *la-mo-s*.

synchronically monosyllabic and with a final cluster. (It may also be noted that the Tungusic item would seem to contain an original long vowel. The question concerning the nature and origin of vowel length in Tungusic is, however, complicated and will not be discussed here.)

(3) Proto-Mongolic *\*kalimu* ‘whale’ ← Proto-Tungusic *\*kalimV* id. ← Pre-Proto-Ghilyak *\*kalimV* id. > modern Ghilyak *kalm* (*qalm*) ‘(small) whale’ (NRS 147). This is another maritime term that would seem to have entered Tungusic from Amuric. From Tungusic the word proceeded into Mongolic, which today is a continental language family with no direct connection to the sea. In a larger context the word has parallels also elsewhere in Central and North Asia (Janhunen 2012). The final vowel is ambiguous in the data, but, in any case, Tungusic and Mongolic suggest that the word was originally trisyllabic, and that the vowels of the second and third syllables were both lost in Ghilyak.

Vowel loss itself is, of course, a trivial process attested in many languages in the region, including, for instance, Pre-Proto-Turkic, Pre-Proto-Korean(ic), modern Mongolian, and several forms of Tungusic. It may also be shown to have been active in Pre-Proto-Ainu (Janhunen 2015). However, it is not observed in the geographically closest Tungusic (Amur Tungusic) neighbours of Ghilyak, which is why it is possible that the loss of vowels in Pre-Proto-Ghilyak was a separate process, with no direct areal connection with the other languages in the region. In any case, vowel loss also explains some internal properties of modern Ghilyak. For instance, the contrast of front and back velars in cases like *ask* ‘spider’ vs. *asq* ‘younger sibling’ (as mentioned above) may be explained as being due to the loss of different final vowels, which originally conditioned the type of velar consonant, that is, *ask* < *\*as(+)kE* vs. *asq* < *\*as(+)kA*, where *\*E* stands for any original front/high vowel, while *\*A* stands for any back/low vowel. Note that the rules of vowel harmony (as discussed further below) suggest that words containing both back/low vowels and front velars, or vice versa, may originally have been compounds of two harmonically different components.

It is also important to note that the loss of final vowels in Pre-Proto-Ghilyak was a diachronic process that had a limited extension in time. Loanwords received after the process became no longer active retain final vowels (4). There are also many native words with final vowels; these require, again, other explanations, including the possibility of compound origin.

(4) Tungusic (*\*luuca*) > *luca* ~ *loca* ‘Russian’ → Ghilyak *loca* ~ *loji* id. (NRS 164). The two variants in Ghilyak are likely to represent two separate instances of borrowing, and the immediate source must have been Amur Tungusic (SSTM I: 513–514), though it is difficult to identify the exact donor language(s). The ethnonym has travelled a long way from mediaeval Scandinavia through Siberia

to Manchuria (Old Swedish → Finnic → Nenets → Ewenki → Amur Tungusic and Manchu, cf. Janhunen 1997), and it may have reached the Ghilyak only with the arrival of the Russians on the Lower Amur and Sakhalin in the eighteenth to nineteenth centuries.

### *Vowel loss in the initial syllable*

Ghilyak also has initial clusters of two segments (#CC). Since this is a feature alien to all immediate neighbours of Ghilyak, most of which represent the Ural-Altai language type with very simple phonotactics, it may be assumed that the Ghilyak initial clusters are secondary (Austerlitz 1990: 22–29). In fact, secondary clusters were also formed in Korean, where they subsequently developed into a set of glottalized consonants. The possibility that the Ghilyak initial clusters were formed by the loss of the original vowel of the initial syllable is suggested by both internal (5) and external evidence (6). The exact conditions of this development are, however, unknown, though it is clear that they must be connected with the rules of consonant phonotactics (determining the types of permitted initial clusters), as well as, possibly, morphological and prosodic factors (which may have determined which vowels could be dropped).

(5) Modern Ghilyak *nyrhank* = *ny-rhank* (*ny-rhank*) ‘(one) hundred’ (NRS 218) < \**nyV-tangku* = Manchu *tangku* id. In this etymon, the Manchu numeral *tangku*, also present in Amur Tungusic and Manchurian Ewenic (SSTM II: 163), is represented as *-rhank* in Ghilyak, but the initial fricative rhotic goes back to the strong dental \**t* by the regular process of rhoticization. This process (as discussed further below) was originally active in intervocalic position, which means that the members of the initial cluster were once separated by a vowel that was subsequently lost. In fact, internal information confirms that the element *ny-* represents the numeral ‘one’ and derives from earlier \**ni-* : \**nyé-*, as also attested in many forms containing numeral classifiers, as in *nyé-c* ‘one (board)’, *nyé-x* (*nyé-xh*) ‘one (long object)’, *ni-k* ‘one (small round object)’, *ni-rh* ‘one (sledge)’, etc. (NRS 208–211; cf. also Gruzdeva 2004 *passim*). The Pre-Proto-Ghilyak form for ‘(one) hundred’ may, consequently, be reconstructed as \**nyé+tangku* < \**ni+tangku*. The vowel of the initial syllable is determined regressively by the rules of vowel harmony. (The alternation between archiphonemic *n* [n ~ ɲ] and distinctively palatal *ny* [ɲ] is also determined by complementary factors.)

(6) Sakhalin Ainu *tunakay* (→ Japanese *tonakai*) ‘reindeer’ < \**tunankay* ← Pre-Proto-Ghilyak \**tVla+nga-y* > modern Ghilyak *tlangi* id. (NRS 451). This is an important lexical contact from both the cultural and the phonetic point of view. The item is originally Ghilyak, since it can be explained as a compound containing the elements (\**tle-* ‘to draw’ and (\**nga* ‘(wild) animal’, augmented by a further derivative component that may be tentatively reconstructed as \**y* (Austerlitz 1976). The same element *-ngi* is attested in other animal names, as in *kangi* (*qangi*)

‘navaga’ (fish of the family *Gadidae*) (NRS 141). The possibility that the Ghilyak word for ‘reindeer’ and, consequently, also the root for ‘to draw’, contained originally a vowel between the segments today forming the initial cluster *\*tl*, is suggested by the presence of a vowel in the Ainu shape (from which the item was further transmitted into Japanese). It is, of course, also theoretically possible that the vowel in the Ainu data is simply an addition required by Ainu phonotactics, but this appears less likely. The substitution of Ainu *n* for Ghilyak *l* was conditioned by the differences in the consonant systems. A very similar relationship between Ghilyak and Ainu data is exhibited by Ghilyak *vongi* ‘kabarga’ (Siberian musk deer, *Moschus moschiferus*) (NRS 56) = Ainu *opokay* id. In this case, the original shape of the word must have been *\*obo+nga-y*, with a later loss of the initial vowel in Ghilyak (Austerlitz 1986: 1).

### *The velarization of the final unmarked nasal*

Many Ghilyak nouns end in the velar nasal *ng* [ŋ], which is normally preserved in the Sakhalin dialect but segmentally lost in the Amur dialect at the surface level, though even for the latter it can still be postulated at the deep level because of its impact on the morpheme-boundary segmental alternations (Mattissen 2003: 41–43). Because of its dialectal and morphophonological instability we may call this segment the “weak” (Gruzdeva 1997: 84) or “unstable” nasal and express it notationally by */ng*. In some cases, as in modern Ghilyak *kal/ng* (1), the unstable nasal seems to represent a derivative suffix added secondarily in Pre-Proto-Ghilyak, but in many other cases it corresponds to a similar unstable nasal in Tungusic and/or Mongolic. In the latter two families, this segment is diachronically represented by an unmarked dental nasal *\*n*, though in many varieties of especially Mongolic it has been velarized to *ng*. In view of the correspondence of Ghilyak */ng* to Tungusic and Mongolic *\*n* we may assume that a similar velarization development *\*-n > ng* took place in Pre-Proto-Ghilyak. Some of the items concerned are clearly continental loanwords in Ghilyak (7), while for others the direction of borrowing is impossible to determine (8).

(7) Mongolic *\*morī/n* ~ *\*muri/n* ‘horse’ → Tungusic *\*muri/n* > *\*muri/n* → Pre-Proto-Ghilyak *\*murin* > modern Ghilyak *mur/ng* = Sakhalin *murng*, Amur *mur* (NRS 197). This is a well-known and widespread Central and East Asian cultural item that is present also in Korean (*\*morV*), Chinese (*\*mra*) and Japanese (*uma*) (Janhunen 1998). It is very likely that the word was borrowed into Ghilyak via Tungusic, where the vowel of the initial syllable is *\*u*, originally borrowed from an aberrant form of Mongolic with the development *\*o > \*u*. Since the *\*u* is still preserved in Ghilyak and has not participated in the regular vowel rotation process (as discussed further below), the borrowing may be relatively recent, though it must nevertheless date from the Pre-Proto-

Ghilyak period. It is likely that at the time of borrowing the final nasal was still the unmarked *\*n*, which then underwent velarization to /*ng* before the Proto-Ghilyak stage.

(8) Manchu *aisin* ‘gold’ < *\*aysin* = Pre-Proto-Ghilyak *\*aysin* > modern Ghilyak *ays/ng* = Sakhalin *aysng* (secondarily also > *ayzng*), Amur *ays* (NRS 29). In this case the direction of borrowing remains unclear. Although the Tungusic word is attested in this shape not only in Manchu but also in Amur Tungusic and Manchurian Ewenic (SSTM I: 22), it might nevertheless represent an item of metal terminology borrowed from Amuric into Tungusic. In any case, both Tungusic and Ghilyak have a final nasal in this item, and we observe the development *\*n* > *ng* in Ghilyak. It may be noted that there are two other similar, but different, words for ‘gold’ also attested in Tungusic: *\*ancun* (Jurchen †*ancun* ‘gold’ > Manchu ‘earring’) and *\*altan* (SSTM I: 33). The latter is a well-known loanword from Mongolic (Proto-Mongolic *\*altan*) and ultimately connected with Turkic *\*altin* < *\*altun*, allowing the Pre-Proto-Turkic/Pre-Proto-Mongolic reconstruction *\*alton* (Doerfer 1985: 64; cf. also Rozycki 1994: 13). The relationships between the reconstructions *\*aysin*, *\*ancun*, and *\*alton* are irregular, and it remains unclear whether we are dealing with one, two or three separate etymons. However this may be, the Ghilyak-Manchu item represents a distinct case of areal connection.

### *The rhoticization of the dental stops*

As can be concluded from morphophonological evidence, Ghilyak has undergone a rhoticization development in which the original dental stops *\*t* > [tʰ] vs. *\*d* > [t] developed into the corresponding rhotics *rh* vs. *r*. This development was a part of a more general process in which all intervocalic stops became continuants. Due to the subsequent deletion of vowels, these continuants are synchronically observed in all positions of the word (initial, medial pre- and postconsonantal, final). The development can be verified externally by words in which a Ghilyak continuant corresponds to a stop in a neighbouring language, for instance, Tungusic and/or Ainu (9).

(9) Tungusic *\*urangka-ta* ‘Uryangkhai’ → Pre-Proto-Ghilyak *\*urangata* > *\*orangata* (→ Sakhalin Ainu *orakata*) > Proto-Ghilyak *\*orngarh* > Sakhalin *orngarh*, Amur *ornger* ‘Orok, Ulcha’ (NRS 247). The Ghilyak form of this item, which is ultimately connected with the widespread generic ethnonym *Uryangkhai* ~ *Urangkai* (Janhunen 2014), is likely to be based on a Tungusic derivative (a plural or collective) in *\*-tA*. The form with the dental stop was still preserved in Pre-Proto-Ghilyak at the time when the item was transmitted to Sakhalin Ainu, but later it underwent the process *\*t* (*th*) > *rh*, meaning that the rhoticization process took place only after the contact with Ainu had been established. This word is one of many examples in which a Sakhalin Ghilyak strong (unvoiced)

fricative trill *rh* in final position corresponds to an Amur Ghilyak weak (voiced) “regular” trill *r*. This correspondence may have been caused by two opposite processes: devoicing in the Sakhalin variety and voicing in the Amur variety, which is why it is not always easy to reconstruct the type of the original segment (strong vs. weak). However, the fact that the synchronic trills *r rh* can diachronically represent the dental stops *\*d \*t* is clearly evident from the external data.

### *The diachrony of the system of liquids*

Synchronically, Ghilyak has three liquids, a lateral *l* and the two rhotics *r* and *rh*. As a member of the non-obstruent (sonorant) class of consonants, the lateral does not participate in morphophonological alternations and seems to represent the direct trace of a Pre-Proto-Ghilyak lateral. Since several neighbouring languages, notably Koreanic and Japonic (as well as Chinese), also have a single-liquid system we might conclude that Pre-Proto-Ghilyak also had no other liquids except *\*l*. Tungusic and Mongolic have two liquids, the lateral *\*l* and the trill *\*r*, but they have phonotactic restrictions, and the Tungusic languages in the Amur region, including Neghidal (an Ewenic idiom), have, at least in certain positions, tended to eliminate the trill *\*r* either by deleting it altogether (>  $\emptyset$ ) or by merging it with the palatal glide (> *y*) (Pevnov 1994). Ainu is also basically a single-liquid language, though the situation for Proto-Ainu is still being disputed (Vovin 1993: 20–22; Alonso de la Fuente 2012: 17–18). However this may be, for Pre-Proto-Ghilyak we have to reconstruct two liquids, for the trill *\*r* occurs also in external data, in which it does not derive from a dental stop (10).

(10) Manchu *orho* ‘grass, hay, plant’: *orho-da* ‘ginseng’ = Ewenic *oroogto* id. (SSTM II: 24) < Proto-Tungusic *\*oro-g-ta* > *\*orota* → Pre-Proto-Ghilyak *\*orota* > modern Ghilyak *ororb* ‘ginseng’ (NRS 247). This is an unambiguous Tungusic loanword in Ghilyak, for the Tungusic original is a derivative containing the complex class marker *\*-gt-A* as used for individualizable (countable) objects, such as plants (Benzing 1956: 71–72). It may be noted that the class marker appears in Ghilyak in the simplified shape *\*-ta*, which may be due to a secondary development in Pre-Proto-Ghilyak. The original meaning of the word in Tungusic must have been simply ‘grass, plant’, but in Manchu and Amur Tungusic it has also received the specialized meaning ‘ginseng’, in which meaning it was borrowed into Ghilyak. Here we see both an original trill *\*r* > *r*, still preserved in intervocalic position, and a secondary fricative rhotic *\*t* > *rh* in final (originally intervocalic) position. We may, therefore, reconstruct the *\*r* also for Pre-Proto-Ghilyak in general. This means that the modern occurrences of *r* are diachronically opaque, as they may represent both the original trill *\*r* and the original weak stop *\*d*. The two sources of *r* can only be distinguished when external evidence is available. By contrast, the fricative trill *rh* is, in prin-

cept, diachronically transparent, as it always represents the original strong stop *\*t*. Unfortunately, due to the dialectal devoicing of final *r* to *rh*, the fricative trill can, in practice, represent three diachronic sources: *\*t*, *\*d* and *\*r*.

### *The diachrony of the system of palatals*

Palatals form a distinct group of modern Ghilyak consonants comprising a glide (*y*), a nasal (*ny*), three stops realized as palatal affricates (*c j jj*), and two continuants realized as dental sibilants (*sz*). It may be seen that although the sibilants are normally realized as dentals, they function in the system as palatals and are, in fact, in a regular morphophonological relationship to the palatal stops (*s : c vs. z : j : jj*). This could imply that Pre-Proto-Ghilyak had only two segments of this category, which may be reconstructed as a strong and a weak palatal obstruent, *\*c* vs. *\*j* (cf. Austerlitz 1990: 19–20). These can have been phonetically realized both as stops/affricates and as continuant sibilants depending on the context. A similar situation is known from, for instance, modern Chukchee, in which the segment realized as a dental or laminal sibilant [s] ~ [ç] occupies the position of a palatal obstruent /c/ in the system and can also be realized as an alveopalatal or laminal affricate [tʃ] ~ [tʃç] depending on the context (cf. e.g. Skorik 1961: 28). External evidence shows that this may really have been the case in Pre-Proto-Ghilyak, for a Ghilyak palatal stop/affricate can correspond to a continuant sibilant in a neighbouring language (11).

(11) Manchu *saman* ‘shaman’ = Ewenic *\*samaa*n id. (SSTM II: 59) < Proto-Tungusic *\*samaa*n = Pre-Proto-Ghilyak *\*saman* > Proto-Ghilyak *camng* > Sakhalin *cam/ng*, Amur *cam* (NRS 442). This item, which is the source of the international term for ‘shaman’ (transmitted into Russian from Ewenki dialects with the phonetic development *\*s* > [ʃ]), is normally considered to be of Tungusic origin, but the fact that it is homonymic with Ghilyak ‘eagle’ might indicate that the origin is to be searched on the Amuric side – though the matter remains unconcluded (Austerlitz 1984b: 233, 236; Janhunen 2005: 25–26). In any case, the Tungusic continuant sibilant *\*s* corresponds here to the Ghilyak palatal stop/affricate *c*. Since Tungusic has a primary opposition between *\*s* and *\*c*, we must assume that either the Ghilyak segment was also originally pronounced as a continuant or it was automatically replaced by the Ghilyak affricate *\*c*. The development *\*s* > *c* would correspond to the rules of modern Ghilyak morphophonology, which require that initial obstruents are always stops in underived nominals (cf. also Austerlitz 1982: 81–82). (The question concerning the long vowel in the Tungusic item will again be left outside of the present discussion. It may nevertheless be noted that there is no trace of the long vowel on the Ghilyak side, where this segment is reflected as zero.)

It has to be concluded that external data do not provide conclusive evidence of a distinction between continuant and non-continuant palatals or sibilants in Pre-Proto-Ghilyak. For the Proto-Ghilyak stage, only one set of segments of this type (either *\*c \*j* or *\*s \*z*) can be reconstructed, though it is possible that this set had both continuant (sibilant) and non-continuant (affricate) realizations. On the other hand, since the neighbouring Tungusic languages do have an original distinction between continuant and non-continuant sibilant sounds, it cannot be ruled out that Ghilyak also had this distinction, though it was later lost.

### *The strong labial stop*

As is evident from the information available from the internal reconstruction of the language, Pre-Proto-Ghilyak seems to have had a system of two sets of obstruents, weak vs. strong, the opposition of which may have involved either aspiration or voice, or both. A similar system may be reconstructed for Proto-Tungusic, Proto-Mongolic, and Proto-Turkic, while both Japonic and Ainu seem to have only one primary series of obstruents, at least in initial position. The situation in Korean remains to be clarified. What is, however, important from the areal point of view is that most of these languages have tended to eliminate the strong labial stop (*\*p*) from the system by fricativizing (> [ɸ] > *f*), velarizing (> *x*) or laryngealizing (> *h*) it. The ultimate result is the complete loss of the segment (>  $\emptyset$ ), a development that has an immediate impact on the internal structure of the system of stops. This evolutionary process has taken place at different times in the different languages and language families in the region, and it seems to have proceeded from west (Turkic) to east (Mongolic, Tungusic, Japanese). In this context, Ghilyak belongs to a small group of marginal languages which have not been affected by this tendency (12). These languages comprise Korean, many varieties of Nanai (with Ulcha and Orok), as well as Ainu.

(12) Manchu *fa* ‘window’, Udeghe *paa*, Nanai *paawa*, Neghidal *paagaa* etc. (SSTM II: 31) < *\*paga* (?) = Pre-Proto-Ghilyak *\*paga* > modern Ghilyak *pax* (*paxh*) id. (NRS 283). Since this item is only attested in Manchu, Amur Tungusic, Manchurian Ewenic (Neghidal), and Ghilyak, it is not immediately clear whether it is of Tungusic or Amuric origin. In any case, the Ghilyak data cannot represent a recent borrowing from Orok, since the latter has the shape *paawa* (as in Nanai). The voicelessness of the final consonant in the Ghilyak data (*x* or *xh*) is a regular result of final devoicing. It has to be noted that the concept of ‘window’ is not alien to Ghilyak traditional culture even in the form it existed until recently on the Lower Amur and Sakhalin, which means that a word for ‘window’ does not necessarily have to be a recent loanword, but could as well be heritage from some earlier form of Amuric. The reconstruc-

tion of the original Tungusic shape for ‘window’ is also slightly problematic. A possible protoform could have been *\*pagwa* (cf. Doerfer 1985: 261), but this would presuppose an otherwise controversial internal consonantism. (There is also the unsolved problem of vowel length on the Tungusic side.)<sup>6</sup>

### *Vowel harmony and vowel rotation*

Vowel harmony is attested in Eurasia in two main varieties: palato-velar or horizontal harmony in the west and apertural or vertical harmony, also known as the “retracted tongue root” (RTR) harmony, in the east. The two types are mutually convertible and can occur in a single language family (as in Mongolic) and their diagnostic difference should not be exaggerated. It is impossible to tell which type of harmony is “original”, though claims have recently been made that the apertural type is primary in the eastern realm of the harmony zone (Ko, Joseph & Whitman 2014). It is perhaps relevant to note that the occurrence of back velars, if present in a language, is conditioned by the lower members of the harmonic pairs, suggesting that lowering may be a property secondary to velarity. However, this may be, Ghilyak belongs to the eastern realm of the harmony zone and has a harmony opposing the high vowels *u i e* to the low vowels *o é a*. In this respect, Ghilyak is similar to its neighbours, including both Tungusic and Korean.

Assuming that vowel harmony originally operated on a horizontal basis across Eurasia, we have to postulate a process of vowel rotation that was active in the eastern part of the region (Janhunen 1981). There are, indeed, good arguments, both internal and external, for assuming that vowel rotation was a process that was active in Northeast Asia, and that started relatively late, reaching different languages and language families in different times. One consequence of the rotation process was the neutralization of certain oppositions in the system. Several languages in the region, including Korean, Tungusic, and Mongolic (as well as Turkic), seem originally to have had a system of 8 vowels (*\*a \*e \*i \*i \*o \*ö \*u \*ü*), but due to the rotation process this was typically reduced to 6, in some languages even to 5. The 6-vowel system of Amur Tungusic (with the exception of Orok)

---

6 Incidentally, it has been suggested that the Ghilyak word for ‘door’, *rhe* (NRS 324), could be a Tungusic loanword from Orok *ute* id. (Austerlitz 1989: 23–25). This, however, is chronologically unlikely, since the shape of the Orok word is a very recent result of the process of depalatalization (on which cf. Pevnov 2016, in the present volume) from earlier *\*uce* (as in Ulcha) < *\*uyke* (as in Nanai) < *\*urke* (as in Ewenki) < Proto-Tungusic *\*örke*. It would seem that both the rhoticization of dental stops and the loss of vowels in the initial syllable are relatively early developments in Pre-Proto-Ghilyak, since they are consistently shared by all dialects of the language, including those spoken on the continental side, while depalatalization is a specifically Orok development, absent even in the closest relatives of Orok on the continent.

may be seen as the result of this type of systematic neutralizations. It is therefore possible, though impossible to verify, that Ghilyak also originally had more vowels than its synchronically attested 6.

What makes the Ghilyak vowel harmony specific in its immediate areal context is that it operates synchronically in the regressive direction, as may be seen from examples like *mi-rh* ‘two sledges’ < \**mi+tu* (with *mi* ‘two’ + *tu* ‘sledge’) vs. *me-r* ‘two bundles of slices of dried salmon’ < \**mi+(a)r* (with *mi* ‘two’ + *ar* ‘bundle of slices of dried salmon’), in which the numeral \**mi* ‘two’ takes the allomorphs *mi-* (high) vs. *me-* (low) depending on the vowel originally present in the following syllable (but later lost) (Gruzdeva 2004 *passim*). It may be noted that some of the Chukotic (Northern Kamchukotic) languages, notably Chukchee, also applies vertical vowel harmony regressively, a situation which may or may not have an areal background.

It is somewhat unclear to what extent external evidence can support the postulation of a vowel rotation process in Ghilyak. However, there are occasional examples in which Ghilyak *o* corresponds to *u* in a neighbouring language, suggesting the development \**u* > \**o* (13).

(13) Manchu *ulgiyan* (with cognates in Amur Tungusic and Manchurian Ewenic, cf. SSTM II: 259) = Pre-Proto-Ghilyak \**ulgan* > *olghng* > (metathesis:) *oghlong* > (phonetically:) [o:loŋ] ‘swine’ (NRS 244). This is again a cultural item for which the direction of borrowing is not known: it could just as well be an early Amuric loanword in Tungusic as vice versa. What is well known, however, is that pigs played an important role in the diet of the early populations in Manchuria. In any case, the Amur Tungusic languages also show *o* in this item, as in Nanai *olgian*, suggesting that the Ghilyak item could be a relatively late borrowing from Amur Tungusic, in which case the rotation process would have affected the word already on the Tungusic side. This possibility is contradicted by the fact that the Ghilyak data lack any evidence of the palatal element (-*iy-*) that is present in Tungusic, a situation which suggests that the lexical correspondence is older and the palatal element may even represent a secondary development, perhaps only in Manchu, from where the word would again have been borrowed into Amur Tungusic and Manchurian Ewenic.<sup>7</sup> It may also be noted that the Orok shape of the word is *orgin* (= *orgén*), which is even less compatible with the Ghilyak data.

It is also important to note that the Ghilyak vowel *e* (often transcribed, or transliterated, as <y>) is etymologically equivalent to Tungusic (and Mongolic) \**e*

<sup>7</sup> The issue of palatalization after \**g* is an unsolved problem of Manchu-Mongolic correspondences. It is also observed in three colour terms (Okada 1962): Manchu *fulgiyan* ‘red’ (= Mongolic \**xulaxan* < \**pulagan*), *shanggiyan* > *shanyan* ‘white’ (= Mongolic \**cagaxan* < \**cagagan*), *nyowanggiyan* ‘green’ (= Mongolic \**nogaxan* < \**nyogagan*).

(14), which is the harmonic pair of *\*a*. Supposing that the Ghilyak vowel system underwent a process of rotation, the phonetic development would have been from [e] to [ə] or [ɨ], very much as it was in Tungusic (and Mongolic), though in the latter the resulting vowel tends to have a quality closer to [ə]. The modern Ghilyak vowel é [e], on the other hand, represents the lowered and fronted counterpart of the original unrounded back vowel *\*i*, which in Tungusic is generally represented as a segment with the slightly higher quality [ɪ].

(14) Ghilyak *telgu*, orthographical <tylgu> ‘legend, fairytale’, Sakhalin *telgurb* (NRS 388) = Tungusic *\*telngu* > Nanai *telunggu*, Ewenic (*\*)telung* (*\*teelung*) id. (SSTM II: 233). This item is present in both Amur Tungusic and Ewenic, as well as in Ghilyak, suggesting that the Ghilyak data is a relatively recent loanword from Tungusic, a possibility also suggested by the bisyllabic shape of the Ghilyak item. Even so, the medial vowel has been lost in Ghilyak, and the representation of Tungusic *-ng-* as Ghilyak *-g-* cannot be explained as completely regular. What is important is that the vowel of the initial syllable demonstrates the equivalence of Ghilyak *e* and Tungusic *e*. (The final element *-rb* in the Sakhalin Ghilyak data must be secondary and will not concern us here.)

As an additional detail it may be noted that the synchronic distribution of vowels in bisyllabic or longer words (Shiraishi & Botma 2016, in the present volume) no longer follow the rules of vowel harmony. This is, however, a secondary situation, conditioned by the fact that many bisyllabic words in Ghilyak are compounds in which the components originally belong to different harmonic types. Also, the high vowel *i* is frequently attested in combination with the low vowels *o é a*. This may imply that *i* often has a sequential origin and may derive from *\*Vy*, that is, from sequences of any other vowel plus the palatal glide *\*y*. A possible example is offered by items of the type *tlangi* (6), which may be reconstructed as Pre-Proto-Ghilyak *tVlangay* in view of the Ainu form *tunakai*.

#### 4. CONCLUSION

Both internal and external reconstruction suggest that, at least as far as syllable structure is concerned, Pre-Proto-Ghilyak was less aberrant from its neighbours than modern Ghilyak is today. Basically, the syllable structure of Pre-Proto-Ghilyak corresponded to the norms of the Ural-Altaic language type, with no internal or final clusters and with predominantly bisyllabic roots – though it has to be noted that Ghilyak has a number of monosyllables of the type CV (Austerlitz 1990: 22–23). The original consonant paradigm of Ghilyak was also close to that attested in, for instance, the neighbouring Tungusic languages, with two sets of stop obstruents (weak or basic vs. strong) in four places of articula-

tion (labial, dental, palatal, velar) and two liquids (\**l* vs. \**r*), plus nasals and glides. The presence of a distinct laryngeal consonant (\**h*) in Nivkh is also in accordance with the situation in the neighbouring languages, and there is no need to derive the laryngeal from any other consonant in spite of attempts in this direction (Austerlitz 1990 *passim*, where he proposes the derivation of *h* from a sibilant).

External evidence allows Pre-Proto-Ghilyak in this form to be dated to a period corresponding, more or less, to Proto-Tungusic. Proto-Tungusic, on the other hand, can confidently be dated to the Iron Age, or in absolute terms some 2 000 years back in time — a time depth shared also by some other neighbouring language families, notably Turkic (with Bulgharic) and Japonic (with Ryukyuan), as well as “Macro-Mongolic” (with Khitanic). On the other hand, archaeological evidence suggests that the contacts between Ghilyak and Ainu started not earlier than some 1 000 years ago, perhaps even later, meaning that at least some important sound changes observed in the Ghilyak-Ainu correspondences, including rhoticization (4) and vowel elision (1–2) took place only in the last millennium. This would place Pre-Proto-Ghilyak at a time depth more comparable with “Micro-Mongolic” (without Khitanic) and Koreanic. Even so, some of the loan contacts with, especially, Tungusic may have taken place much earlier at a location substantially different from the current habitat of the Ghilyak speakers.

The developments that took place in the period between Pre-Proto-Ghilyak and Proto-Ghilyak profoundly changed the paradigmatic and syntagmatic properties of Ghilyak. It is interesting to note that many of these developments are shared by Korean including, in particular, the loss of vowels in non-initial syllables (the details for Korean remain to be clarified), the formation of initial clusters, and the development of a stop system with three different sets of segments (though with phonetic differences between the two languages). Ghilyak and Korean also share the important feature of retention of the strong labial stop \**p*, which in all other language families of the region (Tungusic, Mongolic, Turkic, Japonic) has tended to undergo spirantization and/or complete loss. It remains to be investigated whether these parallels between Ghilyak and Korean can have had any contact-induced background. For the time being, almost no lexical parallels between the two languages (implying mutual loanwords) are known.

On the other hand, it has to be noted that from the point of view of synchronic phonology, morphophonology, morphology, morphosyntax, and syntax, Ghilyak shows major differences with regard to the Ural-Altai type. The incorporation of the object in the predicate would seem to have closest parallels in Kamchukotic, possibly also in Ainu (Sato 2016, in the present volume), though the Ghilyak system of incorporation (or polysynthesis) is in some respects idiosyncratic (Mattissen 2003). Other syntactic and morphosyntactic features, like the verbality

of adjectives (static verbs), however, link Ghilyak to the eastern periphery of the Ural-Altai belt, that is, Koreanic and Japonic (as well as Chinese). It may be assumed that modern Ghilyak involves the result of a mixture of several typological layers, some of which may be of a very ancient local (Pacific Rim) origin, while others reflect interaction with the neighbouring languages, most of which have long belonged to the Ural-Altai type.

## REFERENCES

- ALONSO DE LA FUENTE, José Andrés 2012. *The Ainu Languages: Traditional Reconstruction, Eurasian Areal Linguistics, and Diachronic (Holistic) Typology*. PhD thesis, University of the Basque Country.
- AUSTERLITZ, Robert 1972. *Reconstructio interna linguae Ghiliacorum*. Unpublished lecture handout, 21 April 1972. Helsinki: Finno-Ugrian Society.
- AUSTERLITZ, Robert 1976. L'appellation du renne en japonais, aïnou et surtout en ghiliak. In: W. HEISSIG, J.R. KRUEGER, F. OINAS & E. SCHÜTZ (eds), *Tractata Altaica Denis Sinor sexagenario optime de rebus altaicis merito dedicata*: 45–49. Wiesbaden: Harrassowitz.
- AUSTERLITZ, Robert 1980. On the Penumbra of Questions Surrounding the Internal Reconstruction of Gilyak. In: B. COMRIE (ed.), *International Review of Slavic Linguistics* 5: 75–78. Edmonton: Linguistic Research.
- AUSTERLITZ, Robert 1982. Gilyak Internal Reconstruction, I: Seven Etyma. In: H.I. ARONSON & B.J. DARDEN (eds), *Papers from the Second Conference on the Non-Slavic Languages of the USSR* (Folia Slavica 5): 81–88. Columbus: Slavica Publishers.
- AUSTERLITZ, Robert 1983. Préhistoire des Guiliaks: Métaux. Gilyak Prehistory: Metals. *Colloque international Sibérie 1582–1982* (Paris 24–27 mai 1983). Résumés/Abstracts/Резюме: 1–2. Paris: C.R.N.S., Laboratoire de Slavistique.
- AUSTERLITZ, Robert 1984a. Gilyak Internal Reconstruction, II: Iron and Questions Related to Metallurgy. In: H.I. ARONSON (ed.), *Papers from the Third Conference on the Non-Slavic Languages of the USSR* (Folia Slavica 7): 38–48. Columbus: Slavica Publishers.
- AUSTERLITZ, Robert 1984b. On the Vocabulary of Nivkh Shamanism: The Etymon of *qas* ('drum') and Related Questions. In: M. HOPPÁL (ed.), *Shamanism in Eurasia*, Part II, Section IV: *Myth and Language in Culture*: 231–239. Göttingen: Edition Herodot.
- AUSTERLITZ, Robert 1986. Giljakin nisäkästerminologiaa [Ghilyak mammalian terminology]. Unpublished lecture handout, 21 Nov. 1986: 1–7. Helsinki: Finno-Ugrian Society.
- AUSTERLITZ, Robert 1989. Nivkh-Ainu-Orok Linguistic Symbiosis on the Island of Sakhalin. Unpublished manuscript dated 5.IX.[19]89: 1–52
- AUSTERLITZ, Robert 1990. Typology in the Service of Internal Reconstruction: Saxalin Nivx. In: W.P. LEHMANN (ed.), *Systematic Balance in Language: Papers from the Linguistic Typology Symposium*, Berkeley, 1–3 Dec. 1987 (Language Typology 1987): 17–33. Amsterdam: John Benjamins.

- AUSTERLITZ, Robert 1994. Gilyak Internal Reconstruction, III: Ligneous Matter. In: H.I. ARONSON (ed.), *Non-Slavic Languages of the USSR: Papers from the Fourth Conference*: 229–233. Indiana University: Slavica Publishers.
- AVRORIN, V.A. [В.А. АВРОРИН] 1959. *Грамматика нанайского языка, I: Фонетическое введение и морфология именных частей речи*. Moscow: Издательство Академии Наук СССР.
- BENZING, Johannes 1956. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik*. (Akademie der Wissenschaften und der Literatur, Abhandlungen der geistes- und sozialwissenschaftlichen Klasse 1955/11) Wiesbaden: Franz Steiner.
- DOERFER, Gerhard 1978. Urtungusisch ö. In: G. DOERFER & M. WEIERS (eds), *Beiträge zur nordasiatischen Kulturgeschichte* (Tungusica 1): 66–116. Wiesbaden: Harrassowitz.
- DOERFER, Gerhard 1985. *Mongolo-Tungusica*. (Tungusica 3) Wiesbaden: Harrassowitz.
- GRUZDEVA, Ekaterina 1997. Aspects of Nivkh Morphophonology: Initial Consonant Alternation after Sonants. *Journal de la Société Finno-Ougrienne* 87: 79–96.
- GRUZDEVA, Ekaterina 1998. *Nivkh*. (Languages of the World Materials 111). 2nd edn 2003. München: LINCOM Europa.
- GRUZDEVA, Ekaterina 2004. Numeral Classifiers in Nivkh. *Sprachtypologische Universalienforschung* 57(2/3): 300–329.
- JANHUNEN, Juha 1981. Korean Vowel System in North Asian Perspective. *Han-geul* 172: 72–77.
- JANHUNEN, Juha 1996. *Manchuria: An Ethnic History*. (Mémoires de la Société Finno-Ougrienne 222) Helsinki: Finno-Ugrian Society.
- JANHUNEN, Juha 1997. The Russian Monsters: On the Etymology of an Ethnonymic Complex. *Studia Etymologica Cracoviensia* 2: 159–165.
- JANHUNEN, Juha 1998. The Horse in East Asia: Reviewing the Linguistic Evidence. In: V.H. MAIR (ed.), *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia, I*: 415–430. Philadelphia: Institute for the Study of Man.
- JANHUNEN, Juha 2005. The Eurasian Shaman: Linguistic Perspectives. In: J. PENTIKÄINEN & P. SIMONCSICS (eds), *Shamanhood: An Endangered Language* (Instituttet for sammenlignende kulturforskning, Serie B: Skrifter 117): 17–29. Oslo: Novus Forlag.
- JANHUNEN, Juha 2008. Liao: A Manchurian Hydronym and Its Ethnohistorical Connections. *Studia Etymologica Cracoviensia* 13: 89–102.
- JANHUNEN, Juha 2012. Unicorn, Mammoth, Whale: Mythological and Etymological Connections of Zoonyms in North and Central Asia. In: T. OSADA & H. ENDO (eds), *Current Studies on the Indus Civilization* 9: 189–222. New Delhi: Manohar Publishers and Distributors.
- JANHUNEN, Juha 2014. On the Ethnonyms Orok and Uryangkhai. *Studia Etymologica Cracoviensia* 19: 71–81.
- JANHUNEN, Juha 2015. On the Origin and Typological Context of the Ainu Affiliative Form. Paper presented at the workshop “Northeast Asia and the North Pacific as a Linguistic Area”, Hokkaido University, Sapporo, 21 Aug. 2015.

- JANHUNEN, Juha, FU Yuguan & GUO Shuyun 1999. The Kyakala in China: History and Present Situation. *Journal de la Société Finno-Ougrienne* 88: 248–252.
- IKEGAMI, Jirô [池上二良] 1953. ツングース語オロコ方言の母音音素 *ö* について. 言語研究 [Journal of the Linguistic Society of Japan] 22/23: 75–78.
- KIYOSE, Gisaburo N. 1996. Uvular/Post-Velar *q*, *ɣ* and *χ* in Jurchen and Manchu. In: G. STARY (ed.), *Proceedings of the 38th Permanent International Altaistic Conference (PIAC)*, Kawasaki, Japan, 7–12 Aug. 1995: 237–241. Wiesbaden: Harrassowitz.
- KO, Seongyeon, Andrew JOSEPH & John WHITMAN 2014. Comparative Consequences of the Tongue Root Harmony Analysis for Proto-Tungusic, Proto-Mongolic, and Proto-Korean. In: M. ROBBEETS & W. BISANG (eds), *Paradigm Change in the Transeurasian Languages and Beyond* (Studies in Language Companion Series 161): 141–176. Amsterdam: John Benjamins.
- KREINOVICH, E.A. [Е. А. КРЕЙНОВИЧ] 1937. *Фонетика нивхского (гиляцкого) языка*. Moscow: Государственное учебно-педагогическое издательство.
- KREINOVICH, E.A. [Е.А. КРЕЙНОВИЧ] 1955. Гиляцко-тунгусо-маньчжурские языковые параллели. (Доклады и сообщения Института Языкознания АН СССР 8): 133–167.
- MATTISSEN, Johanna 2003. *Dependent-Head Synthesis in Nivkh: A Contribution to a Typology of Polysynthesis*. (Typological Studies in Language 57) Amsterdam: John Benjamins.
- MATTISSEN, Johanna & Werner DROSSARD 1998. *Lexical and Syntactic Categories in Nivkh (Gilyak)*. (Theorie des Lexikons: Arbeiten des Sonderforschungsbereichs 282) Cologne: Institut für Sprachwissenschaft, Universität zu Köln.
- NRS = В.Н. САВЕЛЬЕВА & Ч.М. ТАКСАМИ. *Нивхско-русский словарь*. Moscow: Издательство “Советская энциклопедия”, 1970.
- OKADA, Hidehiro 1962. Color-Names in Manchu. In: N. POPPE (ed.), *American Studies in Altaic Linguistics* (Indiana University Uralic and Altaic Series 13): 225–227. Bloomington: Indiana University.
- PANFILOV, V.Z. [В.З. ПАНФИЛОВ] 1962. *Грамматика нивхского языка, I. Фонетическое введение и морфология именных частей речи*. Moscow: АН СССР, Институт языкознания.
- PANFILOV, V.Z. [В.З. ПАНФИЛОВ] 1973. Нивхско-алтайские языковые связи. *Вопросы языкознания* 1973(6): 3–12.
- PEVNOV, A.M. [А.М. ПЕВНОВ] 1994. Рефлексы вибранта в негидальском на фоне родственных ему языков. *Journal de la Société Finno-Ougrienne* 85: 125–147.
- PEVNOV, Alexandr 2016. On the Specific Features of Orok as Compared with the Other Tungusic Languages. *Studia Orientalia* 117: 47–63 (in the present volume).
- ROZYCKI, William 1994. *Mongol Elements in Manchu*. (Indiana University Uralic and Altaic Series 157) Bloomington: Indiana University Research Institute for Inner Asian Studies.
- SATO, Tomomi 2016. A Classification of the Types of Noun Incorporation in Ainu and Its Implications for Morphosyntactic Typology. *Studia Orientalia* 117: 83–93 (in the present volume).

- SHIRAISHI, Hidetoshi & Bert ВОТМА 2016. Asymmetric Distribution of Vowels in Nivkh. *Studia Orientalia* 117: 39–46 (in the present volume).
- СКОРИК, Р.Уа [П.Я. СКОРИК] 1961. Грамматика чукотского языка, I. Фонетика и морфология именных частей речи. Moscow: Издательство Академии Наук СССР.
- SSTM = В.И. Цинциус (ed.), *Сравнительный словарь тунгусо-маньчжурских языков*. Материалы к этимологическому словарю, I–II. Leningrad: Издательство “Наука”, 1975–1977.
- VOVIN, Alexander 1993. *A Reconstruction of Proto-Ainu*. (Brill’s Japanese Studies Library 4) Leiden: Brill.
- VOVIN, Alexander 2016. On the Linguistic Prehistory of Hokkaidō. *Studia Orientalia* 117: 29–38 (in the present volume).