

Tungusic languages

Past and present

Edited by

Andreas Hölzl

Thomas E. Payne

Studies in Diversity Linguistics 32



Studies in Diversity Linguistics

Editor: Martin Haspelmath

In this series (see the complete series history at <https://langsci-press.org/catalog/series/sidl>):

19. Seržant, Ilja A. & Alena Witzlack-Makarevich (eds.). Diachrony of differential argument marking.
20. Hölzl, Andreas. A typology of questions in Northeast Asia and beyond: An ecological perspective.
21. Riesberg, Sonja, Asako Shiohara & Atsuko Utsumi (eds.). Perspectives on information structure in Austronesian languages.
22. Döhler, Christian. A grammar of Komnzo.
23. Yakpo, Kofi. A Grammar of Pichi.
24. Guérin Valérie (ed.). Bridging constructions.
25. Aguilar-Guevara, Ana, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.). Definiteness across languages.
26. Di Garbo, Francesca, Bruno Olsson & Bernhard Wälchli (eds.). Grammatical gender and linguistic complexity: Volume I: General issues and specific studies.
27. Di Garbo, Francesca, Bruno Olsson & Bernhard Wälchli (eds.). Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies.
28. Unterladstetter, Volker. Multi-verb constructions in Eastern Indonesia.
29. Edwards, Owen. Metathesis and unmetathesis in Amarasi.
30. Bergqvist, Henrik & Seppo Kittilä (eds). Evidentiality, egophoricity, and engagement.
31. Floyd, Simeon, Giovanni Rossi & N. J. Enfield (eds.). Getting others to do things: A pragmatic typology of recruitments.
32. Hölzl, Andreas & Thomas E. Payne (eds.). Tungusic languages: Past and present.
33. Enke, Dankmar W., Larry M. Hyman, Johanna Nichols, Guido Seiler & Thilo Weber (eds.). Language change for the worse.

ISSN: 2363-5568

Tungusic languages

Past and present

Edited by

Andreas Hölzl

Thomas E. Payne



Andreas Hölzl & Thomas E. Payne (eds.). 2022. *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/355>

© 2022, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

Indexed in EBSCO

ISBN: 978-3-96110-395-9 (Digital)

978-3-98554-053-2 (Hardcover)

ISSN: 2363-5568

DOI: 10.5281/zenodo.7025328

Source code available from www.github.com/langsci/355

Errata: paperhive.org/documents/remote?type=langsci&id=355

Cover and concept of design: Ulrike Harbort

Typesetting: Andreas Hölzl, Patryk Czerwinski

Proofreading: Amir Ghorbanpour, Andreas Hölzl, Benjamin Brosig, Carolina C.

Aragon, Christopher Straughn, Elliott Pearl, Jeroen van de Weijer, Ludger

Paschen, Jean Nitzke, Patricia Cabredo, Tom Bossuyt, Vadim Kimmelman,

Yvonne Treis

Fonts: Libertinus, Arimo, DejaVu Sans Mono, Source Han Serif JA, Source Han

Serif ZH

Typesetting software: X_YL^AT_EX

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

<http://langsci-press.org>

Storage and cataloguing done by FU Berlin

Freie Universität  Berlin

Contents

Preface	iii
Acknowledgments	v
1 Introduction	
Andreas Hölzl & Thomas E. Payne	1
2 The causal-noncausal alternation in the Northern Tungusic languages of Russia	
Natalia Aralova & Brigitte Pakendorf	21
3 Tense and insubordination in Uilta (Orok)	
Patryk Czerwinski	63
4 ‘What’s your name?’ in Tungusic and beyond	
Andreas Hölzl	89
5 On some shared and distinguishing features of Nercha and Khamnigan Ewenki dialects	
Bayarma Khabtagaeva	149
6 Functions of placeholder words in Evenki	
Elena Klyachko	199
7 From consonant to tone: Laryngealized and pharyngealized vowels in Udihe	
Elena Perekhvalskaya	227
8 Proto-Tungusic in time and space	
Martine Robbeets & Sofia Oskolskaya	263
9 Historical language contact between Sibe and Khorchin	
Veronika Zikmundová	295
Index	330

Preface

This book originated as the proceedings of a workshop held at the 51st Annual Meeting of the Societas Linguistica Europaea (SLE) in Tallinn (see Hölzl & Payne 2018). For reasons beyond our control, the volume has been delayed for some time and we are grateful to Martin Haspelmath and Language Science Press for this opportunity. The volume is now an independent publication with some chapters having been presented at the workshop and some being later additions. We are happy to present a volume that covers all branches of Tungusic, several endangered languages (e.g., Even, Evenki, Sibe), and includes fresh first-hand data from the very last speakers of several moribund languages (e.g., Negidal, Oroch, Udihe, Uilta). We hope that this book can contribute to the documentation of these languages. We are especially grateful that this is an open access publication that will make the data freely available to all scholars.

References

Hölzl, Andreas & Thomas E. Payne. 2018. *Introduction to the workshop: The Tungusic language family through the ages*. Interdisciplinary perspectives (International Workshop at SLE 51), 2018.08.29–09.01, Tallinn.

Acknowledgments

We would like to thank the organizers of the SLE in Tallinn, the participants of the workshop, the authors of the present volume, the reviewers, Martin Haspelmath, Sebastian Nordhoff, Felix Kopecky, and the proofreaders for making this publication possible.

Chapter 1

Introduction

Andreas Hölzl

University of Potsdam

Thomas E. Payne

University of Oregon

This introduction briefly presents the Tungusic languages, discusses their classification from a meta-perspective, and outlines the contents of the eight individual contributions to this volume.

1 Tungusic languages

Tungusic (sometimes Manchu-Tungusic) is an endangered language family that encompasses approximately twenty languages located in Siberia and northern China (e.g., Janhunen 1996, 2005, 2012). These languages are distributed over an enormous area that ranges from the Yenisey River and Xinjiang in the west to the Kamchatka Peninsula and Sakhalin in the east. They extend as far north as the Taimyr Peninsula and, for a brief period, could even be found in parts of Central and South China (e.g., Hölzl & Hölzl 2019b). Tungusic-speaking peoples played an important role in the history of Northeast and East Asia and were the founders of several large empires, such as the Jin (1115–1234) and Qing dynasties (1636–1912). Recent years have seen considerable interest in this language family. Tungusic linguistics is an extremely active field of study that produced hundreds of new studies in recent years (see, for example, the references listed in Hölzl 2021b). However, the field is also very fragmented with studies being written in several languages, from a wide range of scholarly traditions. Research on Tungusic languages has been published, among others, in Chinese, Czech, English, French, German, Hungarian, Italian, Japanese, Korean, Latin, Manchu,



Polish, and Russian. Many important contributions and entire languages have gone almost unnoticed because of language barriers or the limited availability of some publications. This volume is an attempt to bring researchers from different backgrounds together to provide an open-access publication in English that is freely available to all scholars in the field. The volume emphasizes the diachronic dimension of Tungusic, tracing the development of the language family from pre-history and the earliest attestations, but also includes synchronic descriptions. This introduction briefly introduces the Tungusic languages, presents some recently published and previously overlooked data, and summarizes the individual contributions.

2 Classification and terminology

Tungusic is a top-level language family. The branching structure is open to discussion (see, e.g., Whaley & Oskolskaya 2020 and references therein), but most accounts agree on four mid-level groupings. These are comparable to branches of Indo-European, such as Germanic, Italic, or Slavic, but there is no universally accepted terminology yet. Following Janhunen (2012), the groups are referred to as Ewenic, Udegheic, Nanaic, and Jurchenic. These terms, based on the languages Even (Ewen), Udihe (Udeghe), Nanai, and Jurchen, respectively, are also used in this introduction and the contribution by Hölzl (2022 [this volume]). They are also briefly addressed in Khabtagaeva (2022 [this volume]) and Robbeets & Oskolskaya (2022 [this volume]). Some of the terms are also used by other contributions in this volume (e.g., Czerwinski 2022 [this volume]; Robbeets & Oskolskaya 2022 [this volume]; Zikmundová 2022 [this volume]). Jurchenic (e.g., Janhunen 1996) and Nanaic (e.g., Georg 2004) already have a relatively long history. For Udegheic, Janhunen (e.g., 2015) sometimes uses the term Orochic, based on the closely related language Oroch instead of Udihe. Jurchenic is also referred to as Manchuric in Alonso de la Fuente (2010/11), Jang (2020), Khabtagaeva (2022 [this volume]), or Robbeets & Oskolskaya (2022 [this volume]), a name derived from the Manchu language. In the Japanese tradition, the groups are indicated with the help of Roman numerals from I to IV (e.g., Ikegami 1974; Kazama 2003) that will be used alongside Janhunen's terminology here.

Many alternative terminologies have been proposed. For instance, Ewenic is often called Northern Tungusic (e.g., Aralova & Pakendorf 2022 [this volume]; Khabtagaeva 2022 [this volume]) while this name is reserved by Janhunen for a proposed group that includes Udegheic and Ewenic. Furthermore, many Ewenic languages of China are spoken as far south as Nanaic or Udegheic. A hypothetical branch encompassing Udegheic and Nanaic is sometimes called Amuric (e.g.,

Khabtagaeva 2022 [this volume]). But following Janhunen (1996), Amuric is also often used as a label for varieties of Nivkh. Doerfer (1978: 5) also employs the terms Northern branch for Ewenic (showing a secondary split into a Northeastern and a Northwestern group) and Southern branch for Jurchenic, but Central Eastern group for Udegheic as well as Central Western group for Nanaic, illustrating that Udegheic and Nanaic are believed to belong to one branch. Southern Tungusic in turn is used by Janhunen for a group that consists of Nanaic and Jurchenic. Given that these terminologies presuppose specific classifications of Tungusic that are not accepted by all researchers, a more neutral terminology is needed. Such a terminology is proposed in Table 1.

Table 1: Theory-neutral terminology for the four Tungusic groups

Numbers	Names	Languages
I	Ewenic	Even (Ewen), Evenki (Ewenki) ...
II	Udegheic/Orochic	Udihe (Udeghe), Oroch, ...
III	Nanaic	Nanai, ...
IV	Jurchenic/Manchuric	Jurchen, Manchu, ...

While the four groups can be considered a common ground for most approaches, their internal classification and higher-level relations are a matter of ongoing debate. Within Ewenic, for instance, Negidal is assumed to be closely related to Evenki in Doerfer (1978) or Aralova & Pakendorf (2022 [this volume]), but to the language Even in Robbeets & Oskolskaya (2022 [this volume]). The internal classification of the entire Udegheic branch (e.g., Udihe, Oroch) is investigated in the contribution by Perekhval'skaya (2022 [this volume]), demonstrating a historical continuum, while Oroch problematically is not grouped with Udihe in Oskolskaya et al. (2022). The relationship of Ewenic languages as spoken in Russia (i.e., Even, Evenki, Negidal) is briefly addressed in Aralova & Pakendorf (2022 [this volume]) and Klyachko (2022 [this volume]). Evenki dialects situated around the Chinese-Russian border (particularly Khamnigan Evenki and Nercha Evenki) are discussed in Khabtagaeva (2022 [this volume]).

The internal structure and relationship of the four mid-level groups also face problems through family-internal language mixing. This can be illustrated with the language Kilen that is variously classified as Jurchenic (Oskolskaya et al. 2022, included into the category Hezhe), mixed but basically Nanaic (Hölzl 2022 [this volume]), Udegheic (Kazama 2003, referred to as Hezhe), or as “missing link” be-

tween Udegheic and Ewenic (Kazama 1998, referred to as Kilen or Hezhen).¹ Similar difficulties exist, among others, for Kur-Urmi Nanai (or Kili) and Ussuri (or Bikin) Nanai that are classified as mixed but basically Nanaic in Hölzl (2022 [this volume]), but as related to Jurchenic in Oskolskaya et al. (2022), whereas Kazama (2003) classifies Kili as Ewenic. There is no simple solution to these problems. Doerfer (1978: 4f.) attempted to solve such obstacles by assuming transitional varieties between the four subgroups. But they are perhaps best considered mixed languages (e.g., Janhunen 2012: 6) that are the result of complex secondary interactions and different types of admixture of the four groups around the confluence of the Amur, Sungari, and Ussuri rivers. Dialect mixture and language contact are universal problems of historical linguistics for which Tungusic languages might prove a valuable natural experiment for future studies (e.g., Epps et al. 2013; McMahon 2013).

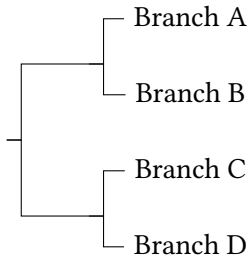
There is currently no generally agreed-upon higher-level classification of Tungusic. Logically speaking, four groups can stand in five types of relationships with each other (Table 2). Three of these represent cases of a twofold primary split, and the other two are cases of three- and fourfold splits, respectively. The exact age and internal diversity of the four groups are irrelevant for this purely topological approach.

Table 2: Logical possibilities for the classification of Tungusic

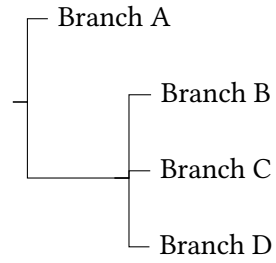
Primary split	Type	Topological schema	Possibilities
Twofold	Type 1	[A B] [C D]	3
	Type 2	A [B C D]	4
	Type 3	A [B [C D]]	12
Threefold	Type 4	[A B] C D	6
Fourfold	Type 5	A B C D	1

Altogether there are 26 logical possibilities for the topology of the Tungusic tree. Only a few of these have been proposed or are widely represented in the literature. For instance, a split into four separate branches (Type 5), sometimes attributed to Ikegami (1974), is not accepted by any current approach. Types 2 and 4 do not appear to be accepted either but remain theoretically possible.

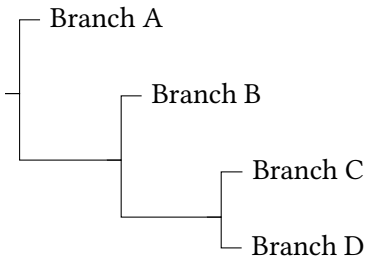
¹Due to the official classification, varieties of Kilen (Chinese *qileng* 奇楞, a mixed language) and Hezhen (Chinese *hezhen* 赫真, a form of southern Nanai) are classified as dialects of the Hezhe 赫哲 language in China (e.g., An 1986). This is similarly problematic as the term “Ewenke” for several Ewenic languages (see below and Khabtagaeva 2022 [this volume]).



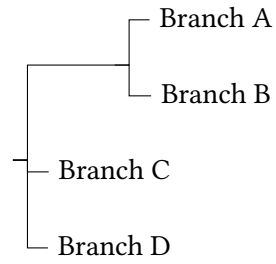
(a) Type 1



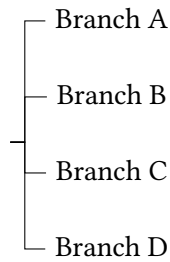
(b) Type 2



(c) Type 3



(d) Type 4



(e) Type 5

Figure 1: Possible topologies

Recent classifications only diverge from each other by few variables, two of which are included here. First, they differ with respect to the position of Udegheic that is either grouped with Ewenic or with Nanaic. Second, they disagree whether Jurchenic is the first branch to diverge from all other branches or is somehow related to Nanaic. Including only these two variables allows a meta-classification of Tungusic as illustrated in Table 3.

Table 3: A simplified meta-classification of Tungusic

	Jurchenic as first branch	Jurchenic related to Nanaic
Udegheic + Nanaic	A: IV [I [II III]]	B: I [IV [II III]]
Udegheic + Ewenic	C: IV [III [I II]]	D: [I II] [III IV]

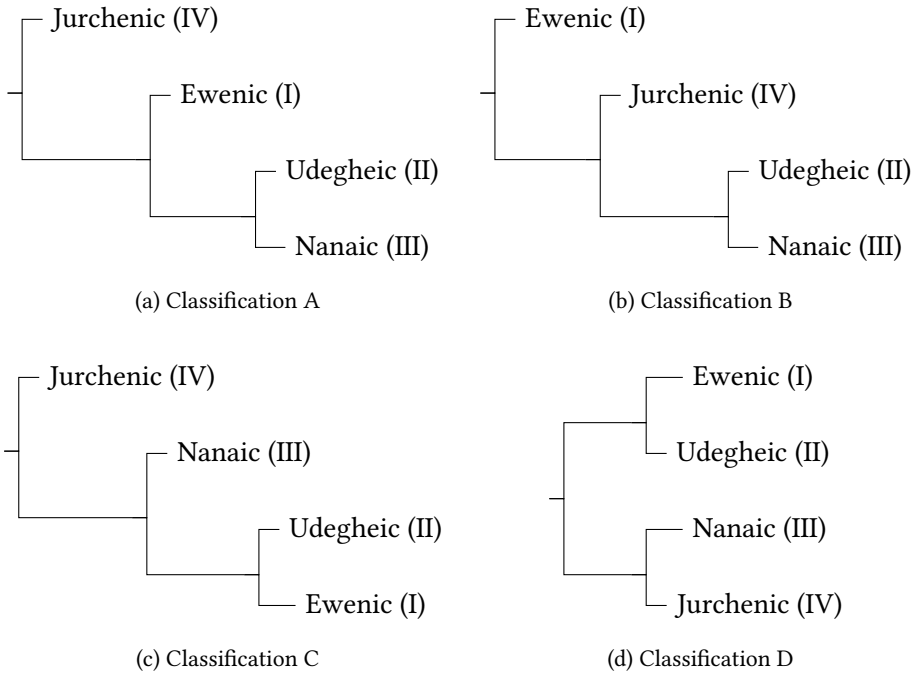


Figure 2: Recent classifications

Three of these represent cases of Type 3 (classifications A, B, C) and one of Type 1 (classification D). All four classifications agree on some points that are, however, explained differently. The well-known similarities between Nanaic and

Udegheic can theoretically be described by shared innovations (classifications A and B) or by convergence (classification D and perhaps C, e.g. Georg 2004; Alonso de la Fuente 2017: 112). The widely acknowledged differences between Jurchenic and the rest of Tungusic can be explained by an early branching (classifications A and C) or by different types of contact with non-Tungusic languages, such as Koreanic, Mongolic, Para-Mongolic, and Sinitic (classification D and perhaps B, e.g., Vovin 2006; Hölzl 2018a).

Some previous studies slightly disagree with the classification into four sub-groups. For instance, Vovin's (1993) tree resembles classification A but assumes that Even forms a separate branch after the split of Jurchenic and before the diversification of the rest of Tungusic. But Vovin (2009: 1103) later accepted classification D as proposed by Georg (2004). Most recent approaches can be categorized according to the meta-classification in Table 3. For example, Robbeets (2015) is a proponent of classification A while Doerfer (1978), although skeptical about tree diagrams, argues for classification B. Kazama (2003) and Pevnov (2017) follow classification C. Georg (2004), Janhunen (2012), and Hölzl (2022 [this volume]) accept classification D that groups Ewenic with Udegheic into a Northern and Nanaic with Jurchenic in a Southern Tungusic branch. Some approaches remain undecided or allow more than one possibility. For instance, Whaley & Oskolskaya (2020: 91) identified classification B as the most likely scenario with classification A also being supported by their study whereas Oskolskaya et al. (2022) tend towards classification D but leave the possibility for an early branching of Jurchenic open.

Whichever classification will eventually be supported by the most evidence, provided that the four groups and the tree model are accepted as a basis, it must be one of the 26 in Table 2 and probably one of the only four possibilities shown in Table 3. All previous classifications are likely to be the object of future revisions due to the development of new methodologies and in the light of newly available data.

3 Availability of new data

Tungusic linguistics has produced several outstanding works, such as the classical comparative dictionary by Cincius (1975/77) that can be considered a milestone in the field. However, it is by now over 45 years old and appeared just before new data became available on languages spoken in China starting from the end of the 1970s, not to mention that the Tungusic languages in Russia have also been increasingly well described over the last decades. Cincius (1975/77) represents only about half of the linguistic varieties (doculects) that are available

by now. It has been supplemented by newer comparative dictionaries, such as Kazama (2003), Doerfer & Knüppel (2004), or Chaoke (2014), but these do not cover all varieties either. A comprehensive review of all available data is beyond this brief introduction that limits itself to briefly presenting some new monographs on Ewenic languages in China from the last couple of years and some previously overlooked Jurchenic languages described during the 1980s.

A comprehensive classification of EWENIC necessarily includes varieties located in Russia (e.g., Arman, Even, Evenki, Negidal) and in China. Except for the dialects of Oroqen, which is called *Elunchun* 鄂伦春 in Chinese, Ewenic languages in China are collectively referred to as *Ewenke* 鄂温克, a cover term for various dialects of Solon and Evenki (e.g., Tsumagari 1992; Janhunen 1996; Khabtagaeva 2022 [this volume]). Several grammars and dictionaries of Ewenic languages spoken in China, many of which were previously underdescribed, have been published over the course of the last couple of years. Recent monographs include, but are not restricted to, two grammars and texts of “Aoluguya Ewenke” (Aoluguya/Yakut Evenki, Chaoke & Sirenbatu 2016; Hasibate’er 2016; Weng & Chaoke 2016), text collections and a grammar of “Tonggusi Ewenke” (Khamnigan/Tungus Evenki, Chaoke & Kajia 2016; Duo & Chaoke 2016), a comprehensive dictionary of “Elunchun” covering several Oroqen dialects (Han & Meng 2019), an extensive phonology of “Ewenke” (Huihe Solon, Wurigexiletu 2018), texts and a dictionary of “Arong Ewenke” (Chaoke & Kalina 2017), texts and a grammar of “Dula’er Ewenke” (Najia 2017; Chaoke & Najia 2020), a dictionary of “Nehe Ewenke” (Chaoke & Kajia 2017) etc. A detailed classification of the latter three varieties remains to be done. Chaoke (2017) is a comparative dictionary of Huihe Solon, Khamnigan/Tungus Evenki (“Morigele” dialect), and Aoluguya/Yakut Evenki.

Apart from some relics, Ewenic languages are unique among Tungusic in preserving an intervocalic *-g-, one common argument for classification B. Table 4 contains examples from the newly published sources. In some Ewenic languages, the -g- is realized as a fricative or approximant, e.g. Aoluguya Evenki [bæ:ʁɑ] ‘moon’ (Hasibate’er 2016), and in a few the -g- disappeared entirely, leading to the emergence of diphthongs and long vowels as in other Tungusic languages. This can be observed, among others, in one Khamnigan Evenki dialect (Urulyungui *tee-*, Borzya *tege-* ‘to sit’, Khabtagaeva 2022 [this volume]), in Oroqen, but also in the language referred to as “Arong Ewenke” that was recorded in Chabaqi 查巴奇 in Inner Mongolia (Chaoke & Kalina 2017). This language, tentatively classified as Solon in Hölzl (2022 [this volume]), also exhibits some features reminiscent of Solon dialects, such as the developments of geminates from consonant clusters. For instance, the cluster -rg- changed to -gg- in the word *iggə* ‘tail’ but is

preserved in *irgi* ‘brain’ (cf. Aoluguya Evenki *irgə* ~ *irgi* ‘tail’, *irgə* ‘brain’, Huihe Solon *iggi* ‘tail’, *iiggi* ‘brain’, Chaoke 2017). The dialects of Solon, Oroqen, and Evenki show an intricate pattern of family resemblances and interaction that is still incompletely understood (e.g., Whaley et al. 1999; Khabtagaeva 2022 [this volume]). This growing number of publications, although difficult to access for the wider public outside of China, represents important progress in the description of the dwindling dialectal diversity of Ewenic.

Table 4: Examples for intervocalic *-g-* in some Ewenic varieties of China (Chaoke 2017; Chaoke & Kajia 2017; Chaoke & Kalina 2017; Han & Meng 2019; Najia 2017)

Variety	fire	four	moon	summer	sun	to sit
Aoluguya Ewenke	togo	digin	beeg(a)	dʒoga	ʃigun	təgə-
Dula’er Ewenke	tog	digin	biaga	dʒogo	ʃiguŋ	təgə-
Huihe Ewenke	tog	digiŋ	beega	dʒog	ʃiguŋ	təgə-
Morigele Ewenke	tog	digin	beega	dʒuga	ʃigun	təgə-
Nehe Ewenke	tog	digin	biag	dʒuag	ʃiguŋ	təg-
Arong Ewenke	too	dijin	bia, be	dʒona	ʃiwən	təgə-
Oroqen (Elunchun)	t’ɔ	tijin	peja	dʒuwaa	ɕiwun	t’ə-

The JURCHENIC branch is of special importance for the history of Tungusic. If classifications A or C should be correct, Jurchenic represents the oldest branch of Tungusic. It is the largest branch in terms of speakers historically and currently. It has produced three distinct writing systems and by far contains the oldest and most numerous records among all Tungusic languages. Today, the last representative of Jurchenic with many speakers is Sibe (Xibe) that is increasingly well described in both its written (e.g., Sary 2017) and spoken forms (e.g., Jang 2020; Jang & Payne 2018; Zikmundová 2013). Despite being studied longest, Jurchenic is sometimes reduced to Jurchen, Manchu, and Sibe. However, Jurchen is a cover term for at least two different varieties (e.g., Kiyose 2000), Zikmundová (2022 [this volume]) points out dialectal differences within Sibe (see also Zheng 2019), and there is a large number of spoken Manchu dialects that were recorded in places such as Aihui (e.g., Shirokogoroff 1924; Wang 2005), Lalin (e.g., Mu 1986b; Ma 1997 [1988]; Wang 2001; Aixinjueluo 2014), Sanjiazi (e.g., Jin 1981; Enhebatu 1995; Kim et al. 2008; Dai 2012), Yanbian (e.g., Zhao 2000), or Yibuqi (e.g., Zhao 1989). In addition, there are at least three outlying Jurchenic varieties called Alchuka, Bala, and Kyakala that were already described in the 1980s but overlooked in comparative studies of Tungusic (Table 5). These three varieties are

probably extinct and have mostly been recorded by a scholar named Mu Yejun (also called Mu'ercha Yejun or Mu'ercha Anbulonga). To avoid confusion, Hölzl & Hölzl (2019a: 90) introduce the names “Chinese Kyakala” for the Jurchenic and “Russian Kyakala” for the Udegheic variety with that name (on which see Perekhval'skaya 2022 [this volume]). The descriptions suffer from inexact transcriptions, some typographic errors, and problematic analyses, but appear to be genuine. At least some of the data have been confirmed through independent recordings (see also Ma 1997 [1984], 1997 [1987], 1997 [1988], 1997 [1990]).

Table 5: Three outlying Jurchenic varieties

Variety	Main studies and sources
Alchuka	Mu 1981: 72; 1985; 1986a; 1986b; 1987; 1988a; Ikegami 1994; 1999: 321–343; Aixinjueluo 2014: <i>passim</i> ; Hölzl 2017, 2019, to appear
Bala	Mu 1984, 1987, 1988b; Ikegami 1999: 321–343; Li et al. 2018; Hölzl 2020, 2021a
Chinese Kyakala	Mu & Ma 1983; Mu'ercha & Mu'ercha 1983; Mu'ercha & Meng 1986; Mu 1987; Gu 2018; Hölzl 2018b; Hölzl & Hölzl 2019a

The term *Manchuric* (with an *r*) as a synonym for Jurchenic (Table 1) should not be confused with *Manchuic* (without the *r*) as used by Hölzl (2017) for one of three hypothetical subgroups of Jurchenic/Manchuric, the others being Alchukaic and Balaic. These have been tentatively proposed in analogy to Janhunen's (2012) Ulchaic subbranch of Nanaic that includes Uilta and Ulcha. *Manchuic* is a cover term for one variety of Jurchen described during Ming dynasty (Kane 1989), written Manchu (including written Sibe), and spoken Manchu dialects recorded in Northeastern (e.g., Aihui, Lalin/Jing, Sanjiazi, Yanbian, or Yibuqi Manchu) and Northwestern China (i.e., spoken Sibe). Following Zikmundová (2022 [this volume]), this last group of Manchurian and Jungarian spoken Manchu dialects that is closely related to the written language can be called Bannermen Manchu (*qiren manyu* 旗人满语 in Chinese).

Alchuka, Bala, and Chinese Kyakala, although all three are sometimes referred to as “Manchu”, do not seem to belong to Bannermen Manchu (e.g., Mu 1987; Hölzl 2017; Hölzl & Hölzl 2019a; Zikmundová 2022 [this volume]). They are characterized by several significant retentions and innovations in phonology, lexicon, and grammar. For instance, all three exhibit cases that lack the sound change $p > f$

found in written Manchu and all Manchu dialects, e.g. Alchuka *p'ut'ia-mei*, Bala *p'ut'ihiaŋ-mi*, Manchu *fucihya-mbi* ‘to cough’. Of the three languages, Alchuka and Kyakala could be more closely related, although the latter appears to show an additional substrate from Udegheic or perhaps Nanaic, e.g. the ocean spirit *taimu* 泰木 (Udihe *temu*, Nanai *temu*). Bala seems to be intricately connected to another Jurchen variety, but a comprehensive comparison and evaluation is still wanting (e.g., Kiyose 1977, 2000; Mu 1987). Both show a number of peculiarities that are otherwise rare or unattested in other Jurchenic languages, e.g. Bala *asəi*, Jurchen <asui> 阿隨 ‘NEG.EX’ (but Manchu *akū*). Bala has an additional admixture from at least one non-Jurchenic language, possibly Kilen (e.g., the word for ‘name’, Hölzl 2022 [this volume]). Alchuka, Bala, and Chinese Kyakala furthermore show influence from Bannermen Manchu or written Manchu as well as complex dialectal and sociolectal variation that remain to be investigated. Together, these three varieties illustrate that the Jurchenic branch of Tungusic is much more diverse and complex than many previous studies assumed. Alchuka, Bala, and Chinese Kyakala exhibit archaic features that are highly relevant for the prehistory of Tungusic and the reconstruction of Jurchen. Their significance cannot be emphasized enough and could be comparable to that of Chuvash and Khalaj among the Turkic languages.

4 Overview of this volume

This volume is based on a workshop held in 2018 at the 51st Annual Meeting of the Societas Linguistica Europaea (SLE) in Tallinn. It includes studies presented at the workshop and a few newly submitted ones. Altogether, it contains eight contributions from ten different scholars and several different countries. All papers were reviewed by three to four people. The contributions cover all branches of Tungusic (Table 6), a wide range of linguistic features, and very different opinions concerning the classification, reconstruction, and cultural background of Tungusic. Some of the contributions are based on first-hand data collected during fieldwork, in some cases from the last speakers of a given language (see Aralova & Pakendorf 2022 [this volume] on Negidal; Czerwinski 2022 [this volume] on Uilta; Perehval'skaya 2022 [this volume] on Udihe and Oroch).

In their contribution entitled *The causal-noncausal alternation in the Northern Tungusic languages of Russia*, Natalia Aralova and Brigitte Pakendorf investigate causative constructions in three endangered Northern Tungusic languages of the Ewenic branch – Even, Evenki, and Negidal. They look at morphological causative/non-causative alternations for 20 verbal meanings in the three

Table 6: An overview of the contributions in this volume

Tungusic languages	Studies in this volume
Ewenic	Aralova & Pakendorf; Khabtagaeva; Klyachko
Udegheic	Perekhval'skaya
Nanaic	Czerwinski
Jurchenic	Zikmundová
Pan-Tungusic	Hölzl; Robbeets & Oskolskaya

languages. For each meaning, the possibilities are marked causative, marked non-causative, equipollence (both alternations marked), or zero marking. They find that equipollence is the dominant strategy in Even and Negidal, whereas in Evenki the logical possibilities are more evenly distributed. This paper contributes significantly to ongoing theoretical discussions of the typology of voice and valence related constructions in the world's languages.

Based on data drawn from published sources spanning over 100 years and fieldwork among the last five speakers of the Nanaic language Uilta, **Patryk Czerwinski** presents a concise and typologically informed overview of the tense system. In his contribution entitled *Tense and insubordination in Uilta (Orok)*, he emphasizes the role of insubordination and verbalization in the emergence of finite verbal categories in all three temporal domains (past, present, future) and illustrates differences between the Northern and Southern dialects. The study is an important contribution in the grammatical description of this critically endangered language and substantially adds to our understanding of diachronic processes in the verbal domain of Tungusic that can also be applied to many other languages.

In *'What's your name?' in Tungusic and beyond*, **Andreas Hölzl** investigates what is referred to as the personal name question (PNQ). The study that is inspired by Frame Semantics and Construction Grammar presents a detailed cross-linguistic analysis of the PNQ that forms the basis of the analysis of the question in Tungusic languages. He identifies two main types that make use of an equational copula (Type A) and a speech act verb (Type B), respectively. Based on a global sample of about 50 languages, he describes several dimensions of variation, such as the use of different interrogatives, the marking of possession, politeness, the presence or absence of a copula, the valency of the speech act verb, etc. Including data from all Tungusic languages, he shows that the PNQ in Proto-Tungusic was of Type A and points out changes that have occurred in the individual languages through language contact.

The contribution by **Bayarma Khabtagaeva** entitled *On some shared and distinguishing features of Nercha and Khamnigan Evenki dialects* is an addition to the author's recent monograph (Khabtagaeva 2017). The study compares data of the probably extinct Nercha Evenki dialect (Castrén 1856) with modern data from Khamnigan Evenki obtained through fieldwork and some of the available literature (Janhunen 1991). It also includes comparative data from a wide range of other Tungusic languages. Through lexical and phonological similarities, she shows a close connection between the two varieties. For instance, she finds that the two varieties share the word *düčin* '40' of Mongolic origin that has a different form or is entirely absent in other Ewenic varieties. The study furthermore points out cases of lexical borrowing from different Mongolic languages, Russian, and Solon.

Placeholder words are items that speakers use to signal that they don't know or can't remember the correct word for something. Examples in English include "whatchamacallit" and "thingamajig". In *Functions of placeholder words in Evenki*, **Elena Klyachko** looks at placeholder words in terms of their morphological and syntactic behavior. In addition to providing valuable background information on Evenki varieties, including their morphological characteristics, Klyachko's study finds that placeholder words can substitute for items in almost any word class. As such they reflect the morphological character of the word they replace. A detailed discourse study of the use of placeholder words is included, showing that they have additional uses beyond the expected placeholder function. For example, they can be used as hesitation particles, and as discourse initiators.

Udihe is a highly endangered group of Tungusic varieties spoken in the Russian far east. Varieties of Udihe are famous for their multiple series of vowels, including short, long, laryngealized and sometimes pharyngealized sets. *From consonant to tone: Laryngealized and pharyngealized vowels in Udihe* by **Elena Perekhvaskaya** contains detailed discussion of the special political and sociolinguistic history of the various Udegheic varieties. Valuable spectrographic data from all recorded varieties, including data on allegro vs. full modes of pronunciation, forms the core of Perekhvaskaya's contribution. One major conclusion is that inter-variety variation in vowel inventories is explained on the basis of contrasting prosodic patterns.

In *Proto-Tungusic in time and space*, **Martine Robbeets** and **Sofia Oskolskaya** address some of the fundamental and important problems of Tungusic linguistics concerning the age, original location, and classification. They summarize and discuss the results of a recent Bayesian analysis of the Tungusic languages (Oskolskaya et al. 2022) that identifies a form of classification D as the most likely scenario but leaves the possibility of an early branching of Jurchenic open.

They assume a rough age of Proto-Tungusic at the beginning of the first millennium CE. Based on the modern distribution of the Tungusic languages and comparison with recent results from archaeology and genetic analyses of modern and prehistoric populations, they argue for a location of the Proto-Tungusic homeland somewhere around lake Khanka. They furthermore speculate that a hypothetical form of pre-Proto-Tungusic might have been spoken by incoming farmers that interacted with the distant ancestors of the modern Nivkh several millennia before Proto-Tungusic times.

With 20,000 or more native speakers, the Jurchenic language Sibe is the only modern Tungusic language that is not yet seriously endangered. There is a long-standing controversy over the ethnic identity of the Sibe people and the linguistic lineage of the Sibe language. Some, mostly linguists and outsiders to the culture, consider the spoken language to be a variety of Manchu. Others, in particular many Sibe speakers, consider the language and culture to be distinct from Manchu, arguing partly on the basis of a large number of words and concepts with clear origin in the Khorchin Mongol language. In her contribution, *Historical language contact between Sibe and Khorchin*, **Veronika Zikmundová** investigates several Mongolic features of Sibe and concludes that indeed Sibe is genetically closely related to Manchu, but that the Mongolic features can be explained on the basis of documented historical contact with Khorchin Mongol in the 15th and 16th centuries CE.

References

- Aixinjueluo, Yingsheng 爱新觉罗瀛生. 2014. *Manyu kouyu yindian* 满语口语音典. Beijing: Huayi chubanshe.
- Alonso de la Fuente, José Andrés. 2017. An Oroch word-list lost and rediscovered: A critical edition of Tronson's 1859 pseudo-Nivkh vocabulary. *Bulletin of SOAS* 80(1). 97–117.
- Alonso de la Fuente, José Andrés. 2010/11. Written Manchu uju 'head'. *Ural-Altische Jahrbücher, Neue Folge* 24. 204–226.
- An, Jun 安俊. 1986. *Hezheyu jianzhi* 赫哲语简志. Beijing: Minzu chubanshe.
- Aralova, Natalia & Brigitte Pakendorf. 2022. The causal-noncausal alternation in the Northern Tungusic languages of Russia. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 21–62. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053361.
- Castrén, M. Alexander. 1856. *Grundzüge einer tungusischen Sprachlehre nebst kurzem Wörterverzeichnis*. St Petersburg: Buchdruckerei der kaiserlichen Akademie der Wissenschaften.

- Chaoke, 朝克 D. O. 2014. *Mantonggusi yuzu yuyan cihui bijiao* 满通古斯语族语言词汇比较. Beijing: Zhongguo shehui kexue chubanshe.
- Chaoke, 朝克 D. O. 2017. *Ewenke zu san da fangyan cihui bijiao* 鄂温克族三大方言词汇比较. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, 朝克 D. O. & 卡佳 Kajia. 2016. *Tonggusi ewenkeyu huihua* 通古斯鄂温克语会话. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, 朝克 D. O. & 卡佳 Kajia. 2017. *Nehe ewenkeyu jiben cihui* 讷河鄂温克语基本词汇. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, 朝克 D. O. & 卡丽娜 Kalina. 2017. *Arong ewenkeyu* 阿荣鄂温克语. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, 朝克 D. O. & 娜佳 Najia. 2020. *Dula'er ewenkeyu huihua* 杜拉尔鄂温克语会话. Beijing: Zhongguo shehui kexue wenxian chubanshe.
- Chaoke, 朝克 D. O. & 斯仁巴图 Sirenbatu. 2016. *Aoluguya ewenkeyu huihua* 敖鲁古雅鄂温克语会话. Beijing: Shehui kexue wenxian chubanshe.
- Cincius, Vera I. 1975/77. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazыkov. materialy k etimologičeskomu slovarju*. 2 vols. Leningrad: Nauka.
- Czerwinski, Patryk. 2022. Tense and insubordination in Uilta (Orok). In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 63–87. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053363.
- Dai, Guangyu 戴光宇. 2012. *Sanjiazi manyu yuyin yanjiu* 三家子满语语音研究. Beijing: Beijing daxue chubanshe.
- Doerfer, Gerhard. 1978. Classification problems of Tungus. In Gerhard Doerfer & Michael Weiers (eds.), *Tungusica, vol.1: Beiträge zur nordasiatischen Kulturgeschichte*, 1–26. Wiesbaden: Harrassowitz.
- Doerfer, Gerhard & Michael Knüppel. 2004. *Etymologisch-ethnologisches Wörterbuch tungusischer Dialekte (vornehmlich der Mandschurei)*. Hildesheim: Olms.
- Duo, Limei 多丽梅 & 朝克 D. O. Chaoke. 2016. *Tonggusi ewenkeyu yanjiu* 通古斯鄂温克语研究. Beijing: Shehui kexue wenxian chubanshe.
- Enhebatu, 恩和巴图 Merden. 1995. *Manyu kouyu yanjiu* 满语口语研究. Hohot: Neimonggu daxue chubanshe.
- Epps, Patience, John Huehnergard & Na'ama Pat-El. 2013. Introduction: Contact among genetically related languages. *Journal of Language Contact* 6. 209–219.
- Georg, Stefan. 2004. Unreclassifying Tungusic. In Carsten Naehrer (ed.), *Proceedings of the First International Conference on Manchu-Tungus studies*, vol. 2 (Tunguso Sibirica 9), 45–57. Wiesbaden: Harrassowitz.
- Gu, Changchun 谷长春 (ed.). 2018. *Qiakala ren de gushi, xiao mo'ergen yiwen* 恰喀拉人的故事, 小莫尔根轶闻. Jilin: Jilin renmin chubanshe.
- Han, Youfeng 韩有峰 & Shuxian 孟淑贤 Meng. 2019. *Elunchunyu hanyu cidian* 鄂伦春语汉语词典. Beijing: Beifang wenyi chubanshe.

- Hasibate'er, 哈斯巴特尔. 2016. *Aoluguya fangyan yanjiu* 敖鲁古雅方言研究. Beijing: Minzu chubanshe.
- Hölzl, Andreas. 2017. New evidence on Para-Mongolic numerals. *Journal de la Société Finno-Ougrienne* 96. 21–37.
- Hölzl, Andreas. 2018a. Constructionalization areas: The case of negation in Manchu. In Evie Coussé, Peter Andersson & Joel Olofsson (eds.), *Grammaticalization meets construction grammar* (Constructional Approaches to Language 21). Amsterdam: Benjamins.
- Hölzl, Andreas. 2018b. Udi, Udihe, and the language(s) of the Kyakala. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 15. 111–146.
- Hölzl, Andreas. 2019. *Language obsolescence in retrospect: The case of Alchuka*. Paper presented at the First Conference on the Endangered Languages of East Asia (CELEA 1), 2020.09.02-03, Venice.
- Hölzl, Andreas. 2020. Bala (China): Language snapshot. *Language Documentation and Description* 19. 162–170.
- Hölzl, Andreas. 2021a. The only known text from Bala, an extinct Tungusic language. *Studia Orientalia Electronica* 9(1). 173–191.
- Hölzl, Andreas. 2021b. Tungusic languages / langues toungouses. In Anne Aarssen, René Genis & Eline van der Veken (eds.), *Linguistic bibliography of the year 2020*. Leiden: Brill.
- Hölzl, Andreas. 2022. 'What's your name?' in Tungusic and beyond. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 89–148. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053365.
- Hölzl, Andreas. to appear. The complexification of Tungusic interrogative systems. In Dankmar W. Enke, Larry M. Hyman, Johanna Nichols, Guido Seiler & Thilo Weber (eds.), *Language change for the worse* (Studies in Diversity Linguistics). Berlin: Language Science Press.
- Hölzl, Andreas & Yadi Hölzl. 2019a. A wedding ceremony of the Kyakala in China: Language and ritual. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 16. 87–144.
- Hölzl, Andreas & Yadi Hölzl. 2019b. *The endangered languages of the Manchus*. Paper presented at Endangered languages in Northern Asia. Freie Universität Berlin, 2019.11.29-30, Berlin.
- Ikegami, Jirō. 1974. Versuch einer Klassifikation der tungusischen Sprachen. In Georg Hazai & Peter Zieme (eds.), *Sprache, Geschichte und Kultur der altaischen Völker* (Schriften zur Geschichte und Kultur des Alten Orients 5), 271–272. Berlin: Akademie Verlag.

- Ikegami, Jirō 池上二良. 1994. Guanyu alechuka manyu de diaocha yanjiu 关于阿勒楚喀满语的调查研究. *Manxue yanjiu* 满学研究 2. 402–404.
- Ikegami, Jirō 池上二良. 1999. *Manshūgo kenkyū* 滿洲語研究. Tōkyō: Kyūko Shoin.
- Jang, Taeho. 2020. Xibe and the Manchuric languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 269–287. Oxford: Oxford University Press.
- Jang, Taeho & Thomas E. Payne. 2018. The modern spoken Xibe verb system. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 15. 147–169.
- Janhunen, Juha. 1991. *Material on Manchurian Khamnigan Evenki* (Castrenianumin toimitteia 40). Helsinki: University of Helsinki.
- Janhunen, Juha. 1996. *Manchuria: An ethnic history* (Mémoires de la Société Finno-Ougrienne 222). Helsinki: The Finno-Ugrian Society.
- Janhunen, Juha. 2005. Tungusic. An endangered language family in Northeast Asia. *International Journal of the Sociology of Language* 173. 37–54.
- Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.
- Janhunen, Juha. 2015. Recent progress in Tungusic lexicography. *Studia Orientalia Electronica* 3. 17–20.
- Jin, Qizong 金启葆. 1981. *Manzu de lishi yu shenghuo: sanjiazitun diaocha baogao* 满族的历史与生活：三家子屯调查报告. Harbin: Heilogjiang minzu chubanshe.
- Kane, Daniel. 1989. *The Sino-Jurchen vocabulary of the Bureau of Interpreters*. Bloomington: Indiana University Press.
- Kazama, Shinjirō. 1998. Genetical position of Hezhen. In Osahito Miyaoka & Fubito Endo (eds.), *Languages of the the North Pacific Rim*, vol. 9, 115–131. Kyōto: Nakanishi.
- Kazama, Shinjirō. 2003. *Basic vocabulary (A) of Tungusic languages* (Endangered Languages of the Pacific Rim A2-037). Kyōto: Nakanishi.
- Khabtagaeva, Bayarma. 2017. *The Ewenki dialects of Buryatia and their relationship to Khamnigan Mongol* (Tunguso Sibirica 41). Wiesbaden: Harrassowitz.
- Khabtagaeva, Bayarma. 2022. On some shared and distinguishing features of Nercha and Khamnigan Ewenki dialects. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 149–197. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053367.
- Kim, Juwon, Dongho Ko, D. O. Chaoke, Youfeng Han, Lianyu Piao & Boris V. Boldyrev. 2008. *Materials of spoken Manchu*. Seoul: Seoul National University Press.

- Kiyose, Gisaburo N. 1977. *A study of the Jurchen language and script: Reconstruction and decipherment*. Kyōto: Shōwado.
- Kiyose, Gisaburo N. 2000. Genealogical relationship of Jurchen dialects and literary Manchu. *Central Asiatic Journal* 44(2). 177–189.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- Li, Guojun 李果钧, Li, Keman 李可漫 & Dong, Yinghua 董英华. 2018. Xiao mo'ergen yiwen 小莫尔根轶闻. In Gu, Changchun 谷长春 (ed.), *Qiakala ren de gushi, xiao mo'ergen yiwen* 恰喀拉人的故事, 小莫尔根轶闻, 107–249. Jilin: Jilin renmin chubanshe.
- Ma, Mingchao 马名超. 1997 [1984]. Acheng diqu minjian wenxue kaocha baogao 阿城地区民间文学考察报告. In Ma Xiujian 马秀娟 (ed.), *Ma Mingchao minsu wenhua lunji* 马名超民俗文化论集, 113–134. Harbin: Heilongjiang renmin chubanshe.
- Ma, Mingchao 马名超. 1997 [1988]. Alechuka “badaogang” minsu kaocha 阿勒楚喀“八道岗”民俗考察. In Ma Xiujian 马秀娟 (ed.), *Ma Mingchao minsu wenhua lunji* 马名超民俗文化论集, 166–174. Harbin: Heilongjiang renmin chubanshe.
- Ma, Mingchao 马名超. 1997 [1990]. Alechuka manzu minsu kaocha baogao 阿勒楚喀满族民俗考察报告. In Ma Xiujian 马秀娟 (ed.), *Ma Mingchao minsu wenhua lunji* 马名超民俗文化论集, 175–181. Harbin: Heilongjiang renmin chubanshe.
- Ma, Mingchao 马名超. 1997 [1987]. Heilongjiang tonggusi yuqu yuanzhu minzu minjian xushi wenxue diaocha 黑龙江通古斯语区原住民族民间叙事文学调查. In Ma Xiujian 马秀娟 (ed.), *Ma Mingchao minsu wenhua lunji* 马名超民俗文化论集, 142–148. Harbin: Heilongjiang renmin chubanshe.
- McMahon, April. 2013. Issues in the genetic classification of contact languages. In Peter Bakker & Yaron Matras (eds.), *Contact languages: A comprehensive guide*, 333–362. Berlin: De Gruyter Mouton.
- Mu, Yejun 穆晔骏. 1981. *Jichu manyu gailun* 基础满语概论. Songhuajiang diqu wenwu guanlizhan.
- Mu, Yejun 穆晔骏. 1984. Juzhu zai zhangguangcailing de manzu “balaren” 居住在张广才岭的满族“巴拉人”. *Heilongjiang wenwu congkan* 黑龙江文物丛刊 2. 64–68.
- Mu, Yejun 穆晔骏. 1985. Alechuka manyu yuyin jianlun 阿勒楚喀满语语音简论. *Manyu yanjiu* 满语研究 1985(1). 5–15.
- Mu, Yejun 穆晔骏. 1986a. Alechuka manyu de shuci yu gezhuoci 阿勒楚喀满语的数词与格助词. *Manyu yanjiu* 满语研究 1986(1). 2–17.
- Mu, Yejun 穆晔骏. 1986b. Lalin manyu yuyin gailun 拉林满语语音概论. *Manyu yanjiu* 满语研究 1986(2). 2–30.
- Mu, Yejun 穆晔骏. 1987. Balayu 巴拉语. *Manyu yanjiu* 满语研究 1987(2). 2–31, 128.

- Mu, Yejun 穆晔骏. 1988a. Alechukayu yuanyin fasheng de yinbian tedian 阿勒楚喀语元音发声的音变特点. *Manyu yanjiu* 满语研究 7. 1–24.
- Mu, Yejun 穆晔骏. 1988b. Lun balayu de yuyin bianhua 论巴拉语的语音变化. *Manyu yanjiu* 满语研究 6. 1–26.
- Mu, Yejun 穆晔骏 & Wenye 马文业 Ma. 1983. Qiakalaren minsu 恰喀拉人民俗. *Heilongjiang minjian wenxue* 黑龙江民间文学 8. 191–209.
- Mu'ercha, Anbulonga 穆尔察安布隆阿 [= Mu Yejun 穆晔骏] & Yiling'a 穆尔察依凌阿 Mu'ercha. 1983. Qiantan "qiakala hejinge" 浅谈《恰喀拉合套歌》. *Heilongjiang minjian wenxue* 黑龙江民间文学 9. 145–153.
- Mu'ercha, Yejun 穆尔察晔骏 [= Mu Yejun 穆晔骏] & Huiying 孟慧英 Meng. 1986. Qiakalaren de gushi 恰喀拉人的故事. *Heilongjiang minjian wenxue* 黑龙江民间文学 19. 1–103.
- Najia, 娜佳. 2017. *Dula'er ewenkeyu yanjiu* 杜拉尔鄂温克语研究. Beijing: Shehui kexue wenxian chubanshe.
- Oskolskaya, Sofia, Ezequiel Koile & Martine Robbeets. 2022. A Bayesian approach to the classification of Tungusic languages. *Diachronica* 39(1). 128–158.
- Perekhval'skaya, Elena. 2022. From consonant to tone: Laryngealized and pharyngealized vowels in Udihe. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 227–262. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053371.
- Pevnov, Alexander M. 2017. On the origin of Uilta (Orok) *nōni* 'he or she'. *Hoppō jibun kenkū* 北方人文研究 10. 71–77.
- Robbeets, Martine. 2015. *Diachrony of verb morphology. Japanese and the Transeurasian languages* (Trends in linguistics 291). Berlin: De Gruyter Mouton.
- Robbeets, Martine & Sofia Oskolskaya. 2022. Proto-Tungusic in time and space. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 263–294. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053373.
- Shirokogoroff, Sergei Mikhailovich. 1924. *Social organization of the Manchus: A study of the Manchu clan organization*. Shanghai: Royal Asiatic Society.
- Sary, Giovanni. 2017. Neue Beiträge zum Sibe-Wortschatz. In Michał Németh, Barbara Podolak & Mateusz Urban (eds.), *Essays in the history of languages and linguistics: Dedicated to Marek Stachowski on the occasion of his 60th birthday*, 703–707. Cracow: Księgarnia Akademick.
- Tsumagari, Toshirō. 1992. A basic vocabulary of Khamnigan and Oluguya Ewenki in northern Inner Mongolia. *Bulletin of the Institute for the Study of North Eurasian Cultures* 21. 83–103.

- Vovin, Alexander. 1993. Towards a new classification of Tungusic languages. *Eurasian Studies Yearbook* 65. 99–113.
- Vovin, Alexander. 2006. Why Manchu and Jurchen look so non-Tungusic? In Alessandra Pozzi, Juha Janhunen & Michael Weiers (eds.), *Tumen jalafun jecen akū. Manchu studies in honour of Giovanni Stary*, 255–266. Wiesbaden: Harrassowitz.
- Vovin, Alexander. 2009. Tungusic languages. In Keith Brown & Sarah Ogilvie (eds.), *Concise encyclopedia of languages of the world*, 1103–1105. Amsterdam: Elsevier.
- Wang, Jing 王晶. 2001. *Lalin alechuka jingzu wenhua* 拉林阿勒楚喀京族文化. Harbin: Ha'erbin chubanshe.
- Wang, Qingfeng 王庆丰. 2005. *Manyu yanjiu* 满语研究. Beijing: Minzu chubanshe.
- Weng, Jianmin 翁建敏 & 朝克 D. O. Chaoke. 2016. *Aoluguya ewenkeyu yanjiu* 敖鲁古雅鄂温克语研究. Beijing: Shehui kexue wenxian chubanshe.
- Whaley, Lindsay, Lenore A. Grenoble & Fengxiang Li. 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75. 286–321.
- Whaley, Lindsay & Sofia Oskolskaya. 2020. The classification of the Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 81–91. Oxford: Oxford University Press.
- Wurigexiletu, 乌日格喜乐图. 2018. *Ewenkeyu yuyin shengxue yanjiu* 鄂温克语语音声学研究. Beijing: Shehui kexue wenxian chubanshe.
- Zhao, Jie 赵杰. 1989. *Xiandai manyu yanjiu* 现代满语研究. Beijing: Minzu chubanshe.
- Zhao, Jie 赵杰. 2000. Yanbian manyu duoshi 延边满语掇拾. *Minzu yuwen* 民族语文 3. 18–24.
- Zheng, Zhonghua 郑仲桦. 2019. Chabucha'er xiboyu wuxiang kouyin chutan 察布查尔锡伯语五乡口音初探. *Manzu yanjiu* 满族研究 134. 68–72.
- Zikmundová, Veronika. 2013. *Spoken Sibe: Morphology of the inflected parts of speech*. Prague: Karolinum.
- Zikmundová, Veronika. 2022. Historical language contact between Sibe and Khorchin. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 295–329. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053375.

Chapter 2

The causal-noncausal alternation in the Northern Tungusic languages of Russia

Natalia Aralova

Kiel University

Brigitte Pakendorf

Dynamique du Langage, UMR5596, CNRS & Université de Lyon

Languages differ widely in the way they code causal-noncausal alternations, in which a verb event is either presented as happening by itself (the noncausal event) or as being instigated by an external causer (the causal event). Some languages, such as English, tend not to make a morphological distinction; rather, the same form of certain verbs can express both a causal and a noncausal event, depending on the context. Other languages, such as Romanian or Russian, have a strong tendency to mark the noncausal event morphologically, while yet others, such as Turkish, tend to code the causal event with morphological means (Haspelmath 1993).

We here investigate the causal-noncausal alternation in Even, Negidal, and Evenki, three Northern Tungusic languages spoken in the Russian Federation, in a cross-linguistic perspective. In these languages, morphological means for decreasing and increasing valency predominate, although equipollence – in which both forms are morphologically marked without one being derivable from the other – is a salient strategy for verbs of destruction. Although we find broadly comparable coding patterns in these and other Tungusic languages that are similar to what is found in other languages of Northern Asia, there are numerous intriguing differences at a fine-grained level.

Keywords: Northern Asia, valence, causative, anticausative, equipollence, form-to-frequency correspondence, Tungusic



1 Introduction

The alternation between a causal and a noncausal (sometimes more specifically called inchoative) form that certain verbs can undergo has drawn a lot of scientific attention, both from a formal perspective – with a focus on only one or two languages, mainly English – and from a typological perspective based on cross-linguistic comparison (see, among many others, Haspelmath 1993; Nichols et al. 2004; Comrie 2006; Schäfer 2009; Koontz-Garboden 2009; Haspelmath et al. 2014; Levin 2015). The verbs involved in this kind of alternation form pairs

which express the same basic situation [...] and differ only in that the causative verb meaning includes an agent participant who causes the situation, whereas the inchoative verb meaning excludes a causing agent and presents the situation as occurring spontaneously. (Haspelmath 1993: 90)

Intriguingly, not all verbs undergo this alternation: while ‘break’ does, ‘cut’ does not (cf. Schäfer 2009: 653). Furthermore, languages differ greatly in the way they code causal-noncausal alternations (e.g. Haspelmath 1993; Nichols et al. 2004). Thus, some languages, such as English, tend not to make a morphological distinction; rather, the same form of some verbs¹ can express both a causal and a noncausal event, depending on the context, e.g., English *break* or *melt*. Other languages have a strong tendency to mark the noncausal event morphologically, as seen by Romanian *se sparge* : *sparge* and Russian *lomát'sja* : *lomát* ‘break’ and Romanian *se topi* : *topi* and Russian *plavít'sja* : *plavít* ‘melt’. Here and throughout the paper the first verb of each pair is the noncausal member (i.e. an intransitive verb) and the second is the causal member (i.e. a transitive verb). A third type of languages, such as Turkish, tends to code the causal event with morphological means,² as shown by the translation equivalents of ‘melt’ and ‘fill’: *eri-* : *erit-* and *dol-* : *doldur-*, respectively (Haspelmath et al. 2014: Appendices). When it is the noncausal member of the pair that is derived morphologically from the causal member, such as Negidal *ɬaptɬaβ-* : *ɬaptɬa-* ‘break’, we will use the term *anti-causative coding*. In contrast, when it is the causal member of the pair that is morphologically derived, as in the Negidal pair *un-* : *uniβkan-* ‘melt’, we will use the term *causative coding*.

¹These are mainly patient-preserving labile verbs denoting a change of state, verbs of motion, and some psych verbs (Zúñiga & Kittilä 2019: 181–182).

²That this is just a tendency and not an obligatory rule is shown by the fact that for ‘break’ Turkish marks the noncausal event morphologically: *kırıl-* : *kır-* (Haspelmath et al. 2014: Appendix A7).

2 The causal-noncausal alternation in Northern Tungusic languages

Non-morphological strategies found to express the causal-noncausal alternation are: 1) syntactic (or: periphrastic) causativization, such as *cause to die* in English (which falls outside the scope of this article); 2) ambitransitivity, as is common in English, where so-called labile verbs can express both the causal and the noncausal event, as illustrated above with ‘break’ and ‘melt’; 3) suppletion (also called lexical causativization, Zúñiga & Kittilä 2019: 25), where different roots are used to express the two events, such as English *die* vs. *kill*; and 4) equipollence, where the causal-noncausal alternation is formally marked, but neither form can be analysed as being derived from the other. This can be illustrated with the Negidal pair *ʃəgdə- : ʃəgdi-* ‘burn’, where the stem ending in *-ə* is intransitive and that ending in *-i* is transitive, and where the bare root *ʃəgd-* does not exist.

These differences in coding have been explained by the so-called degree of spontaneity of the verb event, that is, to what extent an external causer is involved in the event:

[E]vents that are placed on the spontaneous extreme of the scale would be those that can be perceived as internally caused. The occurrence of an external cause in these events is very unlikely. The externally caused events would correspond to a wider portion of the scale of spontaneous occurrence, including not just the events on the non-spontaneous extreme of the scale, but also those in the middle of the scale. (Samardžić & Merlo 2012: 4)

A different approach holds that form-frequency correspondences might account for the coding preferences (Haspelmath et al. 2014): where the noncausal member of a pair occurs more frequently, it will be the causal member that is coded overtly; conversely, if the causal member is used more often, it will be the noncausal member that is marked. In a further development, Haspelmath links the notion of degree of spontaneity to the form-frequency correspondence:

Meanings higher on the spontaneity scale tend to require longer (and more analytic) causative markers because it is less common (and hence less expected) that one uses them in a causal context, so the speaker needs to make a greater coding effort to signal the causal meaning to the hearer. Conversely, meanings lower on the scale tend to have anticausative markers because it is less common and less expected to find them in a noncausal context, so speakers need to expend coding energy to signal the noncausal meaning. (2016: 57)

An additional perspective concerning the actual use of causal vs. noncausal verbs in discourse takes pragmatic considerations into account, with the causal

member of a pair being considered more informative in the description of events that involve an external causer (Levin 2015: 77–78 reporting on Hovav 2014). Thus, speakers are assumed to choose a particular member of a causal-noncausal pair “based on their intentions, their perspective on the situation being described, and the discourse context” (Levin 2015: 78).

The preferred means of coding the alternation has been shown to be relatively stable over time, at least in some European languages (e.g. Comrie 2006: 314–317; Plank & Lahiri 2015: 45). Nichols (2018), however, argues that in certain contact situations causative coding functions as an “attractor”, that is, languages change their profile towards more causative coding. She explains this with causative coding being more iconic: the added semantic content (an agent who causes the event) is expressed by an added element in the verb form; furthermore, causatives can fairly straightforwardly grammaticalize out of phrases with the verb ‘make’. Finally, Creissels (to appear) points out that semantic changes can affect the coding of particular verb pairs. For example, in several sub-Saharan African languages, the pair ‘go out/put out (a fire)’ exhibits a cross-linguistically rare suppletive strategy. This can be explained by the fact that it has lexicalized out of ‘die/kill’, and in doing so has maintained the suppletive coding strategy found for ‘die/kill’.

In this article, we describe the strategies used by the three Northern Tungusic languages spoken in the Russian Federation, namely Even, Evenki, and Negidal, from both a discourse frequency and functional perspective, and discuss them in the light of cross-linguistic studies and comparative data from other languages spoken in Eurasia. We base our study on a twenty-verb meaning list proposed by Creissels (2018) specifically to investigate causal-noncausal alternations (1).

- (1) boil; break; burn; close; run out/use up; dry; fall/drop; get wet/(make) wet; go out/extinguish; increase; melt; move (here: go/bring); open; rise/raise; split; spoil; spread; stop (of humans); turn over; twist

As can be seen, most of the verbs in the list involve an inanimate S/O-argument upon which an animate A-argument can act in the causal state of affairs. In this, the list differs from those used in many of the preceding studies of the causal-noncausal alternation, such as Haspelmath (1993) or Nichols et al. (2004), which included verbs with both inanimate and animate undergoer, or Nichols (2018), which focusses on nine verb pairs with animate undergoer. The impact that the choice of verb meanings has on the results of the study will be addressed in §4.

The remainder of the paper is structured as follows: In the next section we briefly introduce the three languages on which this article is based and describe

our data sources. In §3 we describe the strategies these languages employ to code the causal-noncausal alternation, and in §4 we discuss the differences in frequency and function of these strategies among the three languages. In §5 we discuss the Northern Tungusic data from a genealogical and cross-linguistic perspective, and in §6 we investigate to what extent the form-to-frequency hypothesis set up by Haspelmath et al. holds for Even and Negidal. We end the paper with brief conclusions in §7.

2 The languages and data

Although there is as yet no consensus on the internal branching of the Tungusic family tree (compare, for example, the classifications in Atkine 1997 and Janhunen 2012), all classifications agree that Even, Evenki and Negidal belong to one branch, which we here label with the traditional term “Northern Tungusic”. Within this unit, Evenki and Negidal are more closely related to each other than either is to Even.

Even and Evenki are spoken by small communities scattered over a vast area of Siberia, from the Yenisey in the west to the Sea of Okhotsk in the east and from the Taimyr Peninsula in the north to northern China in the south. Evens and Evenks traditionally practised highly nomadic hunting and reindeer herding, with concomitant dispersal of the individual communities, resulting in a high degree of dialectal fragmentation. For Even, we use both published dictionaries representing the so-called standard, and a text corpus comprising data from mainly two dialects:³ Lamunkhin Even spoken in the village of Sebjan-Küöl in central Yakutia and Bystraja Even spoken in central Kamchatka. The total Even corpus comprises largely monologues, especially autobiographical narratives and some folklore, but also includes a few conversations. Sixty-six speakers (44 women and 22 men) of varying proficiency and aged 11 to 78 years at the time of recording contributed to the corpus, which numbers approximately 90,000 words. For Evenki, we base our study on published dictionaries; these represent largely the southern dialects that form the basis of the so-called standard language (cf. Table 1).

Negidal used to be spoken by a very small population of traditional fishermen and hunters settled along the lower reaches of the Amgun’ river (a tributary of the Amur), and used to comprise two dialects (Myl’nikova & Cincius 1931; Khasanova & Pevnov 2003). Nowadays, however, the Lower Negidal dialect is already extinct, and the Upper dialect is spoken with varying proficiency by only

³The corpus also includes a few texts collected from three speakers of the Tompo dialect. We were unfortunately unable to treat the individual dialects separately due to lack of data.

five elderly women (Pakendorf & Aralova 2018).⁴ Our study is based on three types of sources for Negidal (cf. Table 1): 1) We elicited the list of 20 verb meanings with two speakers (one fluent, one less so), and 2) we used the Negidal-Russian dictionary appended in Cincius (1982) to find lexemes that the speakers hadn't been able to remember. 3) We searched for the verb meanings in a corpus of transcribed, translated, and glossed oral recordings of the Upper dialect (Pakendorf & Aralova 2017) numbering approximately 60,000 words at time of writing and comprising fairy tales, everyday stories, descriptions and procedural texts as well as some conversations. These recordings represent nine different speakers, eight women and one man, of whom four women cannot be considered fluent anymore. Five of the women are a mother and her four daughters, and the recordings provided by the mother (now deceased; see footnote 4) and her oldest still living daughter make up the bulk of the corpus. Table 1 summarizes the data sources used for this investigation as well as the abbreviations used in the text to reference the languages.

Table 1: Data sources

Even (Evn)	Negidal (Neg)	Evenki (Evk)
Cincius & Rišes (1952)	List of 20 verbs elicited with 2 speakers	Boldyrev (1994)
Verified with Robbek & Robbek (2005)	Verified and completed using Cincius (1982)	Verified with Boldyrev (2000) and Myreeva (2004)
Dialectal corpora of oral narratives (c. 90k words)	Corpus of oral narratives (Pakendorf & Aralova 2017; c. 60k words)	

⁴Note that Pakendorf & Aralova (2018) list seven speakers; however, one of them (speaker 1 in their Table 1) passed away in April 2019, and another (speaker 5) passed away in February 2020.

3 Strategies of coding the causal-noncausal alternation and further valency changes in Even, Negidal, and Evenki

The most frequent strategy found in the Northern Tungusic languages to code the causal-noncausal alternation is morphological marking, with equipollence being fairly common as well (especially in the domain of verbs of destruction, see below); in contrast, we found only few verb meanings in Negidal and Evenki where an ambitransitive pair coexists with at least one pair showing morphological derivation; see (2a, b) for a Negidal example.

- (2) a. Negidal (Pakendorf & Aralova 2017: GIK_bear: 32–33)

taduk məjga:ja-n itce-kte ni=lə
 then think-NFUT-3SG see-HORT.SG who=FOC
huki-sin-e-n=də ʃaʃa-ŋi-n ti:
 turn.around-TAM1-NFUT-3SG=ADD bear-POSS-PX.3SG thus
daga-ma-tca
 near-VR-PST[3SG]

‘Then he thinks, let me see who it is. He turns around, and the bear [lit. his uncle] has [already] come close like this.’ /

‘Потом думает, давай посмотрю, кто это. Поворачивается, а дядя (=медведь) уже вот подошел.’

- b. Negidal (Pakendorf & Aralova 2017: GIK_shuka: 13)

əsi=gdə odin odi-l-la-n ogda-βa-βun
 now=CONTR wind blow-INCH-NFUT-3SG boat-ACC-PX.1PL.EX
huki-sin-e-n
 turn.around-TAM1-NFUT-3SG

‘... suddenly the wind blew and turned the boat around.’ /

‘... вдруг ветер подул, лодку повернул.’

Although we did not find any suppletive pairs among the 20 verb meanings that form the basis of the study, ‘die’ and ‘kill’ are expressed suppletively in all three languages. While Negidal and Evenki share the same forms (*bu-* ‘die’ vs. *βa:-* ‘kill’), Even has distinct items (Lamunkhin *koke-*, Bystraja *no:me-* ‘die’ vs. *ma:-* ‘kill’ for both dialects, see (3) for an illustration).

(3) Lamunkhin Even (AAS_elk_17)

... *kapkan-du h̄or-t̄a* *t̄o:ki himbi:r [...]* *ti:la-nikan*
 ... trap.R-DAT get.caught-PST.PTCP elk PTL.Y get.thin-SIM.CVB

koke-ji-n *go:-mi* *n̄oŋan pektere:-niken* ***ma:-ri-n***
 die-FUT-3SG say-COND.CVB 3SG shoot-SIM.CVB kill-PST-3SG

‘... because an elk that has gotten caught in a trap [...] will starve and die anyway, he shot and killed (it).’ /

‘... потому что попавший на капкан лось все равно [...] умрет, отощав, он убил, застрелив из ружья.’

Verbs of destruction in the Northern Tungusic languages make notable use of equipollence to distinguish valency (transitive vs. intransitive) and Aktion-sart (semelfactive vs. iterative), with different consonantal endings coding the distinct meanings (Table 2). This is most systematic in Even, where four different endings are found, while in Negidal the distinction between iterative and semelfactive transitives has largely been lost, although the distinction in Aktion-sart has been retained for the intransitive forms. In Evenki, the system appears to be at most vestigial, judging from the lack of mention in descriptions (Konstantinova 1964; Nedjalkov 1997; Bulatova & Grenoble 1999; Boldyrev 2007). The forms we provide in Table 2 are extracted from examples in Myreeva (2004) and Boldyrev (2007), and we indicate our uncertainty about our analysis with the added question marks. The suffix *-rgA*, for example, is described by Nedjalkov (1997: 228) as being a general anticausative morpheme, albeit one that is mostly used with verbs of destruction or change of state. In Negidal, the cognate form *-dgA* functions as a general anticausative as well, but with verbs of destruction it gets a specifically semelfactive reading. In this language, the ending *-nA* occurs very rarely, with *-l* generally expressing both iterative and semelfactive transitive events. Examples (4a–d) show the full system for the Negidal verb *kalta-* ‘split, halve’, one of the few for which a separate transitive-iterative form exists. Note that the root *kalta-* does not exist by itself.

Table 2: Consonantal endings of verbs of destruction and their meanings

	transitive			intransitive		
	Even	Negidal	Evenki	Even	Negidal	Evenki
iterative	-k	(-nA)	?-gA	-m	-m	?-m
semelfactive	-t	-l	?-li ~ -t, (-nA)	-r	-dgA	?-rgA

- (4) a. Negidal (Pakendorf & Aralova 2017: DIN_preparing_hide: 29)
ti_pekomi kaltal-la noŋan-ma-n
 therefore split[TR.SMLF]-NFUT[3PL] 3SG-ACC-PX.3SG
kaltal-la
 split[TR.SMLF]-NFUT[3PL]
 ‘That is why they cut it (the hide) in half.’ /
 ‘Поэтому (шкуру) разрезают на половину.’
- b. Negidal (Pakendorf & Aralova 2017: TIN_stingy_man: 69)
gə osi=gdə noŋan-ma-n halka-l-tca:
 DP now=CONTR 3SG-ACC-PX.3SG to.hammer-INCH-PST[3SG]
moŋi-l-tca: dajama-βa-n ələ kaltana:ja-n
 hit-INCH-PST[3SG] back-ACC-PX.3SG nearly split[TR.ITER]-NFUT-3SG
 ‘... he immediately started to beat and hit him, he nearly split his back.’ /
 ‘... он стал бить, колотить его палкой, спину чуть ему не переломил.’
- c. Negidal (Pakendorf & Aralova 2017: TIN_monokan: 66)
kaltadga-ja-n tik-kə-n ŋa:ləβki oje-la-n
 split[INTR.SMLF]-NFUT-3SG fall-NFUT-3SG wolf top-LOC-PX.3SG
 ‘It split and fell on top of the wolf.’ /
 ‘Треснула и упала на волка.’
- d. Negidal (field data, 04.08.17)
est’ takie mo:l kotorye ma:n-tin kaltam-ma
 exist.R such.R tree-PL which.R self-PX.3PL split[INTR.ITER]-NFUT[3PL]
 ‘There are such trees which split by themselves in several places’ /
 ‘Есть такие деревья, которые сами по себе раскалываются в нескольких местах.’

Table 3 shows the major morphological means by which the Northern Tungusic languages code valency changes, including the causal-noncausal alternation. As can be readily seen, in all three languages both transitive and detransitive derivation is achieved with a polysemous suffix comprising a labial (cf. Nedjalkov 2013: 12; Pakendorf & Aralova 2020: 299; see 5–7); this appears to have been strengthened with the erstwhile diminutive suffix *-kAn* to form the causative suffix *-βkAn* (cf. Li & Whaley 2012).

The labial (anti)causativizing suffix plays a role in the causal-noncausal alternation, since it can express both causative coding (5a, b) and anticausative coding (6a, b). It also functions as a general marker of valency change, such as deriving

Table 3: Major valency changing morphemes in Northern Tungusic

	Even	Negidal	Evenki
(Anti)causativizing	-β/-u	-β	-β
Adversative-passive	-β/-u	[-β]	[-β/-mu]
Medio-passive	-p/-b	-p	-p/-β
Causative	-βkAn	-βkAn	-βkAn

passives (7a, b). In order to cover all these functions in one gloss, Pevnov (2007: 215) calls it “ambivalent voice” in his analysis of this suffix. However, it should be noted that not all the functions are equally productive (Nedjalkov 1993).

- (5) a. Lamunkhin Even (Krivoshapkina_AX_1930s_055)
upe:-ɲʃe hiβkeŋke:n tar tcajnika-n
 grandmother-AUG.DEF silently DIST teapot.R-PX.3SG
huje-l-tce-le-n tcaj-u oŋke-tce-l-tce
 boil-INCH-PST.PTCP-LOC-3SG tea.R-ACC pour-TAM2-INCH-PST[3SG]
 ‘When the teapot started to boil, grandmother quietly started to pour tea.’ /
 ‘Бабушка тихонько, когда вскипел чайник, начала разливать чай.’
- b. Lamunkhin Even (Krivoshapkina_Marta_bear_003)
a:ŋŋa-riʃur tcaj-u igin
 stop.for.night-ANT.CVB.PL tea.R-ACC etc.Y
huj-u-t-tce-le-t
 boil-VAL-TAM2-PST.PTCP-LOC-1PL
 ‘When we had spent the night, when we were making tea, ...’ /
 ‘Переночевав, когда мы вскипятили чай, ...’
- (6) a. Bystraja Even (Egorova_RM_Arishal_127)
aha me:n-ken man-u-wa:t-ta-n
 aha REFL-DIM.INT finish-VAL-GNR-NFUT-3SG
 ‘Mhm, it disappears by itself.’ /
 ‘Ага, сам исчезает.’

2 The causal-noncausal alternation in Northern Tungusic languages

- b. Bystraja Even (Tylkanova_Lidija_Gavrilevna_2_054)
olra-β=da man-ra
 fish-ACC=ADD finish-NFUT[3PL]
 ‘They finished the fish (i.e. they exterminated the fish)!’ /
 ‘Рыбу закончили!’
- (7) a. Negidal (Pakendorf & Aralova 2017: DIN_game: 31)
ta-duk ge: hutə-βa-n jep-pa-n
 DIST-ABL second offspring-ACC-PX.3SG eat-NFUT-3SG
 ‘... then he [the devil] eats the second child ...’ /
 ‘потом второго ребенка съедает ...’
- b. Negidal (Pakendorf & Aralova 2017: DIN_EmeksiKAN: 380)
amban-du jepu-β-ta bi-ja-n
 devil-DAT eat-VAL-PST.PTCP be-FUT-3SG
 ‘... probably he has been eaten by the devil, ...’ /
 ‘наверно, амбан его съел (наверно, он амбаном съеден), ...’

Although the polysemy covering both valency-increasing and -decreasing functions might at first glance seem counter-intuitive, it is cross-linguistically not uncommon, being attested in several languages of East Asia, such as Mongolian, Japanese, and Korean (Kazama 2004: 83–84; Zúñiga & Kittilä 2019: 226); it is also a common phenomenon in the Tungusic languages (Benzing 1955: 122; Sunik 1962: 123–130). Recent studies have shown that the development is likely to have taken place from the causative to the passive function (Li & Whaley 2012; Jang & Payne 2014; Nedjalkov 2014).

The adversative-passive is a construction that “... creates an additional argument – just as the causative does” (Palmer 1994: 131). Furthermore, in contrast to standard passives, the subject is not the promoted direct object of the active transitive verb, but is “... an entity affected by the situation, possibly not being its participant” (Kazenin 2001: 906). This can be seen in the Even example (8a, b), where (8a) shows that the addressee of the bivalent intransitive verb of speech *tore-* ‘speak’ is marked with dative case (which might alternate with allative or be left unexpressed); in contrast, in the adversative construction (8b) the addressee is promoted to the subject position (as seen in the verbal subject agreement).

- (8) a. Lamunkhin Even (beseda_1626)
ebe-di-t tore-ji-p noŋan-du-n
 Even-ADJR-INS speak-FUT-1PL 3SG-DAT-PX.3SG
 ‘We’ll speak in Even to him.’ /
 ‘Ему по-эвенски будем говорить.’

- b. Lamunkhin Even (AEK_childhood_091)
tobɔr go:n-teken emie tore-β-gere-re-m tar
 this say-MULT.CVБ also.Y speak-ADVRS-HAB-NFUT-1SG that
ahi-du
 woman-DAT
 ‘... and again that woman would scold me/says bad things at me.’ /
 ‘... опять эта женщина на меня говорит.’

The adversative-passive is a productive category in Even (cf. Malchukov 1995: 21–26), but in Evenki (Nedjalkov 1997: 220–222) and Negidal (9a–b) it is restricted to environment verbs. As pointed out by Nedjalkov (2013: 3), in Evenki the adversative-passive construction “obligatorily include[s] an animate patient, i.e. the person who is subject to a certain atmospheric phenomenon considered as adversative to this person”, “while the base verbs do not contain any ‘animate’ semantic roles in their predicate frames”.

- (9) a. Negidal (Pakendorf & Aralova 2017: GIK_2tatarskoe: 28)
bu o-ηati-βun ηanə-jə uže
 1PL.EX NEG-DEONT-1PL.EX go-NEG.CVБ already.R
dəlbə-ηati-n
 fall(night)-DEONT-3SG
 ‘We’re not going, it’s already getting night.’ /
 ‘... мы не поедем, уже наступит ночь.’
- b. Negidal (Pakendorf & Aralova 2017: GIK_kljukva: 45)
noŋan goje-βa a:-tɕa-n noŋan ʃali-n bit
 3SG distance-ACC sleep-PST-3SG 3SG because.of-PX.3SG 1PL.IN
dəlbə-β-tɕa-lti
 fall(night)-ADVRS-PST-1PL.IN
 ‘She slept for a long time, because of her we were caught by the night.’ /
 ‘Она долго спала, из-за неё нас застала ночь.’

The medio-passive derivation results in constructions in which no agent is implied (compare 10b with 10a). In Even and Negidal this is marked by a labial stop rather than the labial fricative or glide used in the (anti)causitivizing and (adversative-)passive function, but in Evenki *-p* and *-β* are used interchangeably, e.g. *ula-* ‘make wet, moisten’: *ula-β* ~ *ula-p* ‘become wet’ (Nedjalkov 2013: 13).

2 The causal-noncausal alternation in Northern Tungusic languages

- (10) a. Bystraja Even (RME_Arishal_042)

jan urke-β aŋa-ri-βun
and door-ACC open-PST-1PL.EX

‘And we opened the door.’ /

‘И вот дверь открыли.’

- b. Bystraja Even (RME_Arishal_20)

iami urke aŋa-p-ta-n
PTL door open-MED-NFUT-3SG

‘... suddenly the door opened.’ /

‘... вдруг дверь открылась.’

Finally, the causative marker *-βkan* derives causatives from both intransitives (in which the causee is marked by the accusative case, as illustrated in (11b) where the morpheme appears as the allomorph *-ukeŋ-*) and transitives, with variation between dative- and accusative-marking for the causee (cf. Nedjalkov 2013: 11; Pevnov 2007: 207; Pakendorf & Aralova 2020: 302).

- (11) a. Lamunkhin Even (AXK_1930s_125)

edu tuŋŋan nime:r bi-niken tegetɬ-tɬe-l
here five neighbor be-SIM.CVB live-PST-PL

‘... here they lived as five families.’ /

‘... здесь жили они в пять семей.’

- b. Lamunkhin Even (KKK_history_012)

ebe-sel-bu tɬele-βu-tnen omen tor-du tegetɬ-ukeŋ-gel
Even-PL-ACC all-ACC-PX.3PL one earth-DAT live-CAUS-HORT.PL

go:n-tɬe-l

say-PST-PL

‘Let’s make the Evens all live in one place ..., they said.’ /

‘Давайте всех эвенов заставим жить на одном месте ...’

When both the (anti)causitivizing suffix *-β* and the causative *-βkan* can be used to encode transitivization, the difference in meaning is one of direct vs. indirect causation, as illustrated by the following examples from Negidal (12a–c). Here, the underived verb *ŋəŋə-* (12a) expresses an animate agent moving of his own volition, while the derived verb *ŋəŋə-β-* (12b) means to make something go by exerting direct, physical force, i.e. by carrying it, while *ŋəŋə-βkan-* (12c) means to cause someone to go by exerting only indirect pressure, i.e. by requesting or commanding them to go.

- (12) a. Negidal (Pakendorf & Aralova 2017: GIK_2sluchaj: 23)
man-si ʎənə-kəl ʃul-la bi amar-gida-du-s
 self-PX.2SG go-IMP.SG front-LOC 1SG behind-SIDE-DAT-PX.2SG
ʎənə-ʃa-β
 go-FUT-1SG
 ‘Go first yourself, I will go behind you.’ /
 ‘Сам иди впереди, я сзади буду идти.’
- b. Negidal (Pakendorf & Aralova 2017: DIN_crow: 92)
taj konʃe-βa hena-la:-ja-n
 DIST birchbark.box-ACC carry.on.back-SMLF-NFUT-3SG
ʃo-tki-j ʎənə-β-βə-n
 house-ALL-PRFL.SG go-VAL-NFUT-3SG
 ‘He hoisted the box on his back and brought it home.’ /
 ‘Взял этот короб и понёс домой.’
- c. Negidal (Pakendorf & Aralova 2017: APN_DIN_memories: 235)
nuʃan ətə-dgi-je-n munə(-βə) ʎənə-βkan-a
 3SG come-REP-NFUT-3SG 1PL.EX-ACC go-CAUS-NFUT[3PL]
kamenka-la
 place.name-LOC
 ‘He comes back and they send us to Kamenka.’ /
 ‘Он возвращается, и нас отправляют на Каменку.’

These data confirm Levshina’s (2016) cross-linguistic observation that the morphological marking of indirect causation (here: *-βkan*) is longer than that of direct causation (here: *-β*; cf. Haiman 1983: 784–788).

Thus, to summarize this section, the Northern Tungusic languages predominantly use morphological means to mark causal-noncausal alternations, although equipollence is common in particular with verbs of destruction. Ambitransitivity and suppletion are rare, and the latter does not occur among the 20 verb pairs which form the basis of the next section, namely the investigation of the patterns of use of the different strategies.

4 Patterns of causal-noncausal alternation among the 20 verb pairs

Table 4 summarizes the different coding patterns found in the three languages for each of the verb pairs; for the actual forms see the Appendices A–C.⁵ In the table,

⁵The data files are also downloadable in .csv format from: <http://doi.org/10.5281/zenodo.3911606>.

2 The causal-noncausal alternation in Northern Tungusic languages

nC stands for “noncausal”, C stands for “causal”, and the mathematical operator indicates the direction of derivation: nC > C “causal is derived from noncausal” (causative coding); nC < C “noncausal is derived from causal” (anticausative coding); nC ≈ C “noncausal and causal are equipollent”; nC = C “noncausal and causal are expressed by the same item” (i.e. the verb is labile). As mentioned in the preceding section, we did not find any suppletive verbs among the 20 meanings.

Following the methodology of previous studies (Haspelmath 1993, Comrie 2006), in those cases where we found synonymous pairs with different coding, we included them all in the dataset. However, we excluded verbs with very narrow meanings, such as Negidal *botco(-β)*- ‘dry out’, which refers only to hides that dry out excessively during preparation and then become unworkable. The number of synonyms and different coding patterns can be quite large (for instance, ‘burn’ in Evenki has four different coding patterns), because we tried to cover the dialectal variation and were rather inclusive in our choice. In these cases, we counted the coding patterns proportionally to their number (e.g. each pattern for ‘break’ in Even counts as 0.5 and each pattern for ‘burn’ in Evenki as 0.25; cf. the Appendices A–C).

It should be noted that our choice of meaning was partly determined by the Negidal elicitation, with which we started our data collection. For instance, since the speakers were unable to give a translation equivalent of ‘move’ (of an inanimate object), we changed this meaning to ‘move (of an animate object)’, i.e. ‘go’. Furthermore, we attempted to include only “basic” meanings and excluded stems where the derivation seemed to provide additional semantic content. We thus excluded forms such as Negidal *ηənəβkan*- ‘make someone go’ as the causative counterpart for *ηən*- ‘move (go)’, since the causative suffix *-βkAn* adds a meaning of indirect causation, as explained above (12c). We also excluded Evn *tikuken*- ~ Neg *tikeβkan*- ~ Evk *tikiβkə:n*- ‘make fall, drop intentionally, unload’, since this carries a meaning of voluntary, intentional action that is absent from ‘fall/drop’.

Given the close relationship of the languages included here, it is not surprising that the patterns we find are overall quite similar, with 15 out of the 20 verb pairs showing the same coding pattern for at least one synonym in all three languages. In contrast, what is notable is that we do find differences in the patterns based on such a small sample of verbs. For instance, for the verb pair ‘fall/drop’, Negidal uses equipollence to code the causal-noncausal alternation (*tik*- : *tibgu*-⁶), whereas Even and Evenki use causative coding (Evn *tik*- : *tikəβ*-, Evk

⁶Note that while *tik*- : *tibgu*- is synchronically equipollent, diachronically it is likely to be a causative derivation followed by metathesis: *tibgu*- < **tigbu*- < **tikbu*- < **tiki*-*bu*- (Aleksander M. Pevnov p.c., 28.06.2020).

Table 4: Coding patterns in causal-noncausal verb pairs

Verb meaning	Even	Negidal	Evenki
boil	nC > C	nC > C	nC > C
break	nC < C nC ≈ C	nC < C nC ≈ C	nC < C
burn	nC > C	nC ≈ C	nC ≈ C nC > C nC = C nC < C
close	nC < C nC ≈ C	nC < C	nC < C
run out/use up	nC < C	nC < C	nC < C
dry	nC ≈ C	nC ≈ C	nC ≈ C
fall/drop	nC > C	nC ≈ C	nC > C nC ≈ C
get wet/make wet	nC < C	nC < C	nC < C
go out/put out	nC > C	nC > C (corpus) nC = C (elicit.)	nC < C
increase	nC < C	nC < C	nC < C
melt	nC > C	nC > C	nC > C
move (go)	nC > C	nC > C	nC > C
open	nC < C	nC < C	nC < C
rise/raise	nC < C nC > C	nC = C nC > C	nC < C nC > C
split	nC ≈ C	nC ≈ C nC > C	nC ≈ C
spoil	nC < C nC > C	nC < C nC ≈ C	nC ≈ C
spread	nC < C	nC < C	nC < C
stop (of humans)	nC > C	nC > C	nC > C
turn over (around)	nC ≈ C nC > C	nC = C nC < C	nC > C nC < C
bend (twist)	nC < C	nC < C	nC < C

2 The causal-noncausal alternation in Northern Tungusic languages

tik- : *tikiβ-*). Furthermore, all three languages derive an indirect causative with the causative suffix *-βkAn*, e.g. Negidal *tikeβkanas* ‘you made me fall’. In addition to the pan-Tungusic root *tik-* (Sunik 1962: 87), Evenki has an equipollent non-cognate pair *buru-* : *huri-*. In the case of ‘go out/put out (a fire)’, Even and Negidal⁷ have causative coding (Evn *hi:β-* ~ Neg *siβ-* : Evn *hi:βi-/hi:βuken-* ~ Neg *siβi-*) in contrast to the noncausative coding found in Evenki (*si:β-* : *si:-*). It appears as if Evenki speakers reanalyzed the root-final *-β* of the noncausal form as the (anti)causativizing morpheme and from this derived the causal form by dropping the labial. For ‘rise/raise’, Even and Evenki have a verb pair showing anticausative coding (Evn *ugərəb-* : *ugər-* and Evk *ugi:riβ-* : *ugi:r-*, respectively) where the Negidal cognate is labile (*ugi-*); in addition, all three languages have a synonymous pair with causative coding, but here only the Negidal and Evenki forms are cognate (Neg and Evk *tukti-* : *tuktiβ-* vs. Evn *ojtci-* : *ojtciβkan-*).

Some further pattern differences we find with respect to specific verbs in the dataset are:

- The verb meaning is expressed by different lexemes, which nevertheless show the same coding pattern, e.g. ‘spread’, which shows consistently anticausative coding: *girkəb-* : *girkə-* in Even, *gildeβ-* : *gilde-* in Negidal, *səktəβ-* : *səktə-* in Evenki.
- Differences in coding pattern correlate with differences in lexeme form, e.g. ‘burn’, which is coded causatively in Even: *dur-* : *duruken-*, but which is expressed by the equipollent pair *ʃəgdə-* : *ʃəgdi-* in both Negidal and Evenki; in addition, Evenki has three more synonyms, none of which are cognate to the Even or Negidal form, and each of which shows a different coding pattern (Appendix C).
- The lexemes are cognate, but the coding patterns differ, e.g. ‘fall/drop’ and ‘go out/put out (of fire)’ discussed above, or ‘spoil’, which in Even shows causative coding: *mun-* : *munuken-*, but which is expressed by the equipollent pair *munu-* : *muni:-* in Evenki.

In Table 5 we summarize the frequency of the different coding patterns for the three languages. While in Even anticausative and causative coding occur with approximately equal frequency, in Negidal and Evenki anticausative coding predominates over causative coding. This is particularly pronounced for Evenki, where anticausative coding is nearly twice as frequent as causative coding.

⁷For the Negidal pair ‘go out/put out’, we included two coding patterns in the dataset: one is found in the corpus and the other was obtained during elicitation.

Table 5: Frequency of different causal-noncausal relations in the Northern Tungusic languages (over 20 verb pairs)

Relation	Even	Negidal	Evenki
nC > C	8.5	5.5	5.75
nC < C	8	8.5	10.25
nC ≈ C	3.5	4.5	3.75
nC = C	–	1.5	0.25

These results offer some counterevidence to the findings of Nichols et al. (2004: 180), who state that “[f]rom eastern North America across the Bering Strait and through Siberia there is a large region marked by a strong preference for augmentation [i.e. causative coding]”. These differences in results are likely to be due to the different verb meanings included in the studies: as mentioned in the Introduction, Nichols et al. (2004) based their investigation on 18 verb pairs, of which nine have an animate undergoer such as ‘laugh’ or ‘sit’, and only nine have an inanimate undergoer and therefore partly overlap with the verb meanings included here.

The impact of the verb meanings on the coding patterns can be further seen from data on Even and Evenki presented in a recent follow-up study by Nichols (2018). This is based on only the nine verb meanings with animate undergoers from the original dataset: ‘laugh : make laugh/amuse’, ‘die : kill’, ‘sit : seat/make sit’, ‘eat : feed/give food’, ‘learn/know : teach’, ‘see : show’, ‘be/become angry : anger’, ‘fear : frighten’, and ‘hide’. In this study, 63% of the nine Evenki verbs show causative coding vs. 50% of the Even verbs (Nichols 2018: Table 6). In our study with its mostly inanimate verbs, approximately 34% (11/32) of all Evenki verb pairs (i.e. counting over all synonyms) and about 38% (12/32) of all Even verb pairs show causative coding. When counting how many of the 20 verb meanings included in our study can be expressed with a causative derivation (irrespective of whether there are synonymous pairs using a different coding strategy), we find 40% (8/20) verb meanings with causative coding in Evenki and 50% (10/20) in Even. Not only are the overall proportions of causative coding generally lower in our study than those reported by Nichols (with the sole exception being the proportion of verb meanings in Even), but the pattern is the opposite: in our data, Evenki makes less use of causativization than Even, while Nichols finds that it makes more use.

To summarize this section, the preferred strategies of the Northern Tungusic languages to code the causal-noncausal alternation are anticausativization and

causativization, with a relatively high frequency of equipollence. Even though the languages are very closely related and the list of verb meanings is quite small, there are still noteworthy differences between them. However, the results of such studies depend considerably on the verb meanings they are based on as well as on the data bases used. For instance, the Evenki dictionaries are much more extensive than the Negidal dictionary and include many dialectal forms. In the following section, we compare the coding patterns found in Even, Evenki, and Negidal to their Tungusic relatives as well as to other Eurasian languages.

5 Northern Tungusic causal-noncausal alternations in a genealogical and cross-linguistic perspective

5.1 Cognates across Tungusic languages

In the preceding section we already mentioned that in some cases cognate verbs show different coding patterns across the three Northern Tungusic languages. Some further interesting patterns emerge when comparing Even, Negidal, and Evenki with other Tungusic languages, namely Nanai, Udihe, and Manchu, the data for which come from the World Atlas of Transitivity Pairs (2014) with verification by specialists of these languages (see §5.2 for further details on this dataset). For instance, the equipollent final vowel change of noncausal *-o/-ə* : *-i* (as found in Negidal and Evenki *ʃəgdə* : *ʃəgdi* ‘burn’ and *olgo* : *olgi* ‘dry’) is also found for the Nanai cognates *ʃəgdə* : *ʃəgʃi*- and *holgo* : *holgi(ʔi)*- and for the putative Udihe cognate *ogo* : *wagi* ‘dry’. Although this alternation is synchronically equipollent, etymologically it traces back to a causativizing pattern with the Tungusic causative **-gi* (Benzing 1955: 122; Sunik 1962: 93). However, Udihe has regularized the causal form of ‘burn’ to *ʃəgdə-βənə*, and Manchu has regularized the causal form of ‘dry’ to *olho-bu*, with both languages deriving the causal form with their regular causative suffix. Udihe also derives the causal form of ‘turn (around, over)’ from the base root (*kumtə* : *kumtə-βənə*), while Negidal and Evenki treat the base root as causal and derive the noncausal form (*kumtəβ* : *kumtə*).

Furthermore, some cases of semantic shift appear to have taken place. For example, the Nanai word *dasip* : *dasi*- means both ‘close’ and ‘cover’, while the Northern Tungusic cognate Evn *dasab*- ~ Neg, Evk *dasib*- : Evn, Neg, Evk *das*- means only ‘cover’, with a separate root (Evn *homab*- ~ Neg *samuβ*- ~ Evk *so:miβ*- : Evn *hom*- ~ Neg *sam*- ~ Evk *so:m*-) meaning ‘close’. Likewise, the Nanai word for ‘break’ *kaltalip*- : *kalta*- is cognate to the equipollent root *kalta*- (Evn, Neg) ~

kakta- (Udihe) ‘split’. It is unclear whether this is a semantic shift from ‘split’ to ‘break’ in Nanai, or whether it is an artefact of data collection (since ‘break’ and ‘split’ are very close in meaning).

5.2 Causal-noncausal alternations across Eurasia

For a cross-linguistic comparison of the Northern Tungusic causal-noncausal alternation we also used data from the World Atlas of Transitivity Pairs (2014). This Atlas contains information on coding patterns for 31 verb meanings based on Haspelmath (1993: 104). Thirteen verb meanings overlapped between our list of meanings (1) and that of Haspelmath (1993). However, we decided to exclude the meaning ‘put out/go out’, since we noticed that for the Even WATP dataset the collected meaning was ‘exit’ and not ‘extinguish’. Since other contributors to the WATP may also have misunderstood the targeted meaning, we opted to exclude this from the dataset in order to ensure that we are indeed comparing the same meanings across languages. We thus used only twelve verb meanings per language for our cross-linguistic comparison: *boil, break, burn, close, dry, melt, open, rise/raise, split, spread, stop, turn over*. We included 60 languages of Eurasia in our comparison, as listed in the legend to Figure 1. For each of them we counted the number of coding patterns in the same manner as shown in Table 5 for Even, Negidal and Evenki. It is important to mention that coding decisions might have had an impact on the counts. For example, in Evenki we analyze the verb pair *ula-p-* ‘get wet’ and *ula-* ‘make wet’ as having anticausative derivation with morphological vowel shortening in the root of the derived noncausal verb. But for Nanai we followed the decision of the WATP contributor Kazama, who coded the relation between *kaltaa-* ‘break (intr.)’ and *kaltali-* ‘break (tr.)’ as equipollent, since there is the pair *xətu-ə- : xətu-li-* ‘split’, where the final vowel in the intransitive verb is clearly a separate vowel, not length. This suggests that ‘break (intr.)’ might also be analysed as *kalta-a-*, with the noncausal form in these equipollent pairs being marked by a mid-low vowel and the causal form being marked by *-li*. The resulting frequency table was plotted on the map in Figure 1 in the form of pie charts reflecting the proportions of the different coding patterns in each language.

The coordinates for the languages were obtained mostly from Glottolog (Hammarström et al. 2019), with a few exceptions, such as Domaaki and Burushaski, which had completely overlapping pie charts and were plotted next to each other. For Even we chose the location of Ola, which is the place where Standard Even is spoken, even though our data come predominantly from the Lamunkhin and Bystraja dialects, and not the standard variety. We chose Ola since it is midway

2 The causal-noncausal alternation in Northern Tungusic languages

between the locations where the Lamunkhin and Bystraja dialects are spoken and it is also frequently used in typological maps to represent the location of Even as a whole.

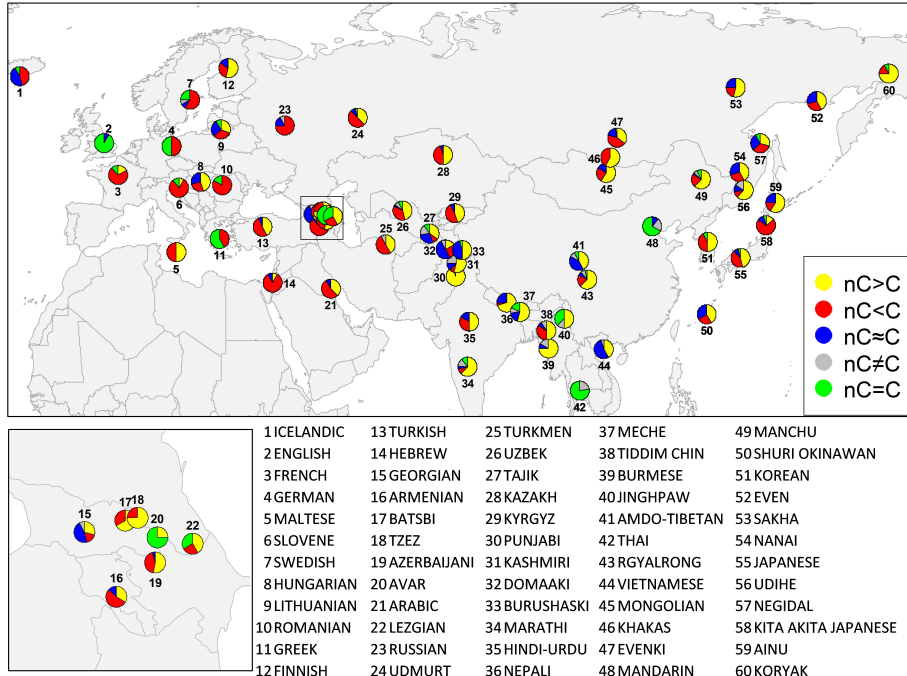


Figure 1: Causal-noncausal alternations in Eurasia; created with R (2020), based on data in WATP (2014)

In Figure 1 the Northern Tungusic languages are labeled with 47 (Evenki), 52 (Even) and 57 (Negidal). In general, they do not stand out in this picture, since they show the most common coding patterns – causativization and anticausativization as well as equipollence – roughly in the same proportion, with Negidal additionally having a small proportion of ambitransitive verbs (for the meanings ‘rise/raise’ and ‘turn over’). Nanai (labelled as 54) matches this distribution as well, whereas Udihe (56) and Manchu (49) show a stronger preference for causativization. With respect to the other languages of the region, the Tungusic languages seem to be rather typical in their marking of causal-noncausal relations: a similar pattern is found in Sakha (53), Mongolian (45), Ainu (59) and the Shuri dialect of Okinawan (50).

The Tungusic languages show a degree of homogeneity of marking the causal-noncausal alternation that is intermediate between that found within the Japonic languages and that found in the Turkic family. In the former, Shuri Okinawan (50), Standard Japanese (55), and the Kita-Akita dialect of Japanese (58) show widely differing proportions of the three major coding patterns, while in the latter, languages as geographically distant as Turkish, Azerbaijani, Central Asian Turkic, and Khakas all show an overall very similar pattern of roughly equal proportions of anticausative and causative coding, with equipollence being very rare. Interestingly, Sakha (Yakut) (53) shows a considerably higher proportion of equipollent coding than its Turkic relatives, a feature that might be due to contact with Tungusic languages.

In general, as seen in Figure 1, while causativization is a feature of Asia as a whole, being quite common in South Asia as well as in some languages of China, it gradually decreases from East to West: indeed, in Europe the only languages with a high proportion of causative strategies are non-Indo-European (Finnish, Hungarian, Maltese, Turkish, some languages of the Caucasus, and Udmurt). Furthermore, as pointed out by Nichols et al. (2004: 180), causativization extends beyond the Bering Strait into North America:

Northern Asia and North America, and to some extent also Central America-Mexico, favor augmentation [i.e. causativization] (and to a lesser extent double derivation [i.e. equipollence]) and disfavor reduction [i.e. anticausativization], ambitransitivity, and auxiliary change.

To summarize this section, the Northern Tungusic languages show quite similar coding strategies to their Tungusic relatives. Some of the patterns are clearly old in the Tungusic family, such as the final vowel alternation in equipollent stems, which goes back to an erstwhile causative pattern, while individual innovation can be shown to have played a role as well, such as the regular causative derivation of formerly equipollent stems in Udihe or Manchu. The Northern Tungusic languages also do not stand out in areal perspective, making use of the most common strategies. To what extent these preferred patterns of coding might be explained by the form-to-frequency hypothesis will be addressed in the next section.

6 The form-frequency correspondence in Even and Negidal

As mentioned in the Introduction, in their paper Haspelmath et al. (2014) focus on the frequency-based motivation for the causal-noncausal alternation. Using large corpora of seven languages they test several predictions. Following their approach, we use data from our Even and Negidal corpora to test the form-to-frequency prediction, which states that unmarked forms are more frequent. This is formulated by Haspelmath et al. (2014: 597) as follows:

In each language, in a causative verb pair, the causal member will be rarer than the noncausal member, while in an anticausative verb pair, the causal member will be more frequent than the noncausal member.

In our count we did not consider the frequencies of labile verbs (one pair for ‘rise/raise’ and one pair for ‘turn over’ in Negidal, both synonymous with morphologically marked pairs), nor did we consider equipollent verbs, as neither of these types is informative for this hypothesis. In Even, the meaning ‘rise/raise’ is expressed with two synonymous verb pairs with opposite coding (see Appendix A), but in our corpus we find only one of these verbs with both causal and noncausal members (*ojtci-* ‘rise’ vs. *ojtciβkan-* ‘raise’). For this reason, we included only the causative coding in our count (see Appendix D). The verb meaning ‘spoil’ was not found in either the Even or the Negidal corpus.

There are ten verb meanings in both Even and Negidal that clearly confirm the form-to-frequency prediction and only four and three, respectively, that do not. If we include those verb pairs where the difference in frequency is very small (only 1–2), so that we cannot say with certainty that one of the forms is truly more frequent than the other (see the cases in the table where “yes” is in brackets), the number of verb pairs confirming the form-to-frequency prediction rises to 12 in both languages. Our data thus do provide some support for the cross-linguistic tendency proposed by Haspelmath et al. (2014).

Haspelmath et al. (2014) suggest that the cross-linguistic tendency for deriving the less frequently used form might in individual languages be overridden by that language’s “macro-type”, i.e a potentially strong preference for causative or anticausative coding (as exemplified by Romanian, which has a distinct preference for anticausative coding and more verb pairs that disconfirm than confirm the prediction, Haspelmath et al. 2014: 599). In order to abstract away from such language-specific particularities they examine the frequencies

of (non)causal uses independently from their coding. They test whether the proportion of noncausal verb uses correlates with the causative prominence scale proposed by Haspelmath (1993). The causative prominence scale ranks the verb meanings included in the study from the most causative-prominent to the most anticausative-prominent and reflects which verb meanings tend to be coded as causatives, and which tend to be coded as anticausatives, across the 21 languages included in Haspelmath's study. Haspelmath et al. (2014) show that the ratio of noncausal uses over all occurrences of a particular verb meaning correlates significantly with the rank of a particular verb on the causative prominence scale: the verb meanings with the least causative prominence (i.e. those where the causal form is the basic form and it is the noncausal form which is derived) tend to have the least noncausal uses in the analysed corpora.

Since we lacked data for all of the verb meanings included by Haspelmath et al. (2014), we did not replicate their test for Even and Negidal; rather, we followed the modified approach proposed by Seifart et al. (2019), who reduce the list of verb meanings to six with different levels of causative prominence cross-linguistically: high (*boil, dry*), mid (*turn, burn*) and low (*break, open*). They modify the causative prominence scale by using data from WATP and by including some data from previous studies (Haspelmath 1993; Nichols et al. 2004) as well as data from their own oral corpora of 14 understudied languages from South America and Papunesia. The results of Seifart et al. (2019) are quite consistent with those of Haspelmath et al. (2014), notwithstanding the fact that they use a modified causative-prominence scale, fewer verb meanings, and much smaller corpora. Both studies confirm that for the verb meanings with lower causative prominence the corpus frequency of the noncausal event is lower, and vice versa, that when the causative prominence is high, the frequency of the noncausal event is higher.

We test whether this tendency holds for the data in the Even and Negidal corpora by plotting the ratio of the noncausal uses over the total number of uses for each verb onto the typological causative prominence scale taken from Seifart et al.'s (2019) study. The results are visualized in Figures 2 and 3. It should be noted that this analysis can only be taken as indicative of tendencies of use in these languages, since it is based on rather few datapoints.

Even and Negidal show different results. In Even, the frequency of use of noncausal verb meanings does not increase with increasing rank on the typological causative prominence scale, while in Negidal it does. The difference between the two patterns is caused by two verbs with mid and high causative prominence: 'burn' and 'dry'. It is remarkable how differently 'burn' and 'dry' are used in the corpora of these closely related languages. In the Negidal corpus, the ratio of

2 The causal-noncausal alternation in Northern Tungusic languages

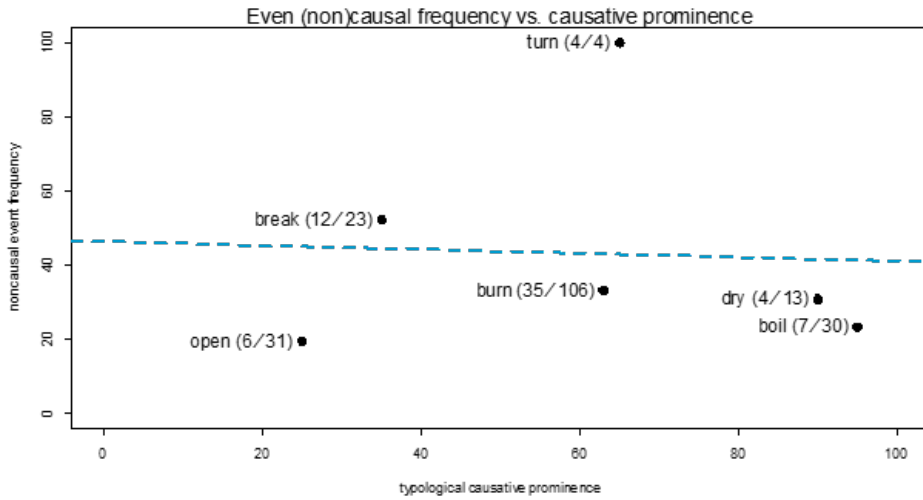


Figure 2: Noncausal uses of six verbs in Even. For each verb the number of the noncausal uses over the total number of uses is shown in brackets; created with R (2020)

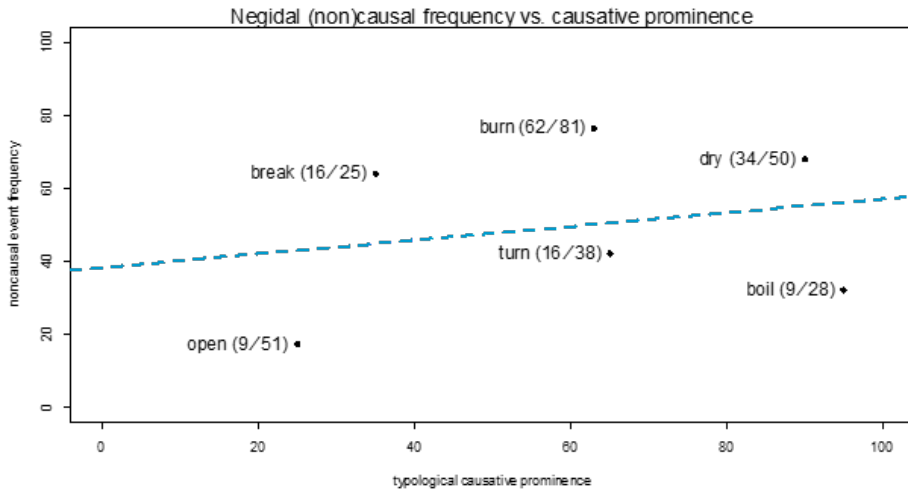


Figure 3: Noncausal uses of six verbs in Negidal. For each verb the number of the noncausal uses over the total number of uses is shown in brackets; created with R (2020)

noncausal usage is 77% for ‘burn’ (62/81) and 68% for ‘dry’ (34/50). In the Even corpus, in contrast, only 33% (35/106) of the occurrences of ‘burn’ have a noncausal meaning⁸ and there are only about 30% (4/13) of noncausal uses of ‘dry’. However, it should be noted that the Even dialects show opposite patterns for ‘burn’: in the Lamunkhin dialect only ~14% (8/56) of the occurrences of ‘burn’ are noncausal, whereas in the Bystraja dialect ~64% (27/42) of the occurrences of this verb are noncausal. Thus it is the Lamunkhin dialect of Even that patterns very differently from both its sister dialect and Negidal. This underlines the high degree of lect-specificity of these patterns of usage.⁹

Another observation concerns ‘boil’, a meaning with high causative prominence: in contrast to what is expected on typological grounds, this verb meaning has a rather low ratio of noncausal usage in both the Even and the Negidal corpora (23% and 32%, respectively¹⁰), and this is the only verb which disconfirms the form-to-frequency prediction in both Even and Negidal (see Appendix D). However, this low frequency of noncausal ‘boil’ is not exceptional cross-linguistically: in several languages of Seifart et al.’s (2019) sample noncausal ‘boil’ occurs with zero or low frequency as well. One can speculate why this pattern emerges for ‘boil’ in several languages spoken in vastly different geographical regions, but not for other verbs with high causative prominence, such as ‘dry’ or ‘freeze’. Whereas freezing and drying can occur spontaneously in natural environments, completely spontaneous boiling is found only in thermal springs or in a volcano crater. Instead, for most boiling events there must be a human who initiates the process by putting a pot with water on a fire. Thus, purely spontaneous boiling is an infrequent event. However, there is a time lapse between the causal event (putting the pot on the fire) and the noncausal event (the water boiling), so that the actual boiling event might be conceptualized as spontaneous and be expressed with a noncausal base form. But in some languages, it seems, people tend to talk more about the causal event because that in general has to precede the noncausal, spontaneous boiling. In addition, in Negidal the verb ‘boil’ appears to be lexicalizing to generalized ‘cook’ – which is of course a causal event and thus adds more causal uses.

⁸Notably, ‘burn’ in Even is also one of the few verbs in Appendix D which does not confirm the form-to-frequency prediction.

⁹All the frequency differences we discuss here are significant: Negidal vs. Even ‘burn’: $\chi^2 = 33.119$, $p < 0.00001$; Negidal vs. Even ‘dry’: $\chi^2 = 4.5207$, $p = 0.03$ (also for Fisher’s exact test, $p = 0.02$); Lamunkhin vs. Bystraja ‘burn’: $\chi^2 = 24.001$, $p < 0.00001$. However, one should keep in mind that usage patterns depend to a large extent on the topic of the text as well as speaker idiosyncrasies, and it is possible that the numbers would change if one were to include a wider range of texts and more speakers.

¹⁰These values do not differ significantly: $\chi^2 = 0.20807$, $p = 0.6483$.

To summarize this section, the causal-noncausal alternations in Negidal and Even confirm the form-to-frequency hypothesis formulated by Haspelmath et al. (2014: 597): most verbs in our sample support the tendency that the derived member of a pair is rarer and the basic one is more frequent. However, some verbs which do not support this hypothesis turn out to be crucial for another prediction, namely that verbs which are higher on the causative prominence scale tend to have a higher ratio of noncausal usage, irrespective of their language-specific coding. The Negidal data support this tendency, whereas the Even data rather contradict it. In both Even and Negidal, as in some languages of South America and Papunesia, the alternation pattern for ‘boil’ deviates from the expected one: this verb has a high rank on the causative prominence scale, but shows a low ratio of noncausal usage. This might be due to the characteristics of the boiling event, which generally needs to be initiated by a human causer, but which manifests itself only after a considerable amount of time.

7 Conclusions

To summarize, the Northern Tungusic languages have a strong preference for morphological marking of the causal-noncausal alternation, with equipollence being a particularly salient strategy for verbs of destruction in Even and Negidal. Ambitransitivity and suppletion, in contrast, are very rare. This observation fits well with the fact that these languages are morphologically rich and express all manner of derivations with a variety of morphemes.

At a broad level the causal-noncausal alternation is fairly stable across languages, as shown by the similarity of the coding patterns found in the Tungusic and especially the Turkic languages. This stability also emerges in the general Asian preference for causativization. However, at a fine-grained level many language-specific particularities emerge, as seen in the different patterns found for cognate verbs in the Tungusic languages, or in the widely different strategies preferred by the Japonic lects included in the WATP dataset.

Lastly, it should be noted that comparative work on the causal-noncausal alternation is rendered quite difficult due to the big impact that the choice of verb meanings and coding decisions can have; the cross-linguistic comparison discussed here should therefore be taken with a grain of salt. For instance, the comparison of our data with those of Nichols (2018) has shown that the choice of verb meanings included in the study can have a notable impact on the preferred coding patterns determined for individual languages. Furthermore, it is not clear whether different studies always collected the same translation equivalents for

all verb meanings, as seen by the fact that in our study we used ‘move (animate being)’, i.e. ‘go’ rather than ‘move (inanimate object)’, or that Kazama obtained the translation equivalent of ‘go.out (exit)’ instead of ‘go.out (extinguish)’. In addition, coding decisions can also play a big role in the resulting overall pattern frequencies. Nevertheless, we hope that the overview of causal-noncausal alternations in Northern Tungusic languages presented here can add some valuable observations about these understudied varieties to the areal and cross-linguistic research on this interesting feature.

Abbreviations

Even, Evenki, and Negidal are abbreviated as Evn, Evk, and Neg, respectively. Russian and Sakha (Yakut) copies are indicated with R and Y. Grammatical abbreviations used in the glosses are:

1, 2, 3	person	HORT	hortative
ABL	ablative	IMP	imperative
ACC	accusative	IN	inclusive
ADD	additive	INCH	inchoative
ADJR	adjectivizer	INS	instrumental
ADVRS	adversative	INT	intensive
ALL	allative	INTR	intransitive
ANT	anterior	ITER	iterative
AUG	augmentative	LOC	locative
CAUS	causative	MED	medio-passive
COND	conditional	MULT	multiplicative
CONTR	contrastive	NEG	negative
CVB	converb	NFUT	non-future
DAT	dative	PL	plural
DEF	definite	POSS	marker of non-canonical possession
DEONT	deontic future		
DIM	diminutive	PRFL	reflexive-possessive
DIST	distal (demonstrative)	PST	past
DP	discourse particle	PTCP	participle
EX	exclusive	PTL	(unspecified) particle
FOC	focus	PX	possessive suffix
FUT	future	REFL	reflexive
GNR	generic	REP	refractive (repetitive)
HAB	habitual	SG	singular

2 *The causal-noncausal alternation in Northern Tungusic languages*

SIM	simultaneous	TR	transitive
SMLF	semelfactive	VAL	valency-changing suffix
TAM	(unspecified) TAM-marker (1 and 2 identify two different morphemes)	VR	verbalizer

Acknowledgements

Very different versions of this paper were presented at the “Atelier morphosyntaxe” at the research unit “Dynamique du Langage”, Lyon, France, on 6 April 2018, and at the “Conference on Uralic, Altaic and Paleoasiatic Languages in the memory of A.P. Volodin” held at the Institute of Linguistic Studies RAS, Saint-Petersburg, on 6 December 2018. We thank the audiences of both events for their comments. We are also grateful to an anonymous reviewer and most especially to Andreas Hölzl and Tom Payne, whose detailed comments helped us improve the paper; needless to say, any remaining errors are our sole responsibility. Furthermore, we thank Galina I. Kandakova and Antonina V. Kazarova for their help in compiling the Negidal dataset. This paper was written in 2019 when Natalia Aralova was a post-doctoral researcher at Dynamique du Langage funded by the Endangered Languages Documentation Programme (ELDP), www.eldp.net. We are very grateful to ELDP for their generous support of our work on Negidal. We also thank the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR).

We would furthermore like to express our gratitude to Elena Perexval’skaya (Udihe), Andreas Hölzl (Manchu), and Sonya Oskol’skaya (Nanai) for checking the WATP datasets and for adding important information, such as synonymous verbal pairs.

Appendix A Causal-noncausal verb pairs in Even

In the following tables, transitivity is abbreviated as “TR” (+: transitive, -: intransitive) and the coding pattern as “Coding”.

Verb meaning		TR	Even verb	Coding
English	Russian			
boil	кипеть	-	huj	nC > C
	кипятить	+	huju:, hjuken	
break	(с)ломаться	-	hajubna, butar	nC < C
	(с)ломать	+	haju:, but	
	(с)ломаться	-	kabar, hokar, кабар, коңдар	nC ≈ C
	(с)ломать	+	kabal, hokak, кабак, коңдак	
burn	гореть	-	dur	nC > C
	сжечь	+	duru:, duruken	
close	закрыться	-	homab	nC < C
	закрыть	+	hom	
	закрыться	-	niṙkəb	nC < C
	закрыть	+	niṙkə	
	закрыться	-	niṙku	nC ≈ C
	закрыть	+	niṙkə	
run out/ use up	израсходоваться израсходовать	- +	manu: man	nC < C
dry	сушиться (сохнуть)	-	olga	nC ≈ C
	сушить	+	olgi	
fall/ drop	падать уронить	- +	tik tikəβ ^a	nC > C

^aIt should be noted that we do not find the form *tikəβ-* in our Even corpus, where we find only *tikuken-*, derived with the causative suffix *-βkAn*. The Even dictionaries don't let us determine whether *tikəβ-* indeed has only the basic meaning 'drop', but we assume so, since *tikuken-* adds specific semantics of a voluntary, intentional action.

2 The causal-noncausal alternation in Northern Tungusic languages

Verb meaning		TR	Even verb	Coding
English	Russian			
get wet/make wet, soak	промокнуть	-	ulab	nC < C
	замочить	+	ul	
go out/put out	погаснуть	-	hi:β	nC > C
	потушить	+	hi:βi:, hiβu:ken ^a	
increase	прибавиться	-	ha:βu	nC < C
	прибавить	+	ha:β	
melt	растаять	-	un, nen	nC > C
	растопить	+	umke, nemkat	
move (go)	идти	-	ɲən, hor/ur	nC > C
	везти	+	ɲənu:, horu/uru	
open	открыться	-	a:ɲa:b	nC < C
	открыть	+	a:ɲa:	
rise (ascend)/ raise (carry up)	подниматься	-	ugəɾəb	nC < C
	поднять	+	ugəɾ	
	подниматься	-	oɟtɕi	nC > C
	поднять	+	oɟtɕiβkan	
split	расколоться	-	kaltar	nC ≈ C
	расколоть	+	kalti:, kaltal	
spoil	испортиться	-	hojib, ha:ju:b	nC < C
	испортить	+	hoj, ha:ju:	
	испортиться	-	mun	nC > C
	испортить	+	munuken	
spread	расстилаться	-	girkəb	nC < C
	расстилать	+	girkə	

^aNote that we cannot be fully certain that the form *hiβu:ken-* does not add any additional semantic component, since we do not find this in our Even corpus, and the dictionaries do not let us determine the precise meaning.

Verb meaning		TR	Even verb	Coding
English	Russian			
stop (of humans)	остановиться	-	il	nC > C
	остановить	+	ilu:kan	
turn over (around)	повернуться, перевернуться	-	hukəlbəŋtɕi	nC ≈ C
	повернуть, перевернуть	+	hukəsən	
	повернуться, перевернуться	-	kumərkin	nC > C
	повернуть, перевернуть	+	kumərkinke:n	
bend (twist)	скрутиться (согнуться)	-	uɕeib	nC < C
	скрутить (согнуть)	+	ut	

Appendix B Causal-noncausal verb pairs in Negidal

Verb meaning		TR	Negidal verb	Coding
English	Russian			
boil	кипеть	-	huj	nC > C
	кипятить	+	hujuβ	
break	(с)ломаться	-	tɕaptɕaβ	nC < C
	(с)ломать	+	tɕaptɕa	
	(с)ломаться	-	tonŋodgə, tonŋam, kilgədgə, kilgam, boktadga, boktam	nC ≈ C
	(с)ломать	+	tonŋol, kilgəl, boktal, boktana:	
burn	гореть	-	ʒəgdə	nC ≈ C
	сжечь	+	ʒəgdi	
close	закрыться	-	samuβ	nC < C
	закрыть	+	sam	
run out/ use up	израсходоваться	-	manaβ	nC < C
	израсходовать	+	mana	

2 The causal-noncausal alternation in Northern Tungusic languages

Verb meaning				
English	Russian	TR	Negidal verb	Coding
dry	сушиться (сохнуть)	-	olgo	nC ≈ C
	сушить	+	olgi	
fall/ drop	падать	-	tik	nC ≈ C
	уронить	+	tibgu	
get wet/make wet, soak	промокнуть	-	olap	nC < C
	замочить	+	ola	
go out/ put out	погаснуть	-	siβ (corpus)	nC > C
	потушить	+	siβi/siβu (corpus)	
	погаснуть	-	siβ (elicitation)	nC = C
	потушить	+	siβ (elicitation)	
increase	прибавиться	-	haβup	nC < C
	прибавить	+	haβ	
melt	растаять	-	un	nC > C
	растопить	+	uniβkan	
move (go)	идти	-	ŋənə	nC > C
	везти	+	ŋənəβ	
open	открыться	-	niβ, a:ŋaβ	nC < C
	открыть	+	ni:, a:ŋa	
rise (ascend) / raise (carry up)	подниматься	-	ugi	nC = C
	поднять	+	ugi	
	подниматься	-	tukti	nC > C
	поднять	+	tuktiβ	
split	расколоться	-	dəlpədgə, dəlpam, kaltadga, kaltam	nC ≈ C
		+	dəlpəl, kaltana:, kaltal	
	расколоть	-	dəlpədgə	nC > C
		+	dəlpədgəβkan	

Verb meaning				
English	Russian	TR	Negidal verb	Coding
spoil	испортиться	–	hajiβ / hajip	nC < C
	испортить	+	haji	
	испортиться	–	hajiβ / hajip	nC ≈ C
	испортить	+	haju	
spread	расстилаться	–	gildeβ	nC < C
	расстилать	+	gilde	
stop (of humans)	остановиться	–	el	nC > C
	остановить	+	eleβkan	
turn over (around)	повернуться, перевернуться	–	kumtəβ, kumtədgə	nC < C
	повернуть, перевернуть	+	kumtə	
	повернуться, перевернуться	–	hukil	nC = C
	повернуть, перевернуть	+	hukil	
bend (twist)	скрутиться (согнуться)	–	otæβ	nC < C
	скрутить (согнуть)	+	ot	

Appendix C Causal-noncausal verb pairs in Evenki

Verb meaning				
English	Russian	TR	Evenki verb	Coding
boil	кипеть	-	huju	nC > C
	кипятить	+	hujuβ	
break	(с)ломаться	-	kapurga, suktɕarga	nC < C
	(с)ломать	+	kapu, suktɕa	
burn	гореть	-	ɟegdə	nC ≈ C
	сжечь	+	ɟegdi:	
	гореть	-	ilaβ	nC < C
	сжечь	+	ila	
	гореть	-	lurgi	nC = C
	сжечь	+	lurgi:	
	гореть	-	badara	nC > C
	сжечь	+	badaran	
close	закрыться	-	so:miβ	nC < C
	закрыть	+	so:m	
run out / use up	израсходоваться	-	manaβ	nC < C
	израсходовать	+	mana	
dry	сушиться (сохнуть)	-	olgo	nC ≈ C
	сушить	+	olgi	
fall / drop	падать	-	tik	nC > C
	уронить	+	tikiβ	
	падать	-	buru	nC ≈ C
	уронить	+	huri:	
get wet / make wet, soak	промокнуть	-	ulap	nC < C
	замочить	+	ula:	
go out / put out	погаснуть	-	si:β	nC < C
	потушить	+	si:	
increase	прибавиться	-	ha:βuβ	nC < C
	прибавить	+	ha:β	

Verb meaning				
English	Russian	TR	Evenki verb	Coding verb
melt	растаять	-	u:n , tɕu:m	nC > C
	растопить	+	u:nɲi:, tɕu:mɲi:	
move (go)	идти	-	suru, ɲənə	nC > C
	везти	+	suruβ, ɲənəβ	
open	открыться	-	ni:β, a:ɲa:β	nC < C
	открыть	+	ni:, a:ɲa:	
rise (ascend) / raise (carry up)	подниматься	-	ugi:riβ	nC < C
	поднять	+	ugi:r	
	подниматься поднять	- +	tukti tuktiβ	nC > C
split	расколоться	-	dəlpərgə	nC ≈ C
	расколоть	+	dəlpəli:	
spoil	испортиться	-	munu	nC ≈ C
	испортить	+	muni:	
spread	расстилаться	-	səktəβ	nC < C
	расстлать	+	səktə	
stop (of humans)	остановиться	-	il, tu:ri:n	nC > C
	остановить	+	iliβka:n, turinmu, tu:ri:n- mukə:n	
turn over (around)	повернуться, перевернуться	-	horol	nC > C
	повернуть, перевернуть	+	horoliβka:n	
	повернуться, перевернуться	-	kumtəβ	nC < C
	повернуть, перевернуть	+	kumtə	
bend (twist)	скрутиться (согнуться)	-	uɕi:β, mataβ	nC < C
	скрутить (согнуть)	+	uɕi:, mata	

Appendix D Corpus frequencies and coding patterns

Corpus frequencies and coding patterns for 20 verbs (beginning with the 12 that overlap with Haspelmath et al. 2014); conf.: confirmed, freq.: frequency, equi.: equipollent, antiC: anticausative, caus: causative.

meaning	Even		Negidal		Hypothesis conf.?	
	freq.	coding	freq.	coding	Even	Negidal
split.intr	4	equi	1	equi	n/a	n/a
split.tr	5		4			
close.intr	4	equi	0	antiC	n/a	yes
close.tr	22		8			
break.intr	7	equi	16	equi	n/a	n/a
break.tr	7		9			
open.intr	6	antiC	9	antiC	yes	yes
open.tr	25		43			
rise	52	caus	71	caus	yes	yes
raise	1		11			
burn.intr	35	caus	62	equi	no	n/a
burn.tr	71		19			
turn.intr	4	caus	7	antiC	yes	no
turn.tr	0		1			
stop.intr	6	caus	5	caus	yes	yes
stop.tr	0		0			
melt.intr	4	caus	4	caus	yes	yes
melt.tr	0		1			
go.out	9	caus	8	caus	yes	yes
put.out	2		6			
dry.intr	4	equi	34	equi	n/a	n/a
dry.tr	9		16			
boil.intr	7	caus	9	caus	no	no
boil.tr	23		19			

meaning	Even		Negidal		Hypothesis conf.?	
	freq.	coding	freq.	coding	Even	Negidal
run.out	25	antiC	22	antiC	no	no
use.up	10		11			
fall	118	caus	95	caus	yes	yes
drop	0		3			
get.wet	12	antiC	11	antiC	no	yes
make.wet	7		31			
increase.intr	0	antiC	0	antiC	yes	(yes)
increase.tr	9		1			
move(go)	886	caus	886	caus	yes	yes
make.move	147		107			
spoil.intr	2	caus			(yes)	
spoil.tr	0					
spread.intr	1	antiC	0	antiC	yes	(yes)
spread.tr	5		2			
bend.intr	2	antiC	1	antiC	(yes)	yes
bend.tr	3		9			
Total YES					10 (12)	10 (12)
Total NO					4	3

References

- Atknine, Victor. 1997. The Evenki language from the Yenisei to Sakhalin. *Senri Ethnological Studies* 44. 109–121.
- Benzing, Johannes. 1955. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik* (Akademie der Wissenschaften und der Literatur. Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 11). Wiesbaden: Akademie der Wissenschaften und der Literatur.
- Boldyrev, Boris V. 1994. *Russko-evenkijskij slovar' [Russian-Evenki dictionary]*. Novosibirsk: Nauka.
- Boldyrev, Boris V. 2000. *Evenkijsko-russkij slovar' [Evenki-Russian dictionary]*. 2 vols. Novosibirsk: Izdatel'stvo SO RAN, filial "GEO".

2 The causal-noncausal alternation in Northern Tungusic languages

- Boldyrev, Boris V. 2007. *Morfologija evenkijskogo jazyka* [The morphology of Evenki]. Novosibirsk: Nauka.
- Bulatova, Nadezhda & Lenore Grenoble. 1999. *Evenki*. Munich: Lincom Europa.
- Cincius, Vera I. 1982. *Negidal'skij jazyk: Issledovanija i materialy*. [Negidal: Analysis and materials]. Leningrad: Nauka.
- Cincius, Vera I. & Ljubov' D. Rišes. 1952. *Russko-evenkij slovar'* [Russian-Even dictionary]. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovarej.
- Comrie, Bernard. 2006. Transitivity pairs, markedness, and diachronic stability. *Linguistics* 44(2). 303–318.
- Creissels, Denis. 2018. *The noncausal / causal alternation*. Talk presented at the research unit Dynamique Du Langage, Lyon (France) within the seminar series “Atelier morphosyntaxe” on 26 January 2018.
- Creissels, Denis. to appear. The noncausal/causal alternation and the limits of ambitransitivity in a sample of sub-Saharan languages. *Linguistique et langues africaines*.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. *Glottolog 4.0*. Jena: Max Planck Institute for the Science of Human History. (Available at <http://glottolog.org>, Accessed 2019.10.04.)
- Haspelmath, Martin. 1993. More on the typology of inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (eds.), *Causatives and transitivity* (Studies in Language Companion Series 23), 87–120. Amsterdam: John Benjamins Publishing Company.
- Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2). 33–63.
- Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyacı. 2014. Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics* 50(3). 587–625.
- Hovav, Malka Rappaport. 2014. Lexical content and context: The causative alternation in English revisited. *Lingua* 141. 8–29.
- Jang, Taeho & Thomas E. Payne. 2014. Conspiring motivations for causative and passive isomorphism: Data from Xibe. In Pirkko Suihkonen & Lindsay Whaley (eds.), *On diversity and complexity of languages spoken in Europe and North and Central Asia* (Studies in Language Companion 164), 91–108. Amsterdam: John Benjamins Publishing Company.
- Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.

- Kazama, Shinjiro. 2004. On the causative forms in Tungus languages. In Carsten Naehrer (ed.), *Proceedings of the First International Conference on Manchu-Tungus studies*, vol. 2, 73–92. Wiesbaden: Harrassowitz.
- Kazenin, Konstantin I. 2001. The passive voice. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2 (Handbooks of Linguistics and Communication Science (HSK) 20), 899–916. Berlin: De Gruyter Mouton.
- Khasanova, Marina & Alexander M. Pevnov. 2003. *Myths and tales of the Negidals* (ELPR Publication Series A2-024). Osaka: Nakanishi.
- Konstantinova, Olga A. 1964. *Evenkijskij jazyk: Fonetika. morfologija. [Evenki: Phonetics. Morphology]*. Moscow: Nauka.
- Koontz-Garboden, Andrew. 2009. Anticausativization. *Natural Language & Linguistic Theory* 27(1). 77–138.
- Levin, Beth. 2015. Semantics and pragmatics of argument alternations. *Annual Review of Linguistics* 1(1). 63–83.
- Levshina, Natalia. 2016. Finding the best fit for direct and indirect causation: A typological study. *Lingua Posnaniensis* 58(2). 65–82.
- Li, Fengxiang & Lindsay Whaley. 2012. The grammaticization cycle of causatives in Oroqen dialects. In Andrej L. Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica), 165–180. Wiesbaden: Harrassowitz.
- Malchukov, Andrei L. 1995. *Even* (Languages of the World/ Materials 12). München: Lincom Europa.
- Myl'nikova, Klavdija M. & Vera I. Cincius. 1931. Materialy po issledovaniju negidal'skogo jazyka [Materials for the study of Negidal]. *Tungusskij sbornik* 1. 107–218.
- Myreeva, Anna N. 2004. *Evenkijsko-russkij slovar' [Evenki-Russian dictionary]*. Novosibirsk: Nauka.
- Nedjalkov, Igor. 1997. *Evenki* (Descriptive Grammars). London: Routledge.
- Nedjalkov, Igor. 2013. *Verb valency classes in Evenki in the comparative perspective*. Available at https://www.eva.mpg.de/lingua/conference/2013_valency_classes/pdf/NedjalkovValency.pdf (Accessed 2019.09.25) (Unpublished manuscript.)
- Nedjalkov, Igor' V. 1993. Causative-passive polysemy of the Manchu-Tungusic -bu/-v(u). *Linguistica antverpiensia* 27. 193–202.
- Nedjalkov, Igor' V. 2014. Grammatikalizacija passiva i dekauzativa v tungusomančžurskix jazykax [Grammaticalization of passive and decausative in

2 The causal-noncausal alternation in Northern Tungusic languages

- Tungus-Manchu languages]. *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies* 10(3). 473–494.
- Nichols, Johanna. 2018. Non-linguistic conditions for causativization as a linguistic attractor. *Frontiers in Psychology* 8. 2356.
- Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8(2). 149–211.
- Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct northern Tungusic language of the Lower Amur*. Endangered Languages Archive. Available at <http://hdl.handle.net/2196/b644db81-725c-4031-935c-f33c763df152>. (Accessed 2019.10.08.)
- Pakendorf, Brigitte & Natalia Aralova. 2018. The endangered state of Negidal: A field report. *Language Documentation and Conservation* 12. 1–14.
- Pakendorf, Brigitte & Natalia Aralova. 2020. Even and the Northern Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 288–304. Oxford: Oxford University Press.
- Palmer, Frank R. 1994. *Grammatical roles and relations* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Pevnov, Aleksandr M. 2007. Glagol'noe osnovnoizmenenie v negidal'skom jazyke [Verbal base formation in Negidal]. *Acta Linguistica Petropolitana* 3(3). 191–244.
- Plank, Frans & Aditi Lahiri. 2015. Macroscopic and microscopic typology: Basic valence orientation, more pertinacious than meets the naked eye. *Linguistic Typology* 19(1). 1–54.
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Robbek, Vasiliy A. & Maria E. Robbek. 2005. *Evensko-russkij slovar' [Even-Russian dictionary]*. Leningrad: Nauka.
- Samardžić, Tanja & Paola Merlo. 2012. The meaning of lexical causatives in cross-linguistic variation. *Linguistic Issues in Language Technology* 7(12). 1–14.
- Schäfer, Florian. 2009. The causative alternation. *Language and Linguistics Compass* 3(2). 641–681.
- Seifart, Frank, Stefan Schnell, Maria Bardají i Farré, Christian Döhler, Sonja Gipper, Katja Hannß, Katharina Haude, Birgit Hellwig, Anna Margetts, Sonja Riesberg, Harriet Sheppard, Nick Thieberger, Claudia Wegener, Catharina Williams van Klinken & Nikolaus P. Himmelmann. 2019. *Typological causative prominence and causal verb text frequencies: Refinements based on language documentation corpora*. Talk presented at the 13th Conference of the Association for Linguistic Typology, 2019.09.04-06, Pavia.

- Sunik, O.P. 1962. *Glagol v tunguso-man'čžurskix jazykax*. [*The verb in the Tungus-Manchu languages*]. Moscow: Izdatel'stvo Akademii nauk SSSR.
- The World Atlas of Transitivity Pairs*. 2014. Tokyo: National Institute for Japanese Language and Linguistics. Available at <http://verbpairmap.ninjal.ac.jp>. (Accessed 2019.09.27).
- Zúñiga, Fernando & Seppo Kittilä. 2019. *Grammatical voice* (Cambridge Textbooks in Linguistics). Cambridge, New York: Cambridge University Press.

Chapter 3

Tense and insubordination in Uilta (Orok)

Patryk Czerwinski

University of Mainz

The paper describes the tense category in Uilta, a critically endangered Tungusic language, from a functional and diachronic perspective. The functional analysis, based on the author's fieldwork, provides a comprehensive typological description of the Uilta tense system. Similarly to other Tungusic languages, the diachronic development of this system and its current shape and complexity are largely the result of the processes of insubordination (replacement of finite verbal forms by non-finite forms in predicative use).

1 Introduction

1.1 The purpose and scope of this paper

The paper offers a comprehensive functional analysis of the tense system of the Tungusic language Uilta (Orok), based largely on the author's own fieldwork, and partially on existing descriptions.¹ The previous descriptions of the Uilta tense system are either incomplete or contradicting, partially due to dialectal differences, as well as diachronic changes. The present analysis aims to account for those differences through different degrees and stages of the processes of insubordination (cf. Evans 2007).

Insubordination, the development of non-finite (participial) into finite (verbal) forms, is a prominent factor in the development of the TAM systems of Tungusic

¹Uilta is the endonym and is strongly preferred by the community over the exonym Orok. Both terms are used in the literature.



languages, which underwent repeated cycles of renewal of finite verbal forms through participles (Malchukov 2013). It will be shown that similar diachronic processes account for the current shape and peculiarities of the Uilta tense system.

1.2 Basic information about Uilta

While there is no universally accepted internal classification of Tungusic, most authors agree on placing Northern Tungusic (represented by Even and Evenki) and Southern Tungusic (the Jurchen/Manchu group) in separate branches, with the remaining groupings, Udegheic and Nanaic, variously assigned to one of the two branches, to a separate (Southeastern) branch, or to branches of their own (Whaley & Oskolskaya 2020). Uilta is a member of the Nanaic (sub-)branch. It is spoken exclusively on the island of Sakhalin, in the Russian Federation. This relative isolation from the rest of the family led to the development of a number of innovations not attested in the languages spoken on the mainland (Pevnov 2016).

The two Uilta dialects, Northern and Southern, are mutually intelligible and historically formed a dialect continuum. The language is critically endangered, with five fluent speakers remaining, all in their seventies, of which four are speakers of the Northern dialect, centered on the village of Val in the Nogliki raion, and one of the Southern dialect, in the city of Poronaysk.²

Uilta has been in close areal contact with Sakhalin Nivkh for at least 300 years (Yamada 2010a), and shares numerous features in the lexical and, to a lesser degree, grammatical domain (Pevnov 2016). Much later, from the mid-19th century onwards, it came into contact with Sakhalin Evenki, a later entrant in the northern part of Sakhalin (Yamada 2010a). Contact with Sakhalin Evenki accounts for a number of distinct features of the Northern dialect compared to the Southern dialect of Uilta (Ikegami 2001 [1994]).

1.3 Insubordination in Tungusic

“Canonical” insubordination, as introduced into linguistic typology by Evans (2007), involves “conventionalized main clause use of what, on prima facie grounds, appear to be formally subordinate clauses” (Evans 2007: 367). A variant

²Historically, different Uilta clans lived as reindeer herders along different rivers on the east coast of central and northern Sakhalin, and migrated yearly between the coast and the mountains in the central part of the island. They were forcibly settled in the 1950s, around a collective farm in Val, Nogliki raion, and Yuzhnyj ostrov, Poronaysk.

of this process, labelled “verbalisation” in Malchukov (2013), involves reanalysis of a nominal (participial) predicate into a verbal predicate. Both scenarios are illustrated below for Even (Northern Tungusic), after Malchukov (2013).

Insubordination “proper”: Reanalysis of a sentential argument as a main clause:

[S PART-AGR.POSS] [COP] → [S PART-AGR.POSS] ∅ → [S] [V-AGR.POSS]

(1) Even

- a. [*Bej-il hör-ri-ten*] *bi-d'i-n.*
 man-PL go-NFUT(PART)-3PL(POSS) be-FUT-3SG
 ‘The men probably left.’ (Literally: ‘The men’s leaving will be.’)
- b. *Bej-il hör-ri-ten.*
 man-PL go-PST-3PL(POSS)
 ‘The men left.’ (Malchukov 2013: 182)

Verbalization: Reanalysis of a nominal predicate into a verbal predicate:

[S] [N/PART] [COP] → [S] [V2 AUX] (→ [S] [V]).

(2) Even

- a. *Bej [hör-če] [bi-si-n].*
 man go-PFV be-PST-3SG
 ‘The man was gone.’
- b. *Bej [hör-če bi-si-n].*
 man go-PFV be-PST-3SG
 ‘The man had left.’ (Malchukov 2013: 181)

In the first scenario, the subject complement clause followed by the existential verb is reanalysed as an independent clause. Typically for Tungusic, the non-finite complement clause has the form of a nominal possessive phrase. Possessive agreement on the participle indicates the subordinate subject. In (1b), the same participial form now forms the predicate of the verbal clause, but retains the (nominal) possessive subject agreement. In the second scenario, the nominal (participial) predicate followed by the existential verb is reanalysed as a periphrastic verbal (pluperfect) construction.

The two processes exemplified above for Even led to gradual replacement of finite TAM forms by forms of participial origin in the verbal paradigms of all branches of Tungusic, and account for a number of peculiarities in their grammatical structures: weak distinction between nominal and verbal forms; inherent

ambiguity of certain tense forms despite rich inventories of distinct markers; the presence of nominal (possessive) agreement paradigms in the (finite) verbal domain (Malchukov 2013).

Furthermore, as demonstrated by Robbeets (2009, 2015) and Malchukov & Czerwinski (2020), the process of replacement of finite forms by participles in repeated cycles of insubordination is prevalent in all “Macro-Altaic”³ languages, and its preponderance can be viewed as one of their characteristic features.

More broadly, as demonstrated by Malchukov (2013) and Malchukov & Czerwinski (2021), this tendency is not limited to “Macro-Altaic”, and instead constitutes an areal feature (diachronic isogloss) of Siberian languages generally, including Paleosiberian (Chukotko-Kamchatkan, Eskimo-Aleut, Nivkh, Yeniseian and Yukaghir) and Uralic languages.

The gradual replacement of finite (verbal) through non-finite (participial) forms leads to competition between old and new forms, often resulting in functional shifts in the relevant verbal categories. This is well documented for South-eastern Tungusic languages (i.e. Udegheic and Nanaic, see §1.2 above), which all retain forms of both finite and participial origin, to varying degrees. As the imperfective and perfective participles acquire predicative function and general present/past meaning respectively, the erstwhile finite forms are pushed out from general present/past use and acquire direct evidential, and later affirmative-emphatic, meaning through a process known as markedness reversal (Croft 2002 [1990]). In the past domain, the development from resultative through perfect to (indirect evidential) past is a universal grammaticalisation path, well-attested cross-linguistically (Bybee et al. 1994). Competition between forms at each stage leads to further development from perfect to (non-witnessed) past to general past, and the parallel development of erstwhile finite forms from (unmarked) indicative first into direct evidential, and later into affirmative-emphatic. Different Tungusic languages display different stages of this development. This is illustrated below for Southeastern Tungusic (Udegheic and Nanaic; Figure 1, adapted from Malchukov 2000: 454).⁴

This competition between forms, with the resulting functional shifts, occurred in Uilta in all three temporal domains, past, present and future, and is a key factor in understanding both the diachronic development and the current shape of the Uilta tense system.

³Here and elsewhere, “Macro-Altaic” is used as an areal-typological label, without any claims regarding genetic relatedness of the families in question (Turkic, Mongolic, Tungusic, Koreanic and Japonic).

⁴The figure in Malchukov (2000) listed Uilta as representing the final, fourth stage, based on a previous description. It was modified to reflect the fact that the finite past form is marginally retained in Uilta, as per other descriptions and as confirmed by the present author (see §3.3 below).

3 Tense and insubordination in Uilta (Orok)

	1 st stage		2 nd stage		3 rd stage		4 th stage
Participial:	(Resultative)	>	Perfect/indirect evid.	>	Preterite	>	General past
Finite:	General past	>	Imperfect/direct evid.	>	Validational	>	∅
			Udihe		Nanai		

Figure 1: Evolution of past tense forms in Southeastern Tungusic (adapted from Malchukov 2000: 454)

§2 of the paper outlines the Uilta tense system. §2.2 lists previous descriptions, with the relevant information on the attested forms, the period of data collection and the dialect they pertain to. §3, §4 and §5 provide functional analysis of the past, present and future tense forms respectively, as well as their diachronic development through different scenarios of insubordination. §6 provides a summary and conclusions.

2 Uilta tense system

2.1 Overview

The contemporary Uilta tense system consists of nine (Northern dialect) or eight forms (Southern dialect; the general future form in *-li* is attested only in the Northern dialect). They are listed below according to their origin. The forms in the right-hand column are the old finite forms. They are mono-functional, i.e. can be used exclusively as the predicate of a main clause, and take subject agreement of the verbal type (see below). The forms in the left-hand column, grammaticalised from the perfective, imperfective and future participles, are poly-functional (retain their function as participles/nominalisations on top of their function as the main clause predicate), and take agreement of the nominal (possessive) type.

2.2 Existing descriptions

Existing descriptions of tense in Uilta go back over a hundred years (see Yamada 2013 for a comprehensive overview). Table 2, adapted from Yamada (2013: 90), lists them all, specifying which dialect they pertain to, the period of data collection, and the forms attested.

Table 1: Tense forms in Uilta

	Participial	Finite
Past	general past in <i>-xAn</i> pluperfect in <i>-xA- bi-čči</i>	direct evid./affirm.-emph. past in <i>-tAA</i>
Present	general present in <i>+RI</i>	direct evid./emph./mirative present in <i>+RAkkA</i>
Future	general future in <i>-li</i> (N dialect) probable future in <i>+Rlli</i>	immediate spontaneous future in <i>+RIIA</i> likely/anticipated future in <i>+RAηA</i>

Table 2: Existing descriptions of tense in Uilta (adapted from Yamada 2013: 90)

Author	Period	Dialect	Past	Present	Future
Piłsudski (Majewicz 2011)	1900	S	<i>-xAn</i>	<i>+RI</i>	–
Nakanome 1917	1910	S	<i>-xAn</i>	<i>+RI</i>	–
Magata 1981	1920–1930	S	<i>-xAn</i> <i>-tAA</i>	<i>+RI</i> <i>+RAkkA</i>	<i>+RIIA</i>
Petrova 1967	1930–1940	N	<i>-xAn</i>	<i>+RI</i>	<i>-llee</i> <i>+RIIA</i>
Ikegami 2001 [1959]	1940–1950	S	<i>-xAn</i> <i>-tAA</i>	<i>+RI</i> <i>+RAkkA</i>	<i>+RIIA</i> <i>+RAηA</i> <i>+Rlli</i>
Tsumagari 2009	1940–1980	S	<i>-xAn</i> <i>-tAA</i>	<i>+RI</i> <i>+RAkkA</i>	<i>+RIIA</i>
Ozolinja 2013	1990–2000	N	<i>-xAn</i>	<i>+RI</i>	<i>+RIIA</i>

The next three sections describe the category of tense in Uilta from a functional perspective, based on the author's own fieldwork.⁵ The analysis by the author will be reconciled with existing descriptions, particularly with regard to diachronic development. It will be shown that this development is best explained through the processes of insubordination. This part is based on and expands on the work on insubordination in Tungusic by Malchukov (2000, 2013).

3 Past domain

In the past domain, Uilta has three forms, general past in *-xAn*, pluperfect in *-xA- bi-čči* [-PST-AGR be-PFV], and direct evidential/affirmative-emphatic in *-tAA*. The forms in *-xAn* and *-xA- bi-čči* grammaticalised from the perfective participle in *-xAn*, and retain the person/number agreement paradigm of the nominal (possessive) type, in contrast with the form in *-tAA* which takes person/number agreement of the (finite) verbal type (Table 3).⁶

Table 3: Person/ number agreement paradigms of Uilta past tense forms.

Number	Person	<i>-xAn</i>	<i>-tAA</i>
Singular	1 st	<i>-xA-mbi</i>	<i>-tA-mmee</i>
	2 nd	<i>-xA-si</i>	<i>-tA-ssee</i>
	3 rd	<i>-xA-ni</i>	<i>-tAA</i>
Plural	1 st	<i>-xA-pu</i>	<i>-tA-ppOO</i>
	2 nd	<i>-xA-su</i>	<i>-tA-ssOO</i>
	3 rd	<i>-xA-či</i>	<i>-tAA-l</i>

3.1 General past in *-xAn*

The general past form in *-xAn* is by far the most frequent past form, with the other two forms limited to specific contexts (see §3.2 and §3.3 below). In some conjugational classes the perfective participle/general past tense takes the form *-či*. It is unclear whether the forms in *-xAn* and *-či* are cognate or heteroclitic (Malchukov 2000). The form in *-xAn* is used in recent (3) and remote past contexts (4), and with punctual (5), durative (6) and habitual meanings (7):

⁵Unless otherwise stated, Uilta data and findings come from the author's own fieldwork.

⁶Unlike most Tungusic languages, Uilta has no inclusive/exclusive first person plural distinction.

- (3) *Pakčira-du-xa-ni. Siweeskə-bi dagjita.*
 get.dark-REITER-PST-3SG candle-1SG.POSS burn(TRANS)+1SG.HORT
 ‘It got dark. Let me light a candle.’
- (4) *Bii nuuč-i-jji ηələ-xə-mpi ηinda-l-ji.*
 1SG little-INSTR.REFL fear-PST-1SG dog-PL-INSTR
 ‘When I was little, I was afraid of dogs.’
- (5) *Ča-du bi-čči-ndulə-ni bii əigə-təkki gumaaska*
 that-LOC be-PFV(PST)-LOC-3SG 1SG older.sister-ALL.REFL money
buu-xə-mpi.
 give-PST-1SG
 ‘Because he had lived there, I gave my sister money.’
- (6) *Tari ənu-či narree goroo daputa-xa-či okči-či-kku*
 this fall.ill-PFV man+ACC long.time+EMPH hold-PST-3PL heal-DUR-place
duku-du.
 house-LOC
 ‘They kept this sick man in the hospital for a long time.’
- (7) *Niməri-ηəssəə-wwee, mittəi aptauli-mba tøyə-xə-či.*
 visit-CONCUR.PST.CONV-1SG 1SG.ALL tasty-ACC treat-PST-3PL
 ‘When I visited [them], they always treated me to something tasty.’

It is also the form most often used in narratives, as in (8):

- (8) *Niiwənikəən balji-xa-ndulli xaali=ddaa suunəə*
 Niiwənikəən(PN) grow(INTR)-PFV(PST)-LOC.REFL how=FOC sun+ACC
ə-čči-ni ittəə.
 NEG.AUX-PST-3SG see+CONN
 ‘When he was growing up, Niiwənikəən never saw the sun.’

On top of its predicative use, the form in *-xAn* retains its original use as the perfective participle (which in all Tungusic languages has double adnominal/nominal function; example 9, cf. also examples 6, 23 and 29).⁷

- (9) *Tari puttə iiwu-xə-mbə-ni sundattaa əni-ni təldə-xə-ni.*
 that child bring.in-PFV-ACC-3SG fish+ACC mother-3SG.POSS fillet-PST-3SG
 ‘The mother filleted the fish that the son brought.’

⁷As in other Tungusic languages, participles are also the main strategy for relative clauses, both pre-nominal (cf. 42) and internally headed (9, 49), complement clauses (32) and, with oblique cases, one of the two strategies for adverbial clauses (5, 8, 12, 15, 39, 45).

In line with its origin as the perfective participle, while firmly established as a general past tense form in predicative use, in a limited number of cases the meaning of *-xA_n* is closer to the resultative or perfect than a pure tense form (Yamada 2013: 98):

- (10) *Nu, əsi=ləkə dəgdə-xə-či əmbee.*
 INTJ now=TOP burn-PST-3PL of.course
 ‘Well, now they have burnt of course.’ (Yamada 2013: 99)

3.2 Pluperfect in *-xA- bi-čči*

The perfective participle form in *-xA_n* followed by the copula/existential verb in the past tense forms the periphrastic pluperfect, similar to other Tungusic languages:

- (11) *Buu gasa-ttai-ppoo gubernator sinda-xa-ni bi-čči.*
 1PL village-ALL-1PL.POSS governor come-PST-3SG be-PFV
 ‘A governor had come to our village. [He had already left.]’

In Uilta, with atelic verbs, the same form can also be used to express past progressive meaning:

- (12) *Bii gyauli-du-wwee bii mapa-ŋu-bi eekkuta-xa-ni*
 1SG row+IPFV(PRES)-LOC-1SG 1SG old.man-AL-1SG.POSS steer-PST-3SG
bi-čči.
 be-PFV
 ‘While I was rowing, my husband was steering.’

Either the lexical verb or the copula can take subject agreement marking, i.e. both *sinda-xa-ni bi-čči* [come-PST-3SG be-PFV] and *sinda-xa bi-čči-ni* [come-PFV be-PST-3SG] are correct (but not **sinda-xa-ni bi-čči-ni* or **sinda-xa bi-čči*).

3.3 Direct evidential/affirmative-emphatic past in *-tAA*

The direct evidential/affirmative-emphatic past form in *-tAA* is marginal in present-day Uilta. It does not appear naturally in narratives or dialogue, and all attestations were obtained through elicitation. It is used overwhelmingly in the third, occasionally in the second, and very rarely in the first person. In the third person, its main use is direct visual evidential as in (13):

- (13) *Sii ŋinda-si bii nakku-ŋŋoo-wwee puktuu-təə.*
 2SG dog-2SG.POSS 1SG chicken-AL+ACC-1SG carry.away-DIREVID.PST.3
 ‘Your dog carried away my chicken.’ [The hearer cannot retort ‘it wasn’t my dog’ because the speaker saw it.]

It can also combine direct evidential with emphatic meaning as in (14):

- (14) *Ərəə, aya bara nari-sal. Əsi sinda-taa-l ulaa-ji.*
 INTJ very many people-PL now come-DIREVID.PST.3-PL reindeer-INSTR
 ‘Wow, how many people. They just came by reindeer.’

Rarely, it can be used purely emphatically, without clear evidential connotation (although not incompatible with it), as in (15):

- (15) *Seryozha uumbu-čči-du-ni sundatta tarttəə*
 Seryozha(PN) fish-PFV(PST)-LOC-3SG fish suddenly
iktəmə-təə.
 bite-DIREVID.PST.3
 ‘When Seryozha was fishing, a fish suddenly bit.’

It is overwhelmingly used in immediate past (just witnessed) contexts, with adverbs like *tarttəə* ‘there (emphatic), right now’. It is incompatible with indirect reported speech, only with direct reported speech as in (16):

- (16) *Sergei mittəi uč-či-ni: “Attaa, tari nari pastuuxi-tai*
 Sergei(PN) 1SG.ALL say-PST-3SG grandmother this man herder-ALL
ŋənə-təə”.
 go-DIREVID.PST.3
 ‘Sergei said to me: “Grandma, he left to join the reindeer herders”.’

In the second person, the form in *-tAA* has affirmative-emphatic meaning as in (17), typically reinforced by the emphatic use of the adverb *goči* ‘again, indeed’.

- (17) *Sii dəptu-tə-ssee goči!*
 2SG eat-DIREVID.PST-2SG EMPH
 ‘You have already eaten though!’

Finally, very rarely, the form in *-tAA* can also be used in the first person, also with affirmative-emphatic meaning as in (18).

- (18) *Buu təə-wu-tə-ppəə goči čaa duwa-du kartooskkaa.*
 1PL sit-TRANS-DIREVID.PST-1PL EMPH that summer-LOC potato+ACC
 ‘We did plant potatoes that summer.’

3.4 Diachronic development of Uilta past tense forms

Earlier descriptions of Uilta past tense forms (Ikegami 2001 [1959]; Tsumagari 2009) describe the finite form in *-tAA* as fully productive, with a complete person/number paradigm. Already at that stage it was restricted to direct evidential contexts (Ikegami 2001 [1959]), and as is clear from the above description, it has become even more restricted in present-day Uilta, with the participial form in *-xAn* used predicatively in almost all contexts.

Together with the fact that the form in *-xAn* retains resultative/perfect meaning (cf. example 10 above), this points to a diachronic development where the perfective participle gradually replaced the erstwhile finite form, through resultative and perfect stages. This mirrors the development observed in other Tungusic languages (cf. Malchukov 2000: 447), along a well-attested grammaticalisation path (Bybee et al. 1994: 105).

4 Present domain

In the present domain Uilta displays competition between two forms, the general present form in *+RI*,⁸ and the direct evidential/emphatic/mirative in *+RAkkA*. The form in *+RI* grammaticalised from the imperfective participle, while *+RAkkA* is the original finite form.⁹ The person/number agreement paradigms for both forms are shown in Table 4 (the form in *+RAkkA* is only attested in the 3rd person in my data).

⁸The form in *+RI* has irregular conjugation and alternates between *-ri*, *-si*, *-ji* and consonant reduplication and/or vowel reduplication and/or alternation. See Ikegami (2001 [1959]) for a full breakdown of alternations by conjugational class of the verb stem. For this and other forms, irregular inflection is marked by a plus sign and capital letters throughout this paper (capitalised vowels indicate vowel harmony).

⁹*+RA* is cognate with the Tungusic aorist form in *-rA*. *-rA* in combination with the emphatic particle in *=k(k)A* is attested as an (emphatic) confirmative mood form in a number of Tungusic languages (Malchukov 2000: 458). In Uilta the bare form in *+RA* marks the lexical verb (glossed as connegative) in negative constructions with the inflected negative auxiliary in *ə* (cf. examples 8, 34, 38, 41 and 43). In combination with other morphemes, it forms the direct evidential/mirative/emphatic in *+RAkkA* (cf. §4.2), the likely/anticipated future in *+RAŋA* (§5.3), and the different-subject imperfective conditional converb in *+RAi* (cf. examples 34, 41). All forms in *+RA* in Uilta have irregular conjugations and alternate between *-rA*, *-si* and vowel reduplication and/or alternation and/or consonant reduplication. See Ikegami (2001 [1959]) for details.

Table 4: Person/number agreement paradigms of Uilta present tense forms

Number	Person	+RI	+RAkkA
Singular	1 st	+RI- <i>wi</i>	–
	2 nd	+RI- <i>si</i>	–
	3 rd	+RI- <i>ni</i>	+RAkkA
Plural	1 st	+RI- <i>pu</i>	–
	2 nd	+RI- <i>su</i>	–
	3 rd	+RI- <i>či</i>	+RAkkA- <i>l</i>

4.1 General present in +RI

The form in +RI is the most frequent present tense form, used in all present contexts except for direct evidential, emphatic and mirative, where the form in +RAkkA is used instead (see below). It is used for events occurring at the moment of speaking as in (19), events occurring in the present generally (generic present) as in (20), habitual events (21), and general statements (22).

- (19) *Bii dægjitə pukki-mbi. Ĵij nalmakta see-ri-či.*
 1SG burn+HORT.1SG fire-1SG.POSS very mosquitoes bite-PRES-3PL
 ‘Let me start a fire. Mosquitoes are biting really bad.’
- (20) *Bəyø-mbø uĵi-pissəə waa-ri-či.*
 bear-ACC rear-COND.CONV.SS.PL kill-PRES-3PL
 ‘They rear the bear and then kill it.’
- (21) *Nooni purə-ttəi puli-si-ni waa-ŋda-su-si-ni.*
 3SG taiga-ALL walk-PRES-3SG kill-VEN-ITER-PRES-3SG
 ‘He regularly goes into the taiga and hunts.’
- (22) *Suwəə-ĵĵee suunə agbi-ŋĵi-ni, pərxi-ĵĵee tuu-ĵĵi-ni.*
 east-side sun appear-PRES-3SG west-side fall-REITER+PRES-3SG
 ‘The sun rises in the east and sets in the west.’

The form in +RI also appears in narratives as in (23), although less frequently than the general past form in -xAn.

- (23) *Wəɖəɖ-ɸtu-xə əəktə pɛɛččila-gačči təə-si-ni moo*
 lose-INTR-PFV woman lean-ANT.CONV sit-PRES-3SG tree
pəjʃje-du-ni.
 under-LOC-3SG.POSS
 ‘The lost woman sat down leaning against the tree.’

In the Southern dialect, which lacks the general future form in *-li* (see §5.2 below), the form in *+RI* is also used for both near (24) and distant future events (25).

- (24) *Ələ-pee sittəi buu-ri-wi, ələ-pula sundattaa.*
 cook-COND.CONV.SS 2SG.ALL give-NPST-1SG cook-PASS.PFV fish+ACC
 ‘I will cook and give [it] to you, the cooked fish.’ (Southern dialect)
- (25) *Bii mərə-čči-wi xaali=ddaa daayi naa-tai ɣənnəe-wi.*
 1SG think-DUR+NPST-1SG when=FOC big land-ALL go+NPST-1SG
 ‘I am thinking, I will go to the mainland someday.’ (Southern dialect)

On top of its predicative use as the main verbal present form, the form in *+RI* retains its participial (adnominal)/nominal function, as in (26); cf. also example (32).

- (26) *Pəččə-nə-si-l=ddəə, mičči-l=ddəə, naa-wa*
 jump-ITER-IPFV-PL=FOC crawl+IPFV-PL=FOC earth-ACC
xullee-l=ddəə.
 burrow+IPFV-PL=FOC
 ‘Those [insects and worms] that jump, those that crawl, and those that burrow in the ground.’

4.2 Direct evidential/emphatic/mirative present in *+RAkkA*

The direct evidential/emphatic/mirative present form in *+RAkkA*, while far more restricted than the general present in *+RI*, is more frequent than the past evidential/affirmative-emphatic form in *-tAA*, and occurs naturally in everyday speech (Yamada 2013: 114). Previously, it was reported to have 1) direct evidential and 2) experiential meaning, and a full person/number paradigm (Ikegami 2001 [1959]). In present-day Uilta, it is restricted to third person use, and to events witnessed by the speaker, at the moment of speech as in (27):

- (27) *Xəwəɾə-kki bəyətəθ daurakka. Pauri-mi*
 lagoon-PROL bear.cub cross+DIREVID.PRES.3 swim-CONV
aaptu-li-ni=yyuu, xai=yyuu?
 reach-FUT-3SG=Q what=Q
 ‘A bear cub is swimming across the lagoon. Is it going to make it or not?’

Very occasionally, it is used in non-visual direct evidential contexts as in (28).

- (28) *Tarree, čoorə ui-sikkə. Nari-sal*
 that+EMPH bell ring-DIREVID.PRES.3 man-PL
sindaakka-lee.
 come+DIREVID.PRES-PL+EMPH
 ‘There, I can hear a bell. People are coming.’

Typically, it combines direct evidential and emphatic meaning as in (29).

- (29) *Əθ, sindaakka tari nari, sokto-xo čipal!*
 INTJ come+DIREVID.PRES.3 this man get.drunk-PFV completely
 ‘There, this man is coming, completely drunk!’

In some instances, the emphatic meaning is clearly more prominent, and the evidential function secondary at best, as in (30) and (31).

- (30) *Nooni mittəi čii puli-sikkə, čii*
 3SG 1SG.ALL constantly walk-DIREVID.PRES.3 constantly
puli-sikkə.
 walk-DIREVID.PRES.3
 ‘He constantly comes to me [won’t leave me alone].’
- (31) *Bii ulbaaxxoo-wwee taitəkkə goči ai suddəəki tari*
 1SG dress+ACC-1SG.POSS wear+DIREVID.PRES.3 EMPH INTJ shameless this
əəktə.
 woman
 ‘She’s wearing my dress, how shameless, this woman.’

Finally, the form in +*RAkkA* is used to express mirative meaning (the speaker’s surprise at unexpected revelation or new information), as in (32).

- (32) *Ərəəi, tari nurreekka goči ləədənji-wə-ppə!*
 INTJ this write+DIREVID.PRES.3 EMPH talk+IPFV-ACC-1PL
 ‘Oh, it is recording what we are saying!’ [The informants realised that the recording device was on.]

4.3 The effect of insubordination on the Uilta present tense forms

Similar to what we observe in the past domain, the functional distribution of the two present forms in present-day Uilta is consistent with the new form of participial origin gradually replacing the old finite form in most contexts, limiting it to direct evidential, emphatic and mirative uses. This mirrors the development in other languages of the Udegheic and Nanaic groups, where the participial forms, semantically neutral, pushed out the old verbal forms into direct evidential, validational and affirmative-emphatic uses, to varying degrees (markedness reversal). This process is typically further advanced in the past domain than in the present (Tense Hierarchy; Malchukov 2000).¹⁰ This is borne out by the fact that the present finite form in +RAkkA is more frequent than the corresponding past form in -tAA in present-day Uilta.

The fact that the form in +RAkkA is restricted to third person use in present-day Uilta is probably motivated by the fact that the third person is more congruous with direct evidential and mirative semantics.

5 Future domain

The Uilta future tense domain displays the clearest example of insubordination at work. There are three future tense forms in the Southern dialect, two of finite and one of participial origin. The present-day Northern dialect additionally features another participial form. It will be shown, through comparison with previous descriptions, that this new form replaced the old finite forms in most functional domains, to become the most productive future form in the Northern dialect.

The four forms are: general future in *-li* (Northern dialect only), immediate spontaneous future in +RILA, likely/anticipated future in +RAηA, and probable future in +Rili. +RILA and +RAηA are pure verbal forms, i.e. can only be used as predicates of a main clause. They take person/number agreement of the verbal type. The forms in *-li* and +Rili are of participial origin, and retain their function as participles/nominalisations. They take agreement of the nominal (possessive) type, also in predicative use. The agreement paradigms for all four forms are presented in Table 5.

¹⁰Malchukov (2000: 450) postulates the Markedness Hierarchy according to which the process of replacement of unmarked finite forms through marked participial forms is further advanced in the past than the present domain, in the plural further than in the singular, and in the 3rd further than in the 1st and 2nd person.

Table 5: Person/number agreement paradigms of Uilta future tense forms

Number	Person	-li (N dialect)	+Rili	+RILA	+RAŋA
Singular	1 st	-li-wi	+Rili-wi	+RILA-mi	+RAŋŋii
	2 nd	-li-si	+Rili-si	+RILA-si	+RAŋA-si
	3 rd	-li-ni	+Rili-ni	+RILAA	+RAŋŋAi
Plural	1 st	-li-pu	+Rili-pu	+RILA-pu	+RAŋA-pu
	2 nd	-li-su	+Rili-su	+RILA-su	+RAŋA-su
	3 rd	-li-či	+Rili-či	+RILAA-l	+RAŋŋA-l

5.1 Immediate spontaneous future in +RILA

The immediate spontaneous future form in +RILA, from the imperfective participle in +RI plus -LA (< *-lan, of unknown origin; Pevnov 2016), is the most productive future form in the Southern dialect, and the second most productive in the Northern dialect, where it competes with the general future form in -li.

In the Northern dialect, +RILA is restricted to immediate future spontaneous contexts, as in (33), (34) and (35).

- (33) *Sii čaa bičixxəə tauuta-ssee, bii sittəi kampeetka*
 2SG that book+ACC read+COND.PFV.CONV.DS-2SG 1SG 2SG.ALL candy
buu-rilə-mi.
 give-NEARFUT-1SG
 ‘If you read this book, I will give you a candy.’
- (34) *Sii noo<mba>ni ə-siyi-si sommee nooni*
 2SG 3SG<ACC> NEG.AUX-COND.IPFV.CONV.DS-2SG close+CONNeg 3SG
pukči-lləə.
 jump-NEARFUT.3
 ‘If you don’t close it [the door], it [the cat] will run around.’
- (35) *Kooppee umi-gačči, gayai-ŋjila-mi.*
 coffee+ACC drink-ANT.CONV stay.awake-NEARFUT-1SG
 ‘Having drunk coffee, [as a result] I won’t be able to sleep.’

In the most detailed previous description (Ikegami 2001 [1959]) the form in +RILA was characterised as expressing 1) near future, 2) future of which the speaker is sure, and 3) spontaneous action in the future. In all attestations of this form

in my data, both conditions 1) and 3), namely short temporal distance and spontaneity, are met. Furthermore, the form is limited to very near, or immediate, future contexts. “Spontaneous” does not imply agent’s own volition, cf. example (35). The relevant distinction is between spontaneous, as in decided/realised on the spot, and planned, or otherwise predicted or predictable events. The form in +*RIIA* is compatible with durative verbs as in (34) and (35), but for actions and states extending into the future, which conflict with its immediate future semantics, the form in *-li* will be used instead (see below). Similarly, for the epistemic modal function reported previously, future that the speaker is sure of, the forms in *-li* or +*Rlli* (see §5.2 and §5.4 below) will normally be used unless the use of the form in +*RIIA* is specifically conditioned by immediate and spontaneous context.

5.2 General future in *-li*

In the Northern dialect of Uilta, the form in *-li* is the most productive, general future form, with the other forms limited to their specific functions. It is used in all contexts that do not warrant the use of any of the other forms, immediate spontaneous future in +*RIIA*, or the two marginal forms with epistemic modal semantics, +*Rlli* and +*RAŋA* (see §5.3 and §5.4 below). For example, it is used for all planned future events, whether near (36) or distant (37).

- (36) *Igəri sindauta-nnee, buu Naxulakka-tai ɲənə-li-pu.*
 Igor(PN) come+COND.PFV.CONV.DS-3SG 1PL Nogliki-ALL go-FUT-1PL
 ‘When Igor comes, we will go to Nogliki.’ [already planned]
- (37) *Ərkəə nooni xotto-du isu-li-ni duku-takki.*
 next.year 3SG city-ABL come.back-FUT-3SG house-ALL.REFL
 ‘Next year he will return home from the city.’

It is also used for predicted or expected future outcomes (38), (39), or statements about the future that hold generally (40).

- (38) *Məənə boččoo-bi əxəktə-mi, tari andu-l-bi*
 own face+ACC-REFL.POSS take.care-CONV this work-PL-REFL.POSS
ə-mi=ddəə xoŋʒee o-li-si taani.
 NEG.AUX-CONV=FOC finish+CONNNEG do-FUT-2SG likely
 ‘If you are preoccupied with your own face [looks], you won’t finish these works.’

- (39) *Tari xoosa-ḡu-l-bari* *to-ḡḡi-ndulli,*
 this reindeer.leg.hide-AL-PL-REFL.POSS.PL do-REITER+IPFV(PRES)-LOC.REFL
asi=ləkə utta-lu o-li-pu.
 now=TOP boot-POSS become-FUT-1PL
 ‘As we process these reindeer leg hides, now we are going to be in possession of boots.’
- (40) *Nəḡnə boo-du nama-li-ni.*
 spring outside-LOC be.warm-FUT-3SG
 ‘In the spring it is going to be warm outside.’

It is also used instead of the form in +*RIL*A for unplanned, spontaneous events if these are not temporally limited to the immediate future, as in (41).

- (41) “*Sii gaandu-ittaayi-si* *məənə puttə-bi,* *bii*
 2SG go.after-VOL+COND.IPFV.CONV.DS-2SG own child-REFL.POSS 1SG
sindu gəsə ə-li-wi bee”, unḡi-ni nooni sitəu
 2SG.LOC together NEG.AUX-FUT-1SG be+CONNEX say+PRES-3SG 3SG new
mama-ḡu-ni.
 wife-AL-3SG.POSS
 “‘If you want to go and bring your child, I won’t live with you”, says his new wife.’

Finally, as with the forms in -*xAn* and +*RI*, the form in -*li* retains its participial (attributive/nominal) function as in (42).

- (42) *Nooči sinda-li-či ulaa-l-ba uidu-xə-či.*
 3PL come-FUT-3PL reindeer-PL-ACC send-PAST-3PL
 ‘They₁ dispatched the reindeer by which they₂ are coming.’

5.3 Likely/anticipated future in +*RAḡA*

The form in +*RAḡA* (cf. footnote 9) combines temporal and epistemic/deontic modal meaning, expressing future that the speaker considers very likely, for example through inference from past experience or common knowledge. In a previous description (Ikegami 2001 [1959]) it was characterised as follows: 1) distant future, 2) possible future, 3) action in the future the doer is compelled or obliged to perform. In present-day Uilta, this form has no inherent temporal distance value, its use being conditioned exclusively by its epistemic/deontic modal

function.¹¹ It is exemplified below in the 1st, 2nd and 3rd person use, expressing likelihood based on inference from circumstances (43), common knowledge (44), and past experience (45). Example (46) shows the use of the form in +*RAŋA* in the deontic modal function (obligatoriness). As is clear from the below examples, it is not limited to distant future contexts.

- (43) *Gəə, balaa, gata-mari ə-ŋə-pu*
 INTJ fast+EMPH pick-CONV.PL NEG.AUX-DISTFUT-1PL
kulpee. Kusal-ji gitu-mari ɲənneesu.
 make.it.in.time+CONNeg fast-INSTR walk-CONV.PL go+HORT.1PL
 ‘Come on, faster, we won’t finish picking in time. Let’s walk faster.’
- (44) *Čii tagda-na-mi, čii tagda-na-mi,*
 constantly be.angry-ITER-CONV constantly be.angry-ITER-CONV
ənu-lləəŋə-si.
 be.ill-INCH+DISTFUT-2SG
 ‘If you are angry all the time, you will fall ill.’
- (45) *Dolbo puttə-ni soŋŋee-du-ni əni-ni*
 night child-3SG.POSS cry+IPFV(PRES)-LOC-3SG mother-3SG.POSS
əmu-mi toŋŋai. Əmu-siŋŋai.
 rock-CONV do+DISTFUT.3 rock-DISTFUT.3
 ‘During the night, when her child cries, the mother will be rocking him.
 She will rock him.’
- (46) *Bii xotto-ttoi ɲənnəəŋŋii, puyə-bi ittəu-ndəəŋŋii.*
 1SG city-ALL go+DISTFUT.1SG wound-1SG.POSS show-VEN+DISTFUT.1SG
 ‘I ought to go to the city and have my wound looked at.’

5.4 Probable future in +*Rlli*

Similar to the form in +*RAŋA*, the form in +*Rlli* (from imperfective participle +*RI* plus future participle *-li*) combines temporal and epistemic modal meaning, expressing future that the speaker considers probable (cf. also Ikegami 2001 [1959]). It is usually accompanied by the adverb *taani* ‘likely, probably’, as in (47) and (48).

¹¹I gloss this form as “distant future” throughout this paper in line with previous descriptions, and to distinguish it from other future forms.

- (47) *Upa-ŋu-bi dabgu-xa-ni, ləpeeskə-buŋji*
 flour-AL-REFL.POSS prepare-PST-3SG make.flatbread-PURP.REFL
to-i-ni. Isu-pee to-ŋjeeli-ni taani.
 do-PRES-3SG come.back-COND.CONV.SS do-REITER+PROBFUT-3SG likely
 ‘She prepared the flour, intends to make flatbread. She will probably
 resume making it when she comes back.’
- (48) *Yə uilə-bi xoŋi-gačči, goi uilə-bi čai-wa*
 this work-REFL.POSS finish-ANT.CONV other work-REFL.POSS tea-ACC
umi-pee otokoo to-ŋjeeli-wi taani.
 drink-COND.CONV.SS later do-REITER+PROBFUT-1SG likely
 ‘Having finished this work, after having tea, I will likely go back to doing
 the other work later.’

Like the forms in *-xAn*, *+RI* and *-li*, the form in *+Rlli* is ultimately of participial origin, and retains its attributive/nominal function, as in (49).

- (49) *Nooni aduli-bi atu-ŋjeeli-wa-ni tari-sal*
 3SG fishing.net-REFL.POSS remove-REITER+PROBFUT-ACC-3SG that-PL
sinda-xa-či, tulə-du-xə-či.
 come-PST-3PL set-REITER-PST-3PL
 ‘They came and set again the fishing nets that he wanted to remove.’

5.5 The effect of insubordination on the Uilta future tense forms

The general future form in *-li*, the most productive future form in the present-day Northern dialect of Uilta, is not attested in the previous descriptions before the 2000s (cf. Table 2). Moreover, it is not attested in the Southern dialect, where the finite form in *+RlIA* is the most productive future tense form, with the form in *+RI* also extended to future use. The most comprehensive description of the Northern dialect, by Petrova (1967), does not mention the form in *-li*, but briefly describes another future form in *-lee*, not mentioned anywhere else. It is unclear whether the forms in *-lee* and *-li* are related, but consonant gemination with vowel lengthening is a prominent feature in Uilta, frequently used for emphasis (cf. example 28), but also for marking grammatical categories like accusative (cf. e.g. examples 6 and 33). With some markers, e.g. the connegative form in *+RA* (from Tungusic aorist in *-rA*, cf. footnote 9), there is free variation between geminated and ungeminated forms in some conjugations.

Nevertheless, the form in *-li* features prominently in the most recent descriptions of the Northern dialect (Pevnov 2016; Yamada 2010b; 2013), as well as the

data from fieldwork in recent years by the present author. It accounts for 60% of all future forms in my data, with the form in *+RIIA* at 40%, and the other two forms being marginal. It apparently developed relatively recently in the Northern dialect, and pushed out the older, finite forms in most functional domains: the old distant future form in *+RAŋA* no longer displays the temporal distance value, and is limited to epistemic modal uses; the form in *+RIIA* is restricted to immediate future, spontaneous events. While tail-end languages are known to undergo substantial grammatical changes (Harrison & Anderson 2008),¹² this rather dramatic shift seems to be another manifestation of the tendency of Tungusic languages (and more broadly, languages of the “Macro-Altaic” areal-typological profile) to renew verbal forms through participles, through the processes of insubordination and verbalisation.

6 Summary and conclusions

As is clear from the above description, the processes of insubordination and verbalisation played a prominent role in the development of the Uilta tense system. The gradual replacement of finite verbal forms through forms of participial origin, with the resulting functional shifts between old and new forms in the relevant verbal categories, is evident across all three temporal domains. In the past domain, the development of the perfective participle in *-xAn* into the general past tense form, through resultative, perfect, and indirect evidential stages, mirrors the development in other Tungusic languages (Malchukov 2000: 447). Uilta represents the last stage of this process as the form in *-xAn* has no discernible evidential meaning; it functions as the general past tense form, with the resultative meaning only partially retained. The erstwhile finite form in *-tAA* is marginally retained, with direct evidential and affirmative-emphatic (particularly in the first and second person, the third person being naturally more congruous with evidential meaning) functions, reflecting its development through the direct evidential and affirmative-emphatic stages, in competition with the finite form. Again, this mirrors the development in other languages of the Nanaic and Udegheic groups: as the participial forms replace the erstwhile finite forms, first in resultative/perfect, then indirect evidential use, the old past forms are restricted to the direct evidential function, and further develop affirmative-emphatic (validational) meaning (stages 2 and 3 in Figure 1 above).

¹² “[L]ast generation speakers of endangered languages [...] can and do introduce grammatical and phonological innovations, [...] including changes resulting in both simplification and in greater complexity. It is often difficult to disentangle whether a particular change is driven by internal restructuring, contact induced change, obsolescence effects, or some combination of these.” (Harrison & Anderson 2008: 243 ff.).

Similarly, in the present domain, the participial form in *+RI* replaced the old verbal form in *+RAkkA* as the general present form, with the old form restricted to third person direct evidential, emphatic and, by extension, mirative uses. The fact that the form in *+RAkkA*, although marginal and restricted to third person use, is still more frequent than the equivalent past form in *-tAA* conforms to the Tense Hierarchy of the patterns of replacement of old verbal forms postulated in Malchukov 2000: 450).

Finally, in the future domain, the participial form in *-li* pushed out the old finite forms in *+RIIA* and *+RAŋA* to become the most productive, general future tense form. This recent development, less advanced than in the past and present domains and limited to the Northern dialect, is yet another example of the tendency of Tungusic languages to renew finite verbal forms through insubordination. It represents the most recent one in the history of repeated cycles of renewal of verbal forms through participles in Tungusic, with most finite forms, including the above forms in *+RA*, ultimately of participial origin (Robbeets 2009).

In fact, this tendency is not limited to Tungusic, with all languages of the “Macro-Altaic” areal-typological type repeatedly undergoing similar development, with some apparent parallels at the proto-languages stage as postulated by Robbeets (2009; 2015), some evident in the diachronic development of individual families, and some still observed in the individual languages (Malchukov & Czerwinski 2020). Note, however, that this tendency is not limited to “Macro-Altaic”, and instead constitutes a general areal feature of Siberian languages, including the Paleosiberian and Uralic languages (Malchukov 2013; Malchukov & Czerwinski 2021). In Uilta, this process played a prominent role in the development, and is largely responsible for the current shape of the Uilta tense system.

Abbreviations

1	1 st person	CONCUR	concurrent
2	2 nd person	COND	conditional
3	3 rd person	CONNEG	connegative
ABL	ablative (case)	CONV	converb
ACC	accusative (case)	COP	copula
AGR	(person/number) agreement	DIREVID	direct evidential
AL	alienable (possession)	DISTFUT	distant future
ALL	allative (case)	DS	different subject
ANT	anterior	DUR	durative
AUX	auxiliary	EMPH	emphatic

3 Tense and insubordination in Uilta (Orok)

FOC	focus	PN	proper name
FUT	future	POSS	possessive
HORT	hortative	PRES	present
INCH	inchoative	PROBFUT	probable future
INTR	intransitiver	PST	past
INSTR	instrumental (case)	PURP	purposive
INTJ	interjection	Q	question particle
IPFV	imperfective	REFL	reflexive
ITER	iterative	REITER	reiterative
LOC	locative (case)	S	subject
N	nominalisation	SG	singular
NEARFUT	near future	SS	same subject
NEG	negative	TOP	topic
NFUT	non-future	TRANS	transitive
NPST	non-past	V	verb
PART	participle	VEN	venitive
PASS	passive	VOL	volitive
PFV	perfective		
PL	plural		

Acknowledgements

I am very grateful to Andrej Malchukov, Walter Bisang, Andreas Hölzl and two anonymous reviewers for their valuable comments. I would like to express my deepest gratitude to my Uilta informants, Elena A. Bibikova, Ljubov' R. Kitazima, Ljubov' N. Konusova, Irina G. Kurušina and Ljudmila X. Minato. This work was partially supported by the Laboratory Program for Korean Studies through the Ministry of Education of the Republic of Korea and the Korean Studies Promotion Service of the Academy of Korean Studies (AKS-2016-LAB-2250004).

References

- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar*. Chicago: University of Chicago Press.
- Croft, William. 2002 [1990]. *Typology and Universals*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

- Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 366–431. Oxford: Oxford University Press.
- Harrison, David K. & Gregory D. S. Anderson. 2008. Tofa language change and terminal generation speakers. In David K. Harrison, David S. Rood & Arienne Dwyer (eds.), *Lessons from documented endangered languages*, 243–270. Amsterdam: John Benjamins.
- Ikegami, Jirō 池上 二良. 2001 [1959]. The verb inflection of Orok. In Ikegami Jirō 池上 二良 (ed.), *Tsungūsugo kenkyū ツングース語研究 [Tungusic languages research]*, 24–66. Tokyo: Kyūko Shoin.
- Ikegami, Jirō 池上 二良. 2001 [1994]. Uirutago-no minami hōgen-to kita hōgen-no sōiten ウイルタ語の南方言と北方言の相違点 [Differences between the Southern and Northern dialects of Uilta]. In Jirō 池上 二良 Ikegami (ed.), *Tsungūsugo kenkyū ツングース語研究 [Tungusic languages research]*, 247–283. Tokyo: Kyūko Shoin.
- Magata, Hisaharu 潤濁 久治. 1981. *Uirutago jiten* ウイルタ語辞典 [*Uilta dictionary*]. Abashiri: Hoppō Minzoku Bunka Hozon Kyōkai.
- Majewicz, Alfred F. (ed.). 2011. *The collected works of Bronisław Piłsudski, vol 4: Materials for the study of Tungusic languages and folklore*. Berlin: De Gruyter Mouton.
- Malchukov, Andrej L. 2000. Perfect, evidentiality and related categories in Tungusic languages. In Lars Johanson & Bo Utas (eds.), *Evidentials: Turkic, Iranian and neighboring languages*, 441–470. Berlin: De Gruyter Mouton.
- Malchukov, Andrej L. 2013. Verbalization and insubordination in Siberian languages. In Martine Robbeets & Hubert Cuyckens (eds.), *Shared grammaticalization: With special focus on the Transeurasian languages*, 177–208. Amsterdam: John Benjamins.
- Malchukov, Andrej L. & Patryk Czerwinski. 2020. Verbal categories in the Transeurasian languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 604–624. Oxford: Oxford University Press.
- Malchukov, Andrej L. & Patryk Czerwinski. 2021. Verbalization/insubordination: A diachronic syntactic isogloss in Northeast Asia. *International Journal of Eurasian Linguistics* 3(1). 83–104.
- Nakanome, Akira 中目 覚. 1917. *Orokko bunten* オロツコ文典 [*A grammar of Orok*]. Tokyo: Sanseidō.
- Ozolinja, Larisa V. 2013. *Grammatika orokskogo jazyka [A grammar of Orok (Uilta)]*. Novosibirsk: Izdatel'stvo GEO.

- Petrova, Taisija I. 1967. *Jazyk orokov (ul'ta)* [The language of Oroks (Uilta)]. Leningrad: Nauka.
- Pevnov, Alexandr. 2016. On the specific features of Orok as compared with the other Tungusic languages. *Studia Orientalia Electronica* 117. 47–63.
- Robbeets, Martine. 2009. Insubordination in Altaic. *Voprosy filologii [Serija: Uralo-altajskie issledovanija]* 1. 61–80.
- Robbeets, Martine. 2015. *Diachrony of verb morphology: Japanese and the Transeurasian languages*. Berlin: De Gruyter Mouton.
- Tsumagari, Toshiro 津曲 敏郎. 2009. Grammatical outline of Uilta (Revised). *Hokkaidō University Journal of the Graduate School of Letters* 4. 1–21.
- Whaley, Lindsay & Sofia Oskolskaya. 2020. The classification of the Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 81–91. Oxford: Oxford University Press.
- Yamada, Yoshiko 山田 淑子. 2010a. A preliminary study of language contacts around Uilta in Sakhalin. *Journal of the Center for Northern Humanities* 3. 59–75.
- Yamada, Yoshiko 山田 淑子. 2010b. Uirutago kita hōgen-ni mirareru dōshi gobi -li ni tsuite ウイルタ語北方言にみられる動詞語尾-liについて [On the verb ending -li in the Northern dialect of Uilta]. In Megumi Kurebito 呉人 恵 (ed.), *Languages of the North Pacific Rim* 15, 85–100. Faculty of Humanities, University of Toyama.
- Yamada, Yoshiko 山田 淑子. 2013. *Uirutago kita hōgen-no bumpō-to gengo sesshoku-ni kan suru kenkyū* ウイルタ語北方言の文法と言語接触に関する研究 [A study of the grammar and contact situation of the Northern dialect of Uilta]. Sapporo: Hokkaidō University Collection of Scholarly & Academic Papers (HUSCAP). (Doctoral dissertation).

Chapter 4

‘What’s your name?’ in Tungusic and beyond

Andreas Hölzl

University of Potsdam

This study investigates questions about personal names, i.e. questions corresponding to *What’s your name?* in English. This potentially universal type of question is referred to as the personal name question (PNQ). The study sketches the typological variation found in the PNQ from a cross-linguistic perspective and analyzes the synchronic typology and diachronic development of the PNQ in Tungusic, a small but important language family spoken in Northeast Asia.

Cross-linguistically, two main types of PNQs are attested. Type A is an equational copula sentence (e.g., *What is your name?*) while Type B contains a speech act verb (e.g., *What are you called?*). Tungusic shows a tendency for Type A but, because of contact languages such as Mongolian and Russian, also has instances of Type B. One of several other dimensions of variation among the world’s languages is the kind of interrogative used in PNQs. Tungusic languages originally used an interrogative meaning ‘who’ (literally *Who is your name?*). The use of ‘what’ in several languages located in the south and of ‘how’ in many languages in the north can be attributed to influence from Chinese, Russian, and other languages.

Historical accounts of Tungusic are usually restricted to individual items (e.g., **si* ‘you (SG)’ **gärbü* ‘name’, **njii* ‘who’, e.g. Benzing 1956), but rarely are larger expressions reconstructed to Proto-Tungusic. This study shows that the Proto-Tungusic PNQ as one idiom can be plausibly reconstructed as **si(n-i) gärbü-si njii?* ‘2SG.(OBL-GEN) name-2SG.POSS who’. Most deviations in modern languages can be explained by contact with surrounding languages.

Keywords: personal name question, typology, Tungusic, reconstruction, frames, construction grammar



1 Introduction

FAUST: What is thy name?

MEPHISTOPHELES: A question small, it seems,
For one whose mind the Word so so much despises;
Who, scorning all external gleams,
The depths of being only prizes.
(Johann Wolfgang von Goethe 2018 [1808])

This study investigates what will be referred to as the *personal name question* (PNQ), i.e. a question about the name of a person, more specifically of an addressee (or second person), such as *What's your name?* in English. Almost every natural language seems to have a conventional way of expressing this question. But despite being a question that occurs in textbooks of many languages, there has been surprisingly little cross-linguistic research on this topic. Even *The Oxford handbook of names and naming* (Hough 2016) only devotes a brief section to this topic (Van Langendonck & Van de Velde 2016: 26). Not many grammatical descriptions mention PNQs and even fewer address it as a topic in its own right. There are some noticeable exceptions, such as Mushin (1995: 8, 19), who noted that Australian languages often employ a personal interrogative meaning 'who' in questions about names. Blust (2013: 509f.) made a similar observation about Austronesian languages. The following examples, therefore, literally mean 'Who is your name?' (see also Hölzl 2014; Gil 2018).¹

- (1) Yankunytjatjara (Pama-Nyungan; Mushin 1995: 19)
nyuntu ini ngana-nya?
2SG.NOM name who-NOM.NAME
- (2) Ngaju Dayak (Austronesian; Blust 2013: 510)
eweh ara-m?
who name-2SG.POSS

Many other languages, such as Aymara spoken in southern Peru or Badaga in India, behave like English and use an interrogative with the meaning 'what' instead.

- (3) Muylaq' Aymara (Aymaran; Coler 2014: 402)
¿kuna suti-ni-v-rak(i)-ta-st(i)?
what name-ATT-COP.V-AD-2SIM-Q

¹Throughout the paper, examples without translation can be translated into English as 'What is your name?' or as an answer thereto.

- (4) Badaga (Dravidian; Balakrishnan 1999: 214)

ninna hesaru e:na?

2SG.GEN name what

Some languages, such as Tok Pisin spoken in Papua New Guinea or Wulai Atayal on Taiwan, allow the use of both ‘who’ and ‘what’.

- (5) Tok Pisin (English-based creole; Wurm & Mühlhäusler 1985: 345)

husat/wanem nem bilong yu?

who/what name GEN 2SG

- (6) Wulai Atayal (Austronesian; Huang 1996: 293: 293)

ima?/nalu? lalu?=su??

who/what name=2SG.POSS

This variation is also addressed in Idiatov (2007: 61–94, *passim*), who, among other things, investigated “NAME-questions” in a large sample of languages. This kind of question is broadly defined, however, and not restricted to the question about personal names. According to Idiatov (2007: 47), the question is based on “non-prototypical combinations of values” because it combines the features THING, IDENTIFICATION, and PROPER NAME (as an expected answer). Prototypical combinations, on the contrary, are said to be PERSON, IDENTIFICATION, PROPER NAME for ‘who’ (e.g., *Who are you? I’m Mike.*) and THING, CLASSIFICATION, and COMMON NOUN for ‘what’ (e.g., *What is this? This is a book.*). Following Idiatov (2007), the fact that some languages like Aymara use ‘what’ and others, such as Ngaju Dayak, ‘who’ in questions about names is a result of the non-prototypical combination of these features that allows both choices. An alternative explanation of the variation, among other things based on the ambiguous nature of the concept NAME itself, will be proposed in this study. The use of other interrogatives, such as *jak* ‘how’ in Polish (asking about the MANNER), is argued to be an “avoidance strategy” (Idiatov 2007: 61). This is a feature common in, but not restricted to, European languages.

- (7) Polish (Indo-European)

Jak masz na imię?

how have.2SG.PRS.IND on name

- (8) Manam (Austronesian; Lichtenberk 1983: 406)

ará-m bá?ara?

name-2SG.POSS how

For some reason, the focus of previous studies has been on the choice of the interrogative in the PNQ. Apart from Idiatov (2007: 63–67), few studies address morphosyntactic patterns on how questions about names are expressed cross-linguistically. But the PNQ also varies on many other dimensions, including the marking of possession, politeness, the presence or absence of a copula, the valency of the speech act verb and many more. These typological features of the PNQ are addressed in §2.

The underlying theoretical background of this study is loosely based on a general form of Frame Semantics and Construction Grammar, especially as it can be applied to historical and areal phenomena (e.g., Fillmore 1985; Langacker 2008; Hilpert & Östman 2014; Trousdale 2014; Lefebvre 2015; Hölzl 2018b). Construction Grammar is built on the idea that the lexicon and the grammar of a language are not clearly distinct, but form a continuum of constructions of different size and complexity. Crucially, idioms and fixed expressions, including the PNQ, are considered constructions in their own right. Construction Grammar allows for partial analyzability and different levels of schematicity. In English, for instance, *What's your name?* is not only a conventional expression, but is at the same time analyzable as an instantiation of more abstract constructions, including *what's X*, where X refers to an open slot. The questions *What's this?* and *What's the problem?* are other instantiations of this partially schematic construction.

This study investigates the personal name question in the Tungusic language family, which allows a detailed analysis of the individual constructions involved in the expression of the question. Tungusic is a small language family of up to twenty different languages spoken in Northeast Asia, especially eastern Russia and northern China. Data from all attested Tungusic languages are included in the study. Its internal classification is a matter of dispute, but four different subgroups can unmistakably be identified. Following Janhunen (2012b), these will be referred to as Ewenic, Udegheic, Nanaic, and Jurchenic. According to one view (e.g., Georg 2004; Janhunen 2012b), the former two together form the Northern Tungusic languages while the latter can be referred to as Southern Tungusic (Table 1). The discussion of the Tungusic PNQ in §4 is divided into subsections on each of the four subgroups. Tungusic is an especially rewarding language family for this study due to the relatively high variability of the personal name question, especially in terms of the interrogative used.

Previous diachronic accounts of Tungusic languages usually focused on phonological, morphological, and lexical aspects (e.g., Benzing 1956; Doerfer 1978 among many others), but have rarely addressed larger expressions. However, similar to lexical items, it is possible to identify COGNATE CONSTRUCTIONS in

4 ‘What’s your name?’ in Tungusic and beyond

Table 1: Possible classification of the Tungusic languages (e.g., Georg 2004; Janhunen 2012b); *languages with highly mixed affiliation

Branch	Subbranch	Languages
Northern Tungusic	Ewenic	Arman, Even (Ewen), Evenki (Ewenki), Oroqen, Solon, Negidal, ...
	Udegheic/ Orochic	Oroch, Udihe (Udeghe), ...
Southern Tungusic	Nanaic	Kilen*, Kili*, Nanai, Samar, Ulcha, Uilta, Ussuri Nanai, ...
	Jurchenic/ Manchuric	Alchuka, Bala, Jurchen A, Jurchen B, Kyakala, written Manchu, spoken Manchu (e.g., Sibe), ...

related languages and, therefore, to reconstruct larger constructions to proto-languages (e.g., Barðdal 2013). A superficial survey of the personal name question in several Romance languages can illustrate this concept.

- (9) a. French Comment t’appelles-tu?
 b. Italian Come ti chiami?
 c. Portuguese Como te chamas?
 d. Romanian Cum te cheamă?
 e. Spanish ¿Cómo te llamas?

Of the five languages mentioned, all can make use of a similar construction with the same elements, e.g. the interrogative *come* ‘how’ in initial position, followed by the personal pronoun *ti* ‘2SG.OBL’, and an inflected second person singular present indicative form of the verb *chiamare* ‘to call’ in Italian (see also 27). Only French has a different verb (*appeler*). Apart from phonological differences, there are also differences in the verbal morphology (e.g., an enclitic personal pronoun *tu* in French, see also 19). Nevertheless, the overall similarity suggests that earlier stages of Romance also had a construction out of which the constructions in the individual languages might have developed.² Changes in the Tungusic PNQ construction and how it can be reconstructed to the proto-language will be addressed in §4 and §5.

²A proofreader pointed out that Brazilian Portuguese also has an innovative construction: *Como você se chama?*

This paper has five sections, including this introduction. §2 sketches a typology of the personal name question from a cross-linguistic perspective. §3 introduces the semantic background of the question from a frame semantic point of view. §4 addresses the expression of the question in Tungusic languages. §4.1 discusses the second person forms and the genitive, §4.2 gives an overview of the word for ‘name’, and §4.3 to §4.6 investigate the PNQ in the four subbranches of Tungusic. The discussion in §5 reconstructs the PNQ to Proto-Tungusic (§5.1) gives some conclusions (§5.2).

2 The personal name question from a cross-linguistic perspective

Personal names are probably a universal or near-universal property of human cultures. An exception could be the Matsigenka in Peru, where “personal names are of little significance” (Johnson 2003: 10). A similar case can be observed in Venezuela, which also illustrates culture-specific functions of personal names:

The Panare, for example, have five personal names for men and seven for women. They are all based on physical characteristics, like ‘big eyes’, ‘cutie’, ‘big one’, ‘lopsided one’ etc. Individuals are more likely to be referred to by kinship and locality, e.g., grandfather of Camana (a place), child of sister, brother (anyone in one’s male peer group), etc. Also, people have different ‘names’ throughout their lifetime. Before about age three, children are just known as ‘baby’. When it looks like they are going to survive, they are given a childhood name. Then when they come of age (ready to marry) they get their adult name. They may also have a Spanish-based name if they are baptised. But none of these ‘names’ are really used all that much as names in the way Europeans use names. Maybe the Christian names come closest. [...] If you ask a Panare person ‘What is your name?’ (in Spanish) you would only get their Christian name in response. (Thomas E. Payne, p.c. 2020)

To my knowledge all Tungusic cultures have personal names. As a rule, Russian and Chinese naming practices can also be found among speakers of Tungusic languages today. Culture-specific details, such as the use of derogatory names among the Manchus (Alonso de la Fuente 2012/2014) or the reference to rivers for the self-identification among the Evenki (Lavrillier 2006), seem to play no significant role for the expression of the PNQ among Tungusic languages. A discussion of specific meanings or functions of names goes beyond the scope of the present paper.

The PNQ could also be a universal or near-universal property, but is expressed differently from language to languages. Cross-linguistically, however, only a limited number of different constructional types is attested (e.g., Idiatov 2007: 63–67). This section gives a brief overview of the typological variation attested in the expression of the PNQ emphasizing those aspects that are relevant for the classification of Tungusic (see also Idiatov 2007 and Gil 2018).

The question ‘What is your name?’ is part of a question-answer sequence, such as in the following well-known Russian dialogue of the explorer Vladimir Arsen’ev with his later friend Dersu Uzala, a member of the Tungusic-speaking Nanai people.

- (10) *Tebja kak zovut? Sprosil ja neznakomca.*
Dersu Uzala, otvečal on.
 “What is your name?” I asked the stranger.
 “Dersu Uzala,” he answered. (Arsen’ev 1921, 2016 [1921]: 18)

More specifically, the sequence consists of a content question with an interrogative, in this case Russian *kak* ‘how’ (see also 26), that is taken up again in the elliptic answer in the form of a personal name, i.e. *Dersu Uzala*.

Pragmatically speaking, there are, of course, many different ways of achieving the same overall meaning as a PNQ, for instance by using an imperative form of a speech act verb (e.g., Schulze 2007: 254). The following is an example from the Tungusic language Evenki (similar to *State your name!*).

- (11) Evenki (Nedjalkov 1997: 148)
si.n-ngi-ve gerbi-ve mi.ne-ve gu:-kel!
 2SG.OBL-GEN-ACC name-ACC 1SG.OBL-ACC say-2SG.IMP
 ‘Tell me your name!’

In certain contexts, even the word *Name!* alone could already be sufficient.

But not only is this much less polite than a question, but cross-linguistically it also is not the usual way of putting the question. CONVENTIONALITY is key in the investigation of the personal name question. While every language is certainly capable of asking for the name of a person, the universal tentatively proposed here is that almost every language might have a conventional way of expressing it.

In some languages, such as German, there are several different ways of putting the question. As in Evenki, an imperative of a speech act verb can be used in certain contexts, for instance when giving vent to one’s impatience.

(12) German

Sag mir (schon) dein-en Name-n!
say.IMP.SG 1SG.DAT already 2SG.GEN-M.SG.ACC name-M.SG.ACC
'Tell me your name (already)!'

Given a certain context, it is also possible to jokingly ask whether somebody actually has a name. Because we know that (in our culture) everybody has a name, we draw the conclusion, by means of pragmatic inference and the intention of being informative, that the appropriate answer to the question is the specific name rather than the answer yes.

(13) German

Have du ein-en Name-n?
have.2SG.PRS.IND 2SG a-M.SG.ACC name-M.SG.ACC
'Do you have a name?'

However, German has two more conventional ways of expressing the question (14) that in most situations would be preferred to the stylistically marked ones above.

(14) German

- a. *Was/Wie ist dein Name?*
what/how is 2SG.GEN.M.SG.NOM name.M.SG.NOM
- b. *Wie heiß-t du?*
how be.called-2SG.PRS.IND 2SG

Conventionality could theoretically be measured by text frequency, but, given that there are no large corpora for Tungusic languages, this method is inapplicable. Most texts that are available to me only contain the question too few times (if at all) to allow any conclusions. The pragmatic approach followed in this study is mostly impressionistic. It is based on the information available in grammar books, dictionaries, some texts, and the information from experts on individual languages.

Cross-linguistically, there are two main ways of expressing this special type of content question that correspond to the two most conventional expressions in German above (14). Consider the following examples from Mandarin and their English translation:

(15) Mandarin (Sino-Tibetan)

- a. [nǐ de míngzi] shì shénme?
 2SG ATTR name COP what
 'What is your name?'
 b. nǐ jiào [shénme (míngzi)]?
 2SG call what (name)
 'What are you called?'

Both examples are directed at a second person and contain an interrogative. Example (15a) is a copula construction that equates 'your name' (the copula subject) with the interrogative (the copula complement, Dixon 2010) while example (15b) contains a speech act verb. These two types of constructions will be referred to as TYPE A and TYPE B, respectively.

Both patterns have several subtypes. Type A, for instance, can take at least two different forms in which the interrogative is either used as an argument of its own (your name = what, see 16) or as an attribute of the noun meaning 'name' (you = what name, see 17). These will be referred to as TYPE A.1 and TYPE A.2.

(16) Jammu/Kashmir Burushaski (Burushaski; Munshi 2006: 148)

[um-e gu-yik] besen ðila?
 2SG-GEN 2SG.POSS-name what be.PRS.3SG[ABST]

(17) Sanjiazi Manchu (Tungusic; Dai 2012: 239)

'ci ['ai 'kəvə]?
 2SG what name

Both types of the personal name question refer to a SECOND PERSON. In many languages, this is overtly marked by a personal pronoun (both types), a possessive marker that also encodes person (especially Type A, see 18), or verbal agreement (especially Type B, see 19).

(18) Nungon (Trans-New Guinea; Sarvasy 2017: 469: 469)

gok maa-ya numa?
 2SG name-2SG.POSS who

(19) French (Indo-European)

Comment t'=appelles=tu?
 how 2SG.OBL=call.2SG.PRS.IND=2SG

In languages with egophoricity, second person can also be encoded indirectly with the help of the *anticipation rule* (Tournadre & LaPolla 2014: 245). In such languages, an egophoric marker usually refers to a first person, but in questions can also refer to a second person because the perspective of the addressee is taken.

- (20) Karlong Mongghul (Khitano-Mongolic; Faehndrich 2007: 114)

ʈɛi-ni nara ja:n-i?

2SG-GEN name what-EGO

- (21) Wutun (Sino-Tibetan; Sandman 2016: 295)

ni-de minze-li ma sho-yek?

2SG-ATTR name-LOC what say-EGO

Among Tungusic languages, only Sibe has been claimed to possess some sort of grammaticalized egophoric system (Li 1984), but to my knowledge, this does not include any marking that would be relevant for the PNQ.

Both types of PNQs usually contain an **interrogative**. A potential exception to this generalization is the language Wari' spoken in Brazil that uses demonstratives instead. Jahai appears to make use of a polar question that also lacks an interrogative (see also Gil 2018).

- (22) Wari' (Chapacuran; Everett & Kern 2007: 40)

cain' cain' ne wixi-um ne?

that.N.DIST that.N.DIST 3N name-2S REC.PST

- (23) Jahai (Austroasiatic; Schebesta & Blagden 1928: 808, 821)

ha kenmo' pai?

Q name you

Literally: 'Your name?'

For Tungusic, only examples with interrogatives are attested. As seen in the Introduction, the kind of interrogative in the name question also differs from language to language. Cross-linguistically, the two most common categories of interrogatives to be found in this question are **THING** (*what*, e.g. English) and **PERSON** (*who*, e.g. Tigre, Pazih), both of which are attested among Tungusic languages.

- (24) Tigre (Afroasiatic; Elias 2014: 227)

man tu səmetka?

who COP.3SG.M name.2SG.POSS.M

Literally: 'Who is your name?'

(25) Pazih (Austronesian; Li & Tsuchida 2001: 44, 46)

ima langat pai siw?

who name Q 2SG.NOM

This variation certainly has several causes, only some of which can be addressed here. In most Tungusic languages, the use of a given interrogative can be explained with language contact. But this does not explain why different interrogatives can be used in the first place.

Table 2 sketches what can be assumed to be some prototypical features of the two interrogatives from a cross-linguistic perspective, although there are language-specific boundaries (based on Nau 1999: 148; Croft 2003: 130; Idiatov 2007: 18).

Table 2: Tentative prototypical combinations of features for ‘who’ and ‘what’. What is referred to as “word class” is not identical to Idiatov’s (2007) feature “expected answer” that is assumed to be “proper name” for ‘who’. Instead, this refers to the word class of the interrogative itself.

	PERSON (<i>who</i>)		THING (<i>what</i>)
referentiality	identification		classification
animacy	human	animate	inanimate
word class	pronoun	proper name	common noun

The frequent use of ‘who’ in PNQs might be explained by the fact that it is a question about an identification of a specific person (*Who are you? I’m Bill.*), but not a classification (*What is that? That is an airplane.*). The two other features are located on well-known typological scales, i.e. pronoun > proper name > common noun and human > animate > inanimate. Perhaps because a PNQ asks about a proper name that is located in the middle of the first of these two scales, ‘who’ (often an interrogative pronoun) and ‘what’ (often an interrogative noun) can both be used. Another factor for the variation might be the ambiguous nature of the concept NAME itself. First, some languages, such as Great Andamanese, treat a name as if it was a body part (Abbi 2013: 80). Second, a name can also be metaphorically conceptualized as a THING that can be possessed (e.g., *I have a book/name, my book/name*). Third, a name can also metonymically stand for the PERSON itself (e.g., *I am Mike*). The first interpretation might allow both ‘who’ and ‘what’ (animate entity), the second favors the use of ‘what’ (inanimate entity), the last of ‘who’ (human being). This represents a slight difference with respect to Idiatov’s (2007: 47) account that assumes that a name generally is a type of THING.

The use of a MANNER (*how*) or other interrogative, such as *come* in Italian or *comment* in French, is less frequent and can possibly be explained with avoidance (Idiatov 2007: 61). This seems to be relatively frequent in southern, central and eastern Europe, but can also be found in other languages (e.g., Gil 2018).

- (26) Russian (Indo-European)
Kak tebja zovut?
how 2SG.ACC call.3PL.PRS.IND
Literally: ‘How do they call you?’

- (27) Italian (Indo-European)
Come ti chiami?
how 2SG.OBL call-2SG.PRS.IND
Literally: ‘How do you call (yourself)?’

As will be shown in §4, many Tungusic languages appear to have calqued the use of a MANNER interrogative on the basis of Russian, i.e. the European pattern spread towards the East.

An interrogative in both types of PNQs may be FOCUSED. Cross-linguistically, there are different means of focusing an interrogative. A strategy common, for instance, in Japonic languages is the use of a morphosyntactic marker.

- (28) Tarama Miyako (Japonic; Aoi 2015: 417)
naa=ju=ba nuu=ti=ga iiga?
name=ACC=TOP what=QUOT=FOC say=q

Except for, perhaps, Uilta, this is not attested in the Tungusic PNQs. Another way of focusing the interrogative is through fronting, also called (full) *wh-movement*, as in English. In Northeast Asia, few languages exhibit this syntactic phenomenon. An indication of fronting is the comparison of the PNQ with its answer. If the personal name appears in the same position as the interrogative (i.e., *in situ*), there is no fronting involved.

- (29) English
a. *What* is [your name]?
b. [My name] is *Anna*.
- (30) Mandarin
a. [nǐ de míngzi] shì shénme?
2SG ATTR name COP what

- b. [wǒ de míngzi] shì ānnà.
 1SG ATTR name COP PN

Northern Tungusic languages are among the very few exceptions with occasional sentence-initial interrogatives in Northeast Asia (Dryer 2013; Hölzl 2018a). Ewenic languages also exhibit other focus positions that are more central for the PNQ. Some Tungusic languages have adopted the European pattern through Russian.

Type A, and sometimes Type B also, contains a DUMMY noun meaning ‘name’. Obviously, there is no generalization on what phonological form this noun has cross-linguistically. It is necessary to distinguish between chance resemblance, a common inheritance, and mutual contact. German *Name* and English *name*, for instance, are similar due to a common Germanic origin. The similarity to Uralic, e.g. Finnish *nimi*, can perhaps best be explained by Indo-European influence (e.g., Anthony 2007: 95). In many other cases, similarities between individual words, such as Persian *nām*, Kurux *na:me*, Japanese *namae*, or Papuan Malay *nama*, is probably the result of chance.

- (31) Papuan Malay (Austronesian; Kluge 2017: 623)
kam pu nama siapa-siapa?
 2PL POSS name who~PL
 ‘What are your names?’

In a few languages, the dummy noun can fuse with other elements. For instance, in the Austronesian language Kilivila, the dummy noun *yaga* ‘name’ (Senft 1986: 420) fused with an interrogative to form the complex stem *amyaga-* ‘what is the name of’ (Senft 1986: 187), which is the basis of the PNQ *amyagam?* that contains a possessive marker *-m* ‘2SG.POSS’ (Senft 1986: 52).

Interrogatives are often reinforced with other elements, such as basic nouns, e.g. Italian *che cosa* ‘what thing > what’ (e.g., Diessel 2003; Hölzl 2018a). Tok Pisin *wanem* ‘what’ seen in (5) is a contraction of English *what* and *name* (Wurm & Mühlhäusler 1985: 210). This reinforcement suggests that the concept NAME is considered, at least by the speakers of this language, a very basic category equivalent to THING.

Depending on the grammar of the individual languages, the dummy noun can belong to a certain class (e.g., animacy, gender, noun class). For instance, it has male gender in German and in the following construction in the Sepik language Abau. In the South American language Panare, it is marked for inanimateness and invisibility.

- (32) Abau (Sepik; Lock 2011: 227)
hwon-o uru po-ho?
2SG-GEN name Q-GL.M
- (33) Panare (Cariban; Payne & Payne 2013: 66)
sinka mën a-yiche?
how IN.INVIS 2-name

In Tungusic, there is no such classification of the dummy noun.

Some languages have more than one dummy that can enter the question. In Standard Korean, for instance, there is a distinction between neutral *ilum* and honorific *sengham* (Song 2005: 95).

- (34) Korean (Koreanic; Sohn 1999: 418)
- a. *ilum i mwe yey-yo?*
name NOM what be-POL
- b. *sengham i ettehkey toy-sey-yo?*
name.HON NOM how become-HON-POL
Literally: ‘How does your name become?’

In this language, the two nouns are part of different constructions. Example (34a) is said to a child or teenager and (34b) is the honorific version. Individual Tungusic languages only have one dummy noun.

An additional distinction in Type A is whether languages make use of an overt COPULA or not. While some languages, such as Sumerian (35), require an overt copula, others, such as Kurux (36) and many Tungusic languages, do not.

- (35) Sumerian (Black 2007: 21)
aba-m mu-zu?
what-3SG.COP name-2SG.POSS
- (36) Kurux (Dravidian; Kobayashi & Tirkey 2017: 242)
ning-hay ender na:me?
2SG-GEN what name

In Type A languages, there is an additional POSSESSIVE RELATIONSHIP, which, depending on the language, can be dependent-marked (e.g., Mongsen Ao, 37), head-marked (e.g., Teiwa, 38), double marked (e.g., Turkish, 39), or unmarked (e.g., Nihali, 40).³

³The PNQ in Mongsen Ao can also be expressed with ‘what’.

- (37) Mongsen Ao (Sino-Tibetan; Coupe 2007: 99)

nə tə-niŋ sópáʔ?

2SG.POSS RL-name who

- (38) Teiwa (Trans-New Guinea; Klamer 2010: 239)

ha-yit amidan?

2SG.POSS-name what

- (39) Turkish (Turkic)

sen-in ad-in ne?

2SG-GEN name-2SG.POSS what

- (40) Nihali (Nagaraja 2014: 116)

ne jumu na:n?

2SG name what

All four types are attested in Tungusic.

In those languages that have POSSESSIVE CLASSIFICATION, there is an additional distinction that refers to the class of the word for 'name'. In Mongsen Ao, for example, the "relational prefix" *tə-* that is seen in (37) is usually found on body parts and kinship terms (Coupe 2007: 84). In Mandarin, *míngzi* 'name' belongs to the set of nouns that is obligatorily possessed with a genitive marker *de*. This marker can be absent with kinship terms. A language that makes a distinction into several different possessive classes is Great Andamanese.

- (41) Great Andamanese (Abbi 2013: 181, 270)

ŋ=er=liu a=fyu bi?

2SG=CL2=name CL1=who COP

In this language, the word *liu* 'name' takes the class 2 possessive marker *er=* ~ *er=* (Abbi 2013: 80, 140, 161) that otherwise attaches to "major body parts that pertain to the 'head', 'brain', 'neck', 'face', 'arms', 'thigh', 'calf', 'knee' and 'bones.'" (Abbi 2013: 141). In addition, the personal interrogative has the class 1 possessive marker *a-* also found on words referring to the mouth and kinship terms, such as mother. As will be shown below, the Tungusic possessive classification marker cannot enter the PNQ.

In Type B constructions, there is variance in the type of SPEECH ACT VERB that is involved. Apart from the language-specific semantics, the most important variation concerns the valency of the verb. In German, *heißen* 'to be called' is an intransitive verb and *nennen* 'to call' is a transitive verb. In Mandarin, *jiào* is an ambitransitive verb that can be either intransitive or transitive (Table 3).

Table 3: Valency of speech act verbs in German and Mandarin. In German, the transitive or causative use of *heißen* is archaic.

	Intransitive (+ NAME)	Transitive (+ NAME)
German	heißen	-
	-	nennen
Mandarin	jiào	jiào

- (42) a. Mandarin (Sino-Tibetan)
wǒ jiào ānnà.
 1SG call PN
 ‘I am called Anna.’ (intransitive)
- b. *tā jiào wǒ ānnà.*
 3SG call 1SG PN
 ‘(S)he calls me Anna.’ (transitive)

English requires a passive, a reflexive, or a third person plural dummy agent in order to use the verb to call as an intransitive verb, e.g. *he is called Joe*, *he calls himself Joe*, *they call him Joe*. A reflexive or a passive of a speech act verb are also possible in German.

- (43) German (Indo-European)
Wie nenn-st du dich?
 how call-2SG.PRS.IND 2SG 2SG.REFL
 Literally: ‘How do you call yourself?’⁴
- (44) German (Indo-European)
Wie wirst du genannt?
 how get.2SG.PRS.IND 2SG called
 Literally: ‘How are you called?’

An impersonal construction is also attested in other languages with Type B constructions.

- (45) Beng (Mande; Paperno 2014: 17)
ouo mi si po?
 3PL.HAB.AFF 2SG call.L what
 Literally: ‘What do they call you?’

⁴This is identical to the original of the question in the quotation from Goethe above.

Changing of valency, reflexives or impersonals are not attested in the few cases of Type B constructions in Tungusic.

POLITENESS is a dimension of variation that plays a larger or smaller role for both types of PNQs depending on the language. In German, there is a two-way politeness distinction that affects the choice of the pronoun and, consequently, the verbal ending. Instead of the usual *du* 'you (SG)', the polite pronoun *Sie* 'you (SG.POL)' is used. Both have suppletive case forms.

- (46) German (Indo-European)
Wie heiß-en *Sie?*
 how be.called-2SG.PRS.IND.POL 2SG.POL
- (47) German (Indo-European)
Was/Wie ist Ihr *Name?*
 what/how is 2SG.POL.GEN.SG.M.NOM name.SG.M.NOM

While German makes use of the same two constructions, there are languages that change the whole construction according to the politeness register. Two such languages that had contact with Tungusic languages are Korean (see above) and Mandarin. Mandarin, apart from the other expressions mentioned throughout this section has the following honorific form that is based on a different pattern.

- (48) Mandarin (Sino-Tibetan)
nín guì xìng?
 2SG.HON honorable surname

In Koreanic languages, apart from the use of a different construction seen above, there is also a distinction in the question marker.

- (49) Soviet Korean (Koreanic; King 1987: 253, 269, slightly adjusted)
- a. *irimi misi-ge-ja?*
 name what-thing-Q.PLAIN
 - b. *irimi misi-ge-mdu?*
 name what-thing-Q.POL

Politeness could also have led to some exceptions from the proposed universal that all languages have a conventionalized way of expressing the PNQ. Jiaomuzu Gyalrong in China, for instance, tends "to avoid direct address", including questions about names. However, even in this language it is possible to ask a PNQ in a polite way:

(50) Jiaomuzu Gyalrong (Sino-Tibetan; Prins 2017: 343)

nənʃo t^{hi} tə-rɲu-n ko?

2SG what 2-be.called-2SG ANX

‘Please, do tell me what is your name?’

Overall, Tungusic languages have few grammaticalized expressions for politeness.⁵

3 The personal name frame

The semantic side of a construction, like that of a lexical item, can be represented by what is often referred to as a frame (e.g., Fillmore 1985). This section introduces the *personal name frame* (PNF) that could be the basis for the personal name question. This frame can be illustrated with dialogues from the Tungusic language Sibe.

(51) Sibe (Jin 1993: 3)

a. **tʃunfu/Chunfu:**

ci χodz=na?

2SG good=Q

‘How are you?’

b. *nənə-m mi.n-j bəji-v bəji-d əmdan taqə-vi-ki.*

first-CVB.IPFV 1SG.OBL-GEN self-ACC self-DAT once know-CAUS-DES

‘First, let me introduce myself.’

c. *mi.n-b tʃunfu ʃi-m.*

1SG.OBL-ACC PN say-IPFV

‘My name is Chunfu.’

d. **tʃaŋmiŋ/Changming:**

ci.n-b taqə-m mutu-xuŋ bi ursun bailə-m.

2SG.OBL-ACC know-CVB.IPFV can-PTCP.PFV 1SG very rejoice-IPFV

‘I am very happy to meet you.’

e. *mi.n-j gəvə-v tʃaŋmiŋ ʃi-m.*

1SG.OBL-GEN name-ACC PN say-IPFV

‘My name is Changming.’

⁵While some Koreanic question markers that show politeness distinctions were possibly borrowed by the Jurchenic branch of Tungusic (Hölzl 2018a: 213), their exact function in Jurchenic still remains unclear.

While this brief dialogue does not contain the personal name question, it is arguably located in a very similar type of situation. While the direct question about the name is avoided by Chunfu, Changming, by means of pragmatic inference, draws the conclusion that, given Chunfu's introduction, it is appropriate to say one's own name in response. In a similar albeit more direct way, one can add a truncated question at the end of one's own introduction:

- (52) Sibe (Jin Ning 1993: 3)
mi.n-j gəvə-v sarasu ʃi-m. ɕi ni?
 1SG.OBL-GEN name-ACC PN say-IPFV 2SG Q
 'My name is Sarasu. What's yours?'

As another example consider the following dialogue:

- (53) Sibe (Jin 1993: 4)
- a. **dzalušan/Zhalušan:**
ɕi mi.n-d əmdan taqə-və-Ø!
 2SG 1SG.OBL-DAT once know-CAUS-IMP
 'Would you introduce me to him please?'
- b. **bəkdəsu/Bekdesu:**
bi so.n-j dzu nanə-v əmdan taqə-vi-ki.
 1SG 2PL.OBL-GEN two person-ACC once know-CAUS-DES
 'Allow me to introduce you.'
- c. *ər əmkən=ni goɕiŋa si-m [...]*
 this one=3SG.POSS PN say-IPFV
 'This is Gosinga.'

In this case, the situation involves not two, but three persons. Apart from the two people making the acquaintance (Zhalusan and Gosinga), there is a third mediating person (Bekdesu).

All three situations above are based on the common background knowledge that everybody has a name. The same is obviously true for the personal name question. But this is only part of the larger personal name frame that contains several subevents and roles tentatively listed in Table 4.⁶

⁶The list presented in Table 4 is probably not exhaustive and the individual subevents could be slightly different depending on the cultural background. For instance, in some societies names can also be removed from a person (e.g., Moutu 2013: 147). Apart from giving, a name can evolve through a process known as onymization (Van Langendonck & Van de Velde 2016: 33). Future studies will have to revise the personal name frame accordingly.

Table 4: The personal name frame and its subparts. The dummy noun meaning ‘name’ is not listed, but is optionally present in all subevents (based on Hölzl 2014)

	Subevents	Roles
1	GIVING A NAME	NAMEE, NAME, NAMER
2	HAVING A NAME	POSSESSOR, NAME
3	ACQUAINTING	
3a	2 persons: ASKING, TELLING	ASKER, ADDRESSEE, NAME
3b	3 persons: INTRODUCING	INTRODUCER, PERSON A, PERSON B, NAMES
4	KNOWING A NAME (common ground)	KNOWER, KNOWN, NAME
5	CALLING BY NAME	CALLER, CALLED, NAME

First, most people do not usually chose their names on their own, but are given the name by somebody else, such as their parents. In this case, there are three different roles, the person giving the name (NAMER), the personal name given (NAME), and the person being named (NAMEE). There are culture- and language-specific conventions and examples for each of these subevents. In this case, this could be a baptism, the acceptance of a new name during a religious initiation, or the change of one’s own name in court.

Second, everybody has or owns a name. Here the roles are the person having the name (POSSESSOR), and the name (NAME). Cross-linguistically, this frame is usually expressed with possessive relationships, e.g. *her name* (attributive possession), *she has a beautiful name* (predicative possession). But because a name is not a concrete and tangible object, these expressions are based on an underlying conceptual metaphor that IDEAS ARE OBJECTS (Lakoff & Johnson 1999: 124f.). This can also be seen in other expressions, e.g. *my plan* or *to have a plan*.⁷ A culture-specific case can be found among the Iatmul in Papua New Guinea who “believe that there is a mystical connection between a name and its bearer” (Moutu 2013: 147).

Third, there are at least two subevents for making the acquaintance of a person that correspond to the two dialogues from Sibe above. These include either

⁷In addition, the conceptual metonymies that the NAME STANDS FOR A PERSON and that the FACE STANDS FOR A PERSON are often combined with this, e.g. in a passport. For instance, when looking at a photo of a person’s face it is possible to say *This is Sam*.

two persons (three roles: ASKER, ADDRESSEE, NAME) or three persons (four roles: INTRODUCER, PERSON A, PERSON B, NAMES).

Fourth, after giving a name or after having made the acquaintance of a person, one has the knowledge of that person's name. This subevent has three roles, the person knowing the name (KNOWER), the person whose name is known (KNOWN), and the name (NAME). Knowing other people's names is part of the common ground. Forgetting somebody's name can lead to severe social awkwardness. Depending on the society, a certain amount of control can for instance be associated with knowing a person's name.

Fifth, when knowing a person's name, one (the CALLER) can refer to that person (CALLED) by his or her name (NAME), either in a direct address (vocative) or in the third person. The name theoretically identifies the exact individual. Depending on the type of naming in a given culture, namesakes can lead to more or less problems (see Moutu 2013: 145ff. for an extreme example). Conversely, one person can have several different names. In certain cases, uttering a specific name can be a taboo.

The PNQ is part of the acquainting subevent, more specifically subevent 3a, but is based on several aspects of the personal name frame. Questions of Type A combine 3a with subevent 2 (HAVING A NAME), and Type B with subevent 5 (CALLING BY NAME). There is a mapping of the roles of the two combined subevents (Table 5). In addition to the roles, the three subevents also contain semantic relations not specified above that can be indicated as ASK (a type of question), CALL (a form of speech act), and HAVE (a possessive relationship), respectively.

Table 5: Combinations of subevents and roles in the two main PNQ types

Type	Subevent 1	Subevent 2	Combined Roles/Relations
Type A	ACQUAINTING (2 persons)	HAVING A NAME	ASKER = \emptyset , ADDRESSEE = POSSESSOR NAME = NAME ASK + HAVE
Type B	ACQUAINTING (2 persons)	CALLING BY NAME	ASKER = CALLER ADDRESSEE = CALLED NAME = NAME ASK + CALL

Using Langacker's (2008: 66) terminology, one could say that different PNQs highlight or *profile* different aspects of the underlying frame that functions as a base. For instance, even though Type B does not necessarily refer to subevent 2 (HAVING A NAME), a speaker must still be aware of it in order to ask the question in the first place.

4 'What's your name?' in Tungusic

The question 'What is your name?' has been recorded for the majority of the Tungusic languages and in a considerable number of dialects. To the best of my knowledge, the PNQ is not documented in Arman, Bala, Lalin/Jing Manchu, the two Jurchen varieties, and Kili (Kur-Urmi Nanai). However, for all these languages, similar constructions or at least individual words, such as 'name' are attested. Only for Chinese Kyakala there is no information on the PNQ at all.

As expected, Tungusic languages show a certain amount of variation in how they express the question. Nevertheless, all constructions exhibit a cognate of the Tungusic word for 'name'. This word functions as some kind of *anchor* around which all PNQs are built. One example with the optional Mongolic word apart, no other word for 'name' is attested in these constructions. This lexical item is addressed in §4.2.

4.1 Second person pronoun and genitive

All Tungusic languages preserve cognates of Proto-Tungusic **si* 'you (SG)' (e.g., Benzing 1956: 109). There are some well-known phonological changes, such as *s* > *ɕ* before *i* in some Jurchenic varieties, or *s* > *h* in some Even dialects. The PERSONAL PRONOUN can often be absent and is less central for the personal name question. Apart from Jurchenic, Tungusic languages also employ a grammaticalized version of this personal pronoun as possessive marker as in the following example from Ulcha (54) (see Ikegami 1985 for details):

- (54) Ulcha (Angina 1993: 3)
si(ə) gəlbʊ-si nguʃ?
2SG name-2SG.POSS who

In Proto-Tungusic, the personal pronoun **si* has an oblique form **si.n-*, for example for the genitive **si.n-i*. The presence of the *-n-* in oblique forms is a phenomenon found throughout the pronominal system of Tungusic and neighbouring languages, such as Mongolic. The genitive is retained, for example, in written

Manchu *si.n-i* '2SG.OBL-GEN' and *suwe.n-i* '2PL.OBL-GEN'. In some languages the genitive *-i* changed to *-u* in the plural pronouns due to a progressive vowel assimilation, e.g. Uilta *si.n-i* 'your (SG)', but *su.n-u* 'your (PL)' (Tsumagari 2009b: 7). In a few languages, for example in Even (*hi.n*) and Bala (*ci.n*), the oblique form was retained in genitive function, although the genitive itself was lost. In several other languages, such as Udihe, the genitive was functionally lost, but still functions as a stem for the possessive forms, e.g. *si.n-i-ŋi* 'yours (SG)', *su.ñ-u-ŋu* 'yours (PL)' (Nikolaeva & Tolskaya 2001: 336). In some languages, possessive forms of this sort developed a meaning similar to a genitive (e.g., 11), which led to a probably erroneous reconstruction of the genitive in Benzing (1956: 79).

4.2 The Tungusic word for 'name'

Traditionally, the Tungusic word for 'name' is reconstructed as **gärbü* (Benzing 1956: 49). While this reconstruction is reasonably robust, it is slightly misleading as the reconstructed **ä* must actually have been pronounced as schwa [ə], as in the majority of the modern languages. Janhunen (1991: 40), perhaps based on Khamnigan Evenki *gärbii*, reconstructs Tungusic **gerbüü* with a long vowel in the second syllable. While a long vowel can also be found in other Evenki dialects, for example Sakhalin Evenki *gärbī* (Bulatova & Cotrozzi 2004) or Nercha Evenki *gärbī* (Khabtagaeva 2022 [this volume]), this seems to be an innovation rather than a retention. Cognates of **gärbü* 'name' are collected, among others, in Schmidt (1923a,b, 1928a,b), Benzing (1956: 49), Cincius (1975/77: 180f.), Lie (1978: 143), Kazama (2003: 68), Doerfer & Knüppel (2004: 336), or Chaoke (2014c: 300f.).

The earliest recordings of Tungusic are in Jurchen, which is a cover term for at least two different varieties that, for lack of better terms and in analogy to similar cases such as Tocharian, can be called Jurchen A (**gebu* 革卜, Kiyose 1977) and Jurchen B (**gebu* 革不, Kane 1989). The word recorded for these two varieties of Jurchen are identical to written Manchu *gebu*, which is attested from the 17th century onward (e.g., Norman 2013). Apart from Jurchen and Manchu, some of the oldest records of the word for 'name' have been made for Evenki and Even. For instance, at the beginning of the 18th century Witsen (1705: 654) mentioned Evenki *gerbisch* 'your name', which can be analyzed as *gerbi-f* 'name-2SG.POSS'. Pallas (1786, 1789: 169) listed *gorbi*/горби for Evenki dialects and *gerbi-n*"/гербинъ for Even. A form *garbi-n* was recorded in 1808 by Koshewin (von Klaproth 1817: 224). To mention but some more examples, the word has been recorded as *gärbī* or *garbi-n* in 1810 by Spassky (Castrén 1856: 107, 128). Schiefner already correctly equated Evenki *gärbī* with Manchu *gebu* (see Castrén 1856: x). Two of the earliest recordings of the word in Nanai (specifically the Ussuri dialect) in the 19th

century are *gerbi*/repби or *gerbu*/repбy (Brylkin" 1861) and *gorbi-ni* (Venukoff 1862; Alonso de la Fuente 2011: 20). The Nanai form *gerbú* listed in the dictionary by Grube (1900) was also collected around the middle of the 19th century. For many other languages, data are only available from the 20th century onward.

The reconstructed **ü* in **gärbü* 'name' underwent a regular sound change to *i* in Northern Tungusic languages (Ewenic and Udegheic) and to *u* in Southern Tungusic (Nanaic and Jurchenic), e.g. Oroqen *gərbi*, Oroch *gəbbi*, but Nanai *gərbu*, Manchu *gebu*. The same sound change can be seen in the interrogative **ŋüi* 'who', e.g. Oroqen *nii*, Oroch *n'ii*, but Nanai *uj* (Uilta *ŋui*), Manchu *we* (see also Hölzl 2018a: 314). Only Even (*gərbə*), Arman (*gerbu*, *gurbu*), and one recording of Oroqen or Solon (*gerbu* in Ivanovskiy 1982 [1894]: 1) might represent special cases in Northern Tungusic. However, other recordings of Oroqen and Solon as well as the Even form *gerbi-* recorded by Pallas (1786, 1789: 169) contain the expected *i* (cf. also Arman *ŋii* and Even *ŋi(i)* 'who'). Apart from that, there have been several language-specific developments. The *r* has been, probably regularly, lost in Jurchenic (e.g., Lalin/Jing Manchu *gəbu*) and changed to *l* in several languages around the lower Amur, including Uilta (*gəlbü*), Ulcha (*gəlbü*), and Lower Negidal (*gölbi* [gəlbɪ], Schmidt 1923a: 18, *gilbi* with additional regressive vowel assimilation, Khasanova & Pevnov 2003: 7). The *l* is already attested in data collected at the beginning of the 20th century, i.e. Uilta *gylbó-ni/gylbú(-ni)*, Ulcha *gýlbü* in Piłsudski (Majewicz 2011: 258, 817) and Ulcha *gölbu* [gəlbɪ] in Schmidt (1923b: 251). The consonant cluster **rb*, possibly via **lb*, developed into a cluster *db* in Upper Negidal (*gədbi*, Natalia Aralova p.c. 2019), *gb* in Bikin Udihe (*gegbi*), and into the geminate *bb* in Oroch (*gəbbi*). Huihe Solon *gəbbi* also has a geminate, but other Solon dialects preserve the consonant cluster *-rb-*, e.g. Ongkor Solon *gerbi* (Aalto 1977: 63). These are mostly regular changes with parallels, for example, in the cluster **lb* as in Proto-Tungusic **dolba* 'night', e.g. Manchu *do-bo(-ri)*, Bikin Udihe *dogbo*, Oroch *dobbo* etc. (Benzing 1956: 46; Kazama 2003: 50; Doerfer & Knüppel 2004: 234).⁸ In a few recordings, an epenthetic vowel seems to have been inserted (either by the speakers themselves or the researchers) to avoid the consonant cluster (e.g., Oroqen or Solon *geribé* in Ivanovskiy 1982 [1894]: 1, Uilta *geribu* in Nakanome 1928: 52). The consonant cluster as such is preserved in several Ewenic (e.g., Evenki *gərbi*) and Nanaic languages (e.g., Samar *gərbu* [gərbɪ], Schmidt 1923a). In Jurchenic, the final vowel was sometimes lost and the *b* underwent regular intervocalic spirantization in several Manchu dialects

⁸Some languages show a slightly different pattern for **lb*. For instance, one subgroup of Jurchenic preserved a reflex of the *l*, i.e. Bala *dəlbə* (Mu 1987: 17), Jurchen A 多羅斡 [duo luo wo] (Kiyose 1977: 101), etc.

both in Dzungaria (e.g., Sibe *gəv(ə)*) and Manchuria (e.g., Aihui Manchu *gəvo* ~ *govo*, Yibuqi Manchu *kowə*, Shenyang Manchu *gef(u)*, Sanjiazi Manchu *gəwu*). Alchuka represents a special case not only in Jurchenic, but in all of Tungusic due its occasional loss of the initial consonant, i.e. *?əwə* (Mu 1986: 14). While the word has also been recorded as *gəbu* (Mu 1987: 14), the form *?əwə* is not necessarily an error (although the *ə* is potentially a misprint for *g*). The language is known to have lost word initial consonants and exhibited a certain amount of internal variation that is poorly understood. Similar variation is known from other dialects, such as that from Sanjiazi. As opposed to the form *gəwu* in Kim et al. (2008) that was collected in 2005/06, Enhebatu (1995) in 1961 recorded the form *gur:bu* instead. While some of the discrepancies are probably a mere byproduct of the transcription (e.g., *ur* instead of *ə*), there are certainly also actual differences in the forms, for example the presence or absence of spirantization. For Chinese Kyakala, no cognate of the word for ‘name’ appears to have been recorded (Hölzl 2018c; Hölzl & Hölzl 2019).

Some languages, in addition to the autochthonous reflex of **gärbü*, have borrowed the Manchu word, but with a special semantics (e.g., Benzing 1956: 18, 49; Alonso de la Fuente 2011: 27; Khabtagaeva 2022 [this volume], Table 6). This led to doublets, such as Udihe *gegbi* ‘name’ vs. *gebu* ‘honor’ (Nikolaeva & Tolskaya 2001). The latter word must represent a borrowing because an intervocalic *b* is otherwise only retained in Jurchenic (e.g., Benzing 1956: 34).

Table 6: Manchu *gebu* ‘name’ in other Tungusic languages

Language	Name	Source
Kili	<i>gəbu</i>	Sunik 1958: 170
Nanai	<i>gəbu</i>	Benzing 1956: 18
Oroch	<i>gəbu</i>	Avrorin & Lebedeva 1978: 175
Udihe	<i>gebu</i>	Nikolaeva & Tolskaya 2001: 930
Ulcha	<i>gəbu</i>	Benzing 1956: 18
Ussuri Nanai	<i>gəbə</i>	Sem 1976: 150

The Manchu borrowing in other Tungusic languages usually has a slightly different meaning, such as ‘honor’, which makes it less important for the purposes of this study. A similar doublet can be found, for instance, in Kili (Kur-Urmi Nanai), i.e. *gərbi* ‘name’ (Sunik 1958: 116) vs. *gəbu* ‘honor, authority, respect’ (Sunik 1958: 170). But in this case, both forms are a borrowing from another language. Apart from Kili, also Bala, Kilen, and Ussuri Nanai must have borrowed

the word for ‘name’ from a Northern Tungusic and more exactly an Ewenic language. For Bala, this was misinterpreted by Mu (1988: 17) as an autochthonous development. But clearly, the words are from a form similar or identical to Evenki (see Table 7). If these were not borrowings, in all four languages the final vowel should be an *u* as in Manchu *gebu* or Nanai *gərbu*.⁹ Brylkin” (1861: 12) recorded both *gerbi* (borrowed) and *gerbu* (autochthonous) among the Ussuri Nanai.

Table 7: The Ewenic word for ‘name’ (e.g., Evenki *gərbi*) in Southern Tungusic

Language	Name	Source
Bala	<i>gərbi</i>	Mu 1987: 14
Kilen	<i>gerbi</i>	Dong 2016: 337
	<i>gərbi</i>	An 1986: 96
Kili	<i>gərbi</i>	Kazama 2003: 68
	<i>gərbi</i>	Sunik 1958: 116
Ussuri Nanai	<i>gərb’i</i>	Sem 1976: 150
	<i>gorbi-ni</i>	Venukoff 1862
	<i>gerbi</i>	Brylkin” 1861: 12

In many languages, **gärbü* is the basis for the derivation of verbs, e.g. Manchu *gebu-le-* ‘to name, to call by name’, Uilta *gəlbulle-* ‘to give a name to’, Udihe *gegbi-si-* ‘to call’, Evenki *gerbi-te-* ‘to be named’ etc., but these are not often encountered in the personal name question.

Among Tungusic languages, only Jurchenic has a gender-like distinction. Even in Jurchenic, this is restricted to a few nouns that show an ablaut phenomenon, e.g. Manchu *haha* ‘man’, *hehe* ‘woman’. The Manchu word *gebu* ‘name’ does not belong to this set of nouns.

All branches of Tungusic except for Jurchenic have a limited system of possessive classification, making use of what is usually referred to as alienable possessive marker, e.g. Udihe *-ŋi*, Uilta *-ŋu* etc. For instance, the noun *dili* ‘head’ in Udihe can be used with and without *-ŋi* (Nikolaeva & Tolskaya 2001: 135). The word for ‘name’ does not belong to the set of nouns that can be marked with the suffix, i.e. it is probably not conceptualized as alienable.

⁹Additionally, the *r* would perhaps have to be absent in the Bala form as in Manchu *gebu*, although Bala is more conservative than Manchu in this particular feature, e.g. Bala *bardi-*, Manchu *banji-* ‘to live’, Bala *dördi-*, Manchu *donji-* ‘to hear’ (Mu 1987, slightly corrected).

4.3 Ewenic

The question is known from all Ewenic languages, with the exception of Arman. Almost all Ewenic examples below are copula sentences (Type A). In Even, two different patterns are attested, but both contain the same interrogative meaning 'who'. Consider the following two question-answer sequences:

(55) Eastern Even (Beryozovka; Kim 2011: 189, corrected)

- a. *hi ŋi gərbə-s?*
2SG who name-2SG.POSS
- b. *mi.n gərbə-w garpək.*
1SG.OBL.GEN name-1SG.POSS PN

(56) Lamunkhin Even (Natalia Aralova p.c. 2019)

- a. *hi gərbə-s ɲi:?*
2SG name-2SG.POSS who
- b. *bi gərbə-w taisiya.*
1SG name-1SG.POSS PN

In both examples, the question makes use of the nominative form of the personal pronoun. In Lamunkhin Even, not even the answer exhibits the genitive. Notably, only the interrogative, but not the personal name of the answer can stand before the word for 'name'. Because the person is already marked on the head noun, the personal pronoun can be absent in Even and, as will be seen, in several other Tungusic languages.

Given the overall similarity of Arman to Even, the question might have been very similar as well. The individual elements of the Even examples above have the following form in Arman: *ši* '2SG', *ni* 'who', *gerbu*, *gurbu* 'name, title etc.', *-s/-SI/-čI* '-2SG.POSS' (Doerfer & Knüppel 2013: 28, 133, 138, 228, 302f., transcription slightly changed). Consequently, the question might have been something like **ši ni gerbu-s?* or **ši gerbu-s ni?* (constructed). However, only the following example with a verb derived from *gerbu* is attested in the material available to me:

(57) Arman (Doerfer & Knüppel 2013: 30, transcription slightly changed)
təməj tək gerbuutte kamčidalal'ji.
therefore now call.NFUT[3PL] PN
'Therefore, they now call themselves Kamchadals.'

The same possibility of the interrogative to precede or follow the word for ‘name’ as in Even is also observed in Evenki. The following example from the Sakhalin dialect has the interrogative after the word for ‘name’ (the same can be found in Konstantinova 1964: 41). As early as the 19th century an example with a preposed interrogative has been recorded.

- (58) Sakhalin Evenki (Bulatova & Cotrozzi 2004: 58)

gərbi-s **ŋi?**
name-2SG.POSS who

- (59) Viljuj Evenki (Maak” 1859: xviii; Schiefner 1878: 144)

ni *gərbi-s?*
who name-2SG.POSS

The absence of the personal pronoun (*si* in Maak” 1859: xix, *sī* in Bulatova & Cotrozzi 2004: 58) is also attested in Even.

- (60) Even (Idiatov 2007: 307)

gerbe-s **pi?**
name-2SG.POSS who

- (61) Even (Benzing 1955: 176)

ŋi *gərbə-s?*
who name-2SG.POSS

This can also be observed in other Evenki recordings, such as the following example from the Eastern dialect:

- (62) Eastern Evenki (Makarova 1999: 16)

- a. **ŋi:** *gərbi-s?*
who name-2SG.POSS
- b. *bi:* *gərbi-v* **ando.**
1SG name-1SG.POSS PN

Similar to Even above, the interrogative stands in a focus position before the dummy noun while the personal name in the answer follows. Seemingly, the same asymmetry of the question and the answer has also been recorded for Aoluguya Evenki in China.¹⁰

¹⁰The analysis by Hasibate’er (2016: 278) is *cini*, i.e. *ci.n-i* ‘2SG.OBL-GEN’, which leads to an example without interrogative, which is unlikely.

(63) Aoluguya Evenki (Hasibate'er 2016: 278)

- a. *ci ni gərbi-ci?*
2SG who name-2SG.POSS
- b. *bi gərbi-w mərə.*
1SG name-1SG.POSS PN

By comparing Even and Evenki dialects with the close relative Oroqen in China, a very similar pattern with the interrogative in second position can sometimes be observed.

(64) Evenki (Boldyrev 2000: 134)

- si ŋi gərbi-s?*
2SG who name-2SG.POSS

(65) Oroqen (Chaoke 2014a: 8)

- shi ni gerbi-shi?*
2SG who name-2SG.POSS

This suggests a relatively high age of this phenomenon among Ewenic languages.

All examples given so far contain a cognate of the Tungusic interrogative **ŋüi* 'who'. The same interrogative can also be found in the personal name question of some Udegheic and Nanaic varieties, but not in Jurchenic. Apart from Even and Evenki, many Ewenic languages also employ different interrogatives. In most Solon dialects, **ŋüi* has been replaced by a selective interrogative meaning 'which (one)' that is also found in the personal name question.

(66) Huihe Solon (Tsumagari 2009a: 15)

- si.n-ii gebbi-si aawu?*
2SG.OBL-GEN name-2SG.POSS who

This latter construction has an exact parallel in the following Dagur example, although the use of the nominative *ši*: 'you (SG)' is also possible.

(67) Tacheng Dagur (Khitano-Mongolic; Yu et al. 2008: 173)

- šin nər-šin anja?*
2SG.OBL.GEN name-2SG.POSS who

Both Solon and Dagur have an innovative personal interrogative that replaced Tungusic **ŋüi* 'who' and Mongolic **ken* 'who', respectively. This innovation in Solon appears to have later spread to Oroqen. This interrogative is already attested in the recordings by Ivanovskiy from the end of the 19th century that are usually taken to represent Solon (e.g., Lie 1978).

- (68) Butkha Solon (Ivanovskiy 1982 [1894]: 1)¹¹
geribé agó?
name who

Unlike Huihe Solon, however, no geminate can be found in the word *geribé* ‘name’. In fact, Ivanovskiy mentions three additional expressions, all of which appear to be closer to Oroqen than Solon:

- (69) “Manegir” (Ivanovskiy 1982 [1894]: 1)
- a. *šl.n-i gerbu ní?*
2SG.OBL-GEN name who
 - b. *ni gerbu bí-či?*
who name COP-?PRS
 - c. *jému gerbi-čí?*
which name-POSS

Notably, two of the examples still have a cognate of **ŋüi* ‘who’ that shows the same syntactic behavior as in Even and Evenki. Alternatively, *neré* ‘name’ is said to be used in (69a), which is the Mongolic word (see examples 20, 67, 73, 119, 125).

Examples (69a) and (69b) are also similar to Even and Evenki, although they appear to lack a possessive marker. The second example is one of the few examples among Tungusic languages that has an overt copula in a Type A construction. A copula is also present in a more recent example from Oroqen that shares the absence of the possessive marker as well as the interrogative of the last example (69c) from Ivanovskiy.

- (70) Xunke Oroqen (Zhang, Yanchang, Li Bing, et al. 1989: 141)
ci: jEma gərbi bi-ci-ni?
2SG which name COP-PRS-3SG

Phonological differences apart, the following two Oroqen sentences are identical to (69c) (see also 106 from Kilen). Some Ewenic languages, such as Oroqen, use the comitative or possessive suffix instead of the second person possessive marker. These are sometimes difficult to differentiate.

- (71) Gankui Oroqen (Sa 1981: 51)¹²
yam gerbi-qi?
which name-POSS

¹¹What is tentatively transcribed as *-g-* here remains partly unclear.

¹²The <q> in this transcription is based on the Chinese Pinyin system, where it stands for [tɕʰ].

- (72) Shengli Oroqen (Han & Meng 1993: 303)

jeema kərpi-tʃi?
which name-POSS

Ivanovskiy (1982 [1894]: 3) mentions two Dagur examples, one of which contains a selective interrogative that might have influenced the choice and position of the interrogative in Oroqen, although the two are probably not etymologically related.

- (73) Dagur (Khitano-Mongolic; Ivanovskiy 1982 [1894]: 3)

si jamár neré?
2SG which name

The same interrogative as in Oroqen is also found in an example from Negidal, albeit in a different syntactic position. This is not the same variation as observed for *ŋüi* 'who', however, because this selective interrogative has an attributive function if preceding the dummy noun. In other words, we are dealing with a Type A.1 construction in Negidal (74), but with a Type A.2 construction in Oroqen (69c, 70, 71, 72).

- (74) Lower Negidal (Kazama 2002: 80)

sii gilbi-si eema?
2SG name-2SG.POSS which

Oroqen and Evenki dialects in China also make use of a THING interrogative, potentially influenced by languages such as Manchu or Chinese. The following two examples likewise are instances of Type A.1 (75) and Type A.2 (76), respectively:

- (75) Khamnigan Evenki (Tsumagari 1992: 96)

ər-nii gərbii ikun?
this-GEN name what
'What is the name of this?'

- (76) Khamnigan Evenki (Chaoke & Kajia 2016: 9)

su ikon gərbii-tsi wee?
2PL what name-POSS CQ

The use of the interrogative *ikun* in (75) might be due to the fact that it does not refer to the name of a person.

One Solon dialect employs *oni* 'how', which might be due to Russian influence (see 26). Given that this interrogative cannot be used attributively, the example contains fronting as in other Ewenic languages.

(77) Arong Solon (Chaoke & Kalina 2017: 17)

fi oni gəbbi-fe?
2SG how name-POSS

The use of MANNER interrogatives is more common in Udegheic and Nanaic but can also be observed in one recording of Negidal. In the following examples, the interrogative *o:n* either stands in the unexpected sentence-initial position even before the personal pronoun or in the same position as the proper name in the answer.

(78) Upper Negidal (Natalia Aralova p.c. 2019)

- a. *o:n si gədbi-s?*
how 2SG name-2SG.POSS
- b. *si gədbi-s o:n?*
2SG name-2SG.POSS how
- c. *bi gədbi-β Antonina*
1SG name-1SG.POSS PN

The sentence-initial position of the interrogative in front of the pronoun, which is otherwise unattested in the PNQ in Tungusic, is clearly due to Russian influence and is a typical European feature (Dryer 2013).

Oroqen and Solon have been more strongly influenced by Mongolic languages than most other Ewenic language. In both languages, there is an alternative Type B construction that is often found in answers to the personal name question. The Type A.2 construction, as in Jurchenic, lacks the genitive in Oroqen.

(79) Nanmu Oroqen (Chaoke 2007: 140, corrected)

- a. *fi ikon gərbi-tfe?*
2SG what name-POSS
- b. *mi.ŋi gərbi-wi tumbutfə gunən.*
1SG.OBL.GEN name-1SG.POSS PN say.3SG

(80) Solon (Chaoke et al. 2014: 8)

- a. *shi.n-i gebbi-shi awu?*
2SG.OBL-GEN name-2SG.POSS who
- b. *mi.n-i gebbi-wi ... gənəŋ.*
1SG.OBL-GEN name-1SG.POSS (PN) say.3SG

This construction appears to be impossible in the PNQ with the transitive verb *gun-* ‘to say’ in Evenki and other Ewenic languages. Another Type B construction, although calqued from Russian, is found in Negidal. Similar to the Arman example above, the verb is derived from the word *gədbi* ‘name’.

- (81) Upper Negidal (Natalia Aralova p.c. 2019)
mi.nə-βə gədbittəə Ton’a
 1SG.OBL-ACC call.NFUT[3PL] PN

Ivanovskiy (1982 [1894]) recorded an answer without a speech act verb.

- (82) “Manegir” (Ivanovskiy 1982 [1894]: 1)
mi.n-í gerbú ...
 1SG.OBL-GEN name (PN)

Although ellipsis cannot be ruled out, this might be additional evidence that the Type B construction is a recent innovation in these languages.

4.4 Udegheic

For both Oroch and Udihe several different expressions have been recorded. Except for the following Type B example, Udegheic makes use of copula sentences. Example (83a) from Udihe seems to be entirely based on Russian while the answer (83b) is similar to Ewenic languages and represents the original Tungusic construction.

- (83) Udihe (Tsumagari 2011: 81, 85)
 a. *si.n-awa ono gegbi-si-ti?*
 2SG.OBL-ACC how name-V-3PL
 b. *bii gegbi-i Tausima.*
 1SG name-1SG.POSS PN

Some of the oldest examples for Udegheic have been recorded around 1900 by Brailovski. Schmidt corrected the sentences, but misinterpreted *ŋi* ‘who’ in (84) as a possessive marker. It is an interrogative that derives from **ŋüi* instead.

- (84) Oroch (Bochi river; Schmidt 1928a: 20, from Brailovski, corrected)
si gabi ŋi?
 2SG name who

- (85) ? Udihe (Samarga river; Schmidt 1928a, from Brailovski, corrected)
si gabi jav?
2SG name what
- (86) Udihe (Nakhtu river; Schmidt 1928a, from Brailovski, corrected)
si gagbi jau?
2SG name what

It is unclear whether the last example (86) might contain a fused second person possessive marker *-(h)i* (< **-si*) as in the following modern examples from the Khor and Bikin dialects (see also Perekhval'skaya 2022 [this volume], on intervocalic *s* and its reflexes in Udegheic):

- (87) Khor Udihe (Elena Perekhval'skaya, p.c. 2019)
si gʌgbi-hi j'əu?
2SG name-2SG.POSS what
- (88) Bikin Udihe (Nikolaeva & Tolskaya 2001: 804)
si gegbi-i j'eu?
2SG name-2SG.POSS what

The use of a personal interrogative (Udihe *ni(i)*, Oroch *n'ii*) seems to be much more restricted than in Ewenic and Nanaic. Apart from *j'ə-u* 'what' (*ja-v* and *ja-u* in Brailovski), which is cognate with Oroqen *i-kon*, and Khamnigan Evenki *i-kun* or *i-kon* above, Udihe can also employ *ono* (< **oni*) 'how' in the same construction.

- (89) Udihe (Tsumagari 2006: 6)
sii gegbi-i ono?
2SG name-2SG.POSS how

Oroch also uses a cognate of this interrogative. In the following example, there is an additional overt copula that is not usually found in the Udihe examples (see §4.5 on Nanaic). As in Ewenic, the personal pronoun can be absent.

- (90) Oroch (Avrorin & Lebedeva 1978: 175)
gʌbbi-si ɔn'i bi?
name-2SG.POSS how COP

While Oroch also has a construction without a copula, according to one author a different interrogative meaning 'how' can be employed.

- (91) Oroch (Lopatin 1957, corrected)
si gaby-si yavanká/yanká?
2SG name-2SG.POSS how

In sum, the Udegheic PNQ shows a strong tendency for Type A and more specifically Type A.1. As opposed to Ewenic, Type A.2 is not attested and one Type B construction in Udihe can be plausibly explained by Russian influence. Apart from this example, fronting of the interrogative is absent in the Udegheic PNQ.

4.5 Nanaic

Brylkin" (1861) very early recorded the following question among the Ussuri Nanai:

- (92) Ussuri Nanai (Brylkin" 1861: 21)¹³
gerbi-si xamaca?
 name-2SG.POSS which

This interrogative (*χamača* 'which (one)') in Sem 1976: 62) is not attested in any other Tungusic PNQ. The question appears to be otherwise unattested for Kili¹⁴ and Ussuri Nanai. But for both languages similar constructions have been recorded.

- (93) Kili (Sunik 1958: 116, 122, shortened)
asi-ni gərbi-ni
 woman-3SG.POSS name-3SG.POSS
 'the name of his wife'

This example from Kili also suggests that a Type A construction might have been used. A PNQ in the third person is attested for Ussuri Nanai.

- (94) Ussuri Nanai (Sem 1976: 38)
s'i am'i-s'i gərb'i-n'i χar χala-n'i χar?
 2SG father-2SG.POSS name-3SG.POSS what clan-3SG.POSS what
 'What's your father's name and what's his surname?'

A similar case, but with a personal interrogative borrowed from Northern Tungusic can be found in Kilen.

¹³The Russian translation was *kak" nazывaetsja?* 'How is (it) called?'

¹⁴For convenience, Kili and Kilen are discussed in this subsection, but they exhibit many features from other Tungusic languages.

- (95) Kilen (Dong 2016: 49, slightly modified)¹⁵
xi hale ni, gerbi ni?
2SG clan who name who

According to Schmidt (1928b: 241), northern Nanai (Samar) has similar questions without a possessive marker, but in the reverse order, perhaps based on Manchu influence. The questions about the clan name in all three languages probably represent cultural influence from Manchu and seem to contain the loanword *hala* ‘clan’.¹⁶

- (96) Samar (Schmidt 1928b: 241)
a. *xai gōrbu?*
what name
b. *xai xala?*
what clan

The personal name question in Ussuri Nanai might have been **s’i gərb’i-s’i χar?* (constructed) as in the following Nanai example. In Nanai, however, both *xaj* ‘what’ and *uj* ‘who’ can be employed (Ussuri Nanai *ui*):

- (97) Nanai (Avrorin 1959: 274)
si gərbu-si xaj/uj?
2SG name-2SG.POSS what/who

The latter example has an exact equivalence in Ulcha.

- (98) Ulcha (Schmidt 1923b: 235)
si gölbu-si uji?
2SG name-2SG.POSS who

Nanai has several different possibilities of expressing the question. Apart from the construction above, there is one influenced by Russian making use of a MANNER interrogative.

- (99) Nanai (Ko & Yurn 2011: 151)
swə gərbu-su xo:ni bi?
2PL name-2PL.POSS how COP
‘What is your (SG.POL) name?’

¹⁵<x> stands for [ç].

¹⁶Ewenic languages of Manchuria also have similar expressions, e.g. Oroqen *shi ikun kal?* ‘What is your surname?’ (Chaoke 2014a: 9).

An almost identical example with a copula is found in Ulcha.

- (100) Ulcha (Angina 1993: 3)
si.n gəɮbu-si xon bi-ni?
 2SG.OBL.GEN name-2SG.POSS how COP-3SG

In answers, Nanai has more or less the same construction as in Ewenic and Udegheic with the personal name following the word for ‘name’:

- (101) Nanai (Ko & Yurn 2011: 151)
mi gəɮbu-i tanja.
 1SG name-1SG.POSS PN
 ‘My name is Tanja.’

Uilta is special among Nanaic languages in showing a regular content question marker that is unattested in the rest of Tungusic and might be a Nivkh borrowing (Hözl 2018a: 39, 302–305).

- (102) Uilta (Nakanome 1928: 52; Ikegami 1997: 67)
 a. *sî.n-i geribu-si hai=ga?*
 2SG.OBL-GEN name-2SG.POSS what=CQ
 b. *si.n-i gəɮbu-si xai=gaa?*
 2SG.OBL-GEN name-2SG.POSS what=CQ

In another recording, an example from Uilta uses a personal interrogative. This suggests that the same synchronic variation as in Nanai might be present. The genitive is obligatory in the southern dialect but absent in the northern (Patrik Czerwinski, p.c. 2020).

- (103) Uilta (Ozolinja 2001: 72)
si gəɮbu-si ŋui=ɣə?
 2SG name-2SG.POSS who=CQ

But all three examples share the special question marker =KA(A) that is only attested in Uilta. This question marker is also found in the following example that contains the interrogative *xooni* ‘how’ (cognate of Solon *oni*, Negidal *o:n*, Udihe *ono*, Oroch *ōn’i*, Nanai *xo:ni*, and Ulcha *xon* above).

- (104) Uilta (Patrik Czerwinski, p.c. 2019)
xooni=ka naa gəɮbu-ni?
 how=CQ INTERJ name-3SG.POSS
 ‘But what’s its name?’

As in Negidal, the sentence-initial position of the interrogative is probably based on Russian.

In Kilen, another special case in Nanaic, one example has been recorded that differs in its interrogative from all the other Tungusic languages. Semantically, however, *yanemi* is a MANNER interrogative and might have been directly or indirectly influenced by Russian. The stem *ya-* ‘what, which’ is cognate with Oroqen *i(-kon)*, Udihe *j’ə(-u)* etc. The combination of the dummy noun with the speech act verb also suggests some Chinese influence.

- (105) Kilen (Dong 2016: 37)¹⁷
xn gerbi-xi ya-ne-mi hudarewye?
2SG.OBL.GEN name-2SG.POSS what-V-CVB.IPFV call

Another Kilen example has an equivalent in Oroqen (§4.3). In fact, not only the dummy noun *gerbi*, but also the interrogative *yama* is from Ewenic.

- (106) Kilen (Chaoke 2014b: 8)
shi yama gerbi-shi?
2SG which name-2SG.POSS

Nanaic, like Ewenic and Udegheic, has a tendency for Type A.1. Isolated Type A.2 constructions in Samar and Kilen are most likely based on Jurchenic or Ewenic influence. Similar to Ewenic, the genitive is only occasionally attested in the PNQ. Fronting is almost entirely absent and based on the Russian pattern.

4.6 Jurchenic

Although the person is not marked on the head noun, the personal pronoun can also be absent in Jurchenic languages. According to one source, Manchu can make use of a personal interrogative *we* ‘who’.

- (107) Manchu (Avrorin 2000: 113)
si.n-i gebu ?we?
2SG.OBL-GEN name who

However, this appears to be a mistake, perhaps based on the author’s knowledge of Nanai, as all other sources invariably give the interrogative *ai* ‘what’ instead. This interrogative is cognate with the Nanaic form encountered above, e.g. Uilta *xai*. In Sibe, an optional question marker can attach at the end of the PNQ.

¹⁷*xn* with initial [ç-] goes back to *si.n-i*.

- (108) Sibe (Sameng et al. 2010: 447)¹⁸
xi.n-ĩ gev ai=ye?
 2SG.OBL-GEN name what=Q

Apart from the universal use of this interrogative, Manchu dialects seemingly show the same variation as the Ewenic languages. The interrogative can precede or follow the noun, the personal pronoun can be absent, and it can take a genitive if the interrogative is postposed. But Jurchenic has a tendency for preposed interrogatives.

- (109) Yibuqi Manchu (Zhao 1989: 127)
εi kowə?
 what name
- (110) Aihui Manchu (Wang 2005: 208)
ci ε gəvo?
 2SG what name
- (111) Sanjiazi Manchu (Enhebatu 1995: 39)
ci ai gu:bu?
 2SG what name

Furthermore, these are Type A.2 constructions in which the interrogative stands attributively to the dummy noun. There is no fronting as in Ewenic.

Manchu in Yanbian close to the North Korean border is only preserved in some isolated words and expressions among which there is the following:¹⁹

- (112) Yanbian Manchu (Zhao 2000: 19)
ai hala (keci)?
 what surname ?
 'What's your surname (clan name)?'

While the same expression *ai hala* is also attested in classical Manchu (e.g., Hauer 2007: 217), the *Qingwen Qimeng*, one of the most influential descriptions of Manchu, also contains the following example with reversed word order:

¹⁸In this example, <x> also stands for [ç].

¹⁹The meaning of *keci* is not clear. It could theoretically correspond to Manchu *se-ci* 'say-CVB.COND', but this is problematic on phonological grounds. It could also correspond to Manchu *o-ci* 'become-CVB.COND', which can be a topic marker. Alchuka is known to have an occasional initial *k-* in this word, i.e. (*k*)*ɔ-* (Mu 1986). A connection to Mongolian *g(e)-* 'to say' is unlikely.

- (113) Manchu (Wuge & Cheng 1730: vol. 2; Wylie 1855: 82)
hala ai?
surname what

According to the same source, questions about personal names have the same structure with the interrogative following the noun.

- (114) Manchu (Wuge & Cheng 1730: vol. 2; Wylie 1855: 82)
gebu ai?
name what

According to Veronika Zikmundová (p.c., 2019), this postposed position of the interrogative is impossible in spoken Sibe. As seen above, it is also not very common in other Manchu dialects.

One special example that contains two copies of the word for ‘name’ (written Manchu *gebu*) is attested for Sanjiazi Manchu.

- (115) Sanjiazi Manchu (Enhebatu 1995: 39)
cin gur:bu [ai gur:bu]?
2SG.OBL.GEN name what name

In a similar example from Sibe that is strongly influenced by the written language, the noun *nalma* ‘person’ (written Manchu *niyalma*) can occur twice. In this case, ‘what name’ seems to function as an attribute to ‘person’.

- (116) Sibe (Kałużyński 1977: 23)
ere nalma [ai gebu nalma]?
this person what name person
‘What is this person’s name?’

The sentence thus literally means ‘A what-named person is this person?’

A major difference of Jurchenic with respect to most other Tungusic languages is the widespread use of questions of Type B. An occasional affricatization of *s* (Manchu *se-* ‘to say’) seen in the following Sibe example is also attested in other Jurchenic varieties (see also Chaoke 2014e: 8).

- (117) Sibe (Chaoke 2006: 206)
si.n-i gəvə-v ai dzi-m?
2SG.OBL-GEN name-ACC what say-IPFV

In the following parallel from written Manchu the optional accusative has been added.

- (118) Manchu (He 2009: 21)
si.n-i gebu(-be) ai se-mpi?
 2SG.OBL-GEN name(-ACC) what say-IPFV

Vovin (2006: 259) argues that Manchu *se-* is a Koreanic loanword. Admittedly, *se-* is unattested outside of Jurchenic and has all the hallmarks of being a borrowing. But Manchu *se-* has almost exactly the same range of functions as Mongolian *g(e)-* 'to say' (Janhunen 2012a: 283–285). On phonological grounds it cannot be a direct borrowing from Mongolian, but the underlying construction in the PNQ is almost identical to the one in Jurchenic. Consider the following answer to a PNQ.

- (119) Mongolian (Janhunen 2012a: 283)
mi.n-ii ner-iig delger+maa ge-deg.
 1SG.OBL-GEN name-ACC PN say-PTCP.HAB
 'My name is Delgerma.'

This parallel with the same word order and the same functional elements suggests that the Jurchenic PNQ has been calqued from Mongolian, but the similarities of the verbs go beyond this construction.

In both languages, this intransitive (+ NAME) speech act verb here has a lexical function but is otherwise frequently used in grammatical functions, for example as a quotative. Depending on how the quotative is embedded into the sentence, it can have different forms that have parallels in both languages. For example, Mongolian *ge-deg* 'say-PTCP.HAB', functionally corresponds to Manchu *se-re* 'say-PTCP.IPFV' and can function as an attribute to a following noun or can take case markers. Mongolian *g-e.j* 'say-CVB.IPFV' functionally corresponds to Manchu *se-me* 'say-CVB.IPFV' and is used adverbially (e.g., Janhunen 2012a: 283). While these parallels cannot rule out a potential Koreanic origin of the Jurchenic verb, they nevertheless illustrate a much more intimate connection with Mongolic.

For instance, *se-* does not have the function of a speech act verb, but that of a quotative in the following example that contains the main verb *hūla-* 'to call'.

- (120) Manchu (Schluessel 2014)
[si.n-i gebu-be ai] se.me hūla-mpi?
 2SG.OBL-GEN name-ACC what QUOT call-IPFV

In the following construction, the same verb is used, but without quotative.

- (121) Sanjiazi Manchu (Kim et al. 2008: 161)
si aj gəwu xola-m?
 2SG what name call-IPFV

In the former sentence, the entire part *sini gebu-be ai* is embedded by means of the quotative *se.me*. In the latter example, the question is not embedded. This example is most likely based on the Chinese construction (e.g., 15b) but it also resembles the Solon and Oroqen answers in §4.3.

While the PNQ is unknown in Bala, the words *ci* ‘you (SG)’, *cin* ‘your (SG)’, *garbi* ‘name’, and perhaps *a(i)*- ‘what’ are all attested (Mu 1987: 14, 25, 31). As seen above, the word *garbi* is of Northern Tungusic origin and must have been transmitted through a form of southern Nanai, such as Kilen.

The sentence is not attested in Alchuka and Lalin/Jing Manchu either. However, a similar construction in the third person has the following form:

(122) Alchuka (Mu 1986: 14)
tʼə.rə-i ?əwə intʼuki.
that-GEN name PN
‘His name is Yentugi.’

(123) Lalin Manchu (Aixinjueluo 1987: 14)
te.re-i gebu yintuhi.
that-GEN name PN
‘His name is Yentugi.’

As seen before, the dummy noun was also recorded as *gəbu* for Alchuka. The cognate of written Manchu *ai* ‘what’ has the form (*k*)*ai* or *ei* in Alchuka and *ai* in Lalin/Jing Manchu. Written Manchu *si* ‘you (SG)’ and *sin-i* ‘your (SG)’ correspond to Alchuka *ci/cin-i* and Lalin/Jing Manchu *si/sin-i*. Written Manchu *se-* ‘to say’ has the form *tsʼə-* in Alchuka and *se-* in Lalin/Jing Manchu (Mu 1986; Aixinjueluo 1987).

The earliest recordings of Tungusic are in Jurchen, but to the best of my knowledge the sentence is not attested in these materials either. In Jurchen B, the second person pronoun is attested as **ši* 失, the genitive as **-i* 亦, and the word ‘name’ as **gebu* 革不 (Kane 1989: 270, 272, 356). In Jurchen A, the second person pronoun apparently is not attested, but the equivalences of Manchu *min-i* ‘my’ and *gebu* ‘name’ have the forms **min-i* 密你 and **gebu* 革卜, respectively (Kiyose 1977: 138, 140, 145). It is likely that a comparable range of different constructions as in modern varieties of Manchu might have been present in these languages.

Jurchenic has several examples of all three types of constructions, Type A.1, Type A.2, and Type B. As seen above, Tungusic has otherwise few cases of A.2 and even fewer of Type B. Jurchenic is also the only subbranch of Tungusic that does not use the personal interrogative in the PNQ. The speech act verb *se-* found in Type B constructions is also unattested in other Tungusic languages. Jurchenic

lost head-marked possession and has extended the scope of the genitive to elements other than the speech act participants. All of these features can best be explained by an unusually strong impact from other languages, such as Khitano-Mongolic and perhaps Koreanic (e.g., Vovin 2006), rather than with an early branching of Jurchenic (e.g., Kazama 2003). As has been shown, the Jurchenic Type B construction is clearly a calque from Mongolian.

5 Discussion

5.1 The (re)construction in Proto-Tungusic

A personal name question must have already existed in Proto-Tungusic. The only element that all Tungusic languages without exception have in common in the PNQ is a cognate of the word **gärbü* 'name'. The second person pronoun **si*, which also functions as a possessive marker *-*si* in languages outside of Jurchenic, can be absent in some constructions, but is also attested in all Tungusic languages. The genitive form can be reconstructed as **si.n-i*.

The interrogative is the element of the question that exhibits the most variation. However, apart from Jurchenic, all three other subbranches of Tungusic have at least some examples with a cognate of the interrogative **ŋüi* 'who'. No other interrogative has such a wide distribution in the PNQs of Tungusic. Instances of **Kooni* 'how' are also found in Ewenic, Udegheic, and Nanaic, but this widespread usage can be more plausibly explained with Russian influence all over the northern half of the Tungusic-speaking areas. The use of Tungusic **Kai*²⁰ in both Nanaic (e.g., Uilta *xai*) and Jurchenic (e.g., Manchu *ai*) could indicate that this is a Southern Tungusic innovation, although it is much more pervasive in Jurchenic than in Nanaic and likely due to language contact. Other interrogatives, such as **ja-* 'which', can only be found in very few languages (e.g., Oroqen *i(-kon)*, Udihe *j'e(-u)*).

The use of 'who' in the North and of 'what' in the South is part of a general areal division between languages around Siberia and Mongolia on the one hand and the surrounding languages (e.g., parts of Europe, China, Japan) on the other (e.g., Idiatov 2007; Gil 2018). Proto-Tungusic most likely was part of an area with 'who' and due to contact with Chinese and other languages changed its typological profile in the South. The increasing use of 'how' in the North is based on the Russian construction that represents a pattern found in many European languages.

²⁰Given the uncertainty of the initial, the abstract label **K-* is used in this reconstruction (e.g., Hölzl to appear).

Table 8: Overview of the interrogatives used in the Tungusic PNQs, including dialects and historical data mentioned in the discussion

Language	*ŋüi 'who'	*Kai 'what'	*ja- 'which'	*Kooni 'how'	other
Arman	?	?	?	?	?
Even	+	–	–	–	–
Evenki	+	–	+	–	–
Oroqen	+	–	+	–	which (one)
Solon	–	–	–	+	which (one)
Negidal	?	–	–	+	which (one)
Oroch	+	–	?+	+	how
Udihe	–	–	+	+	–
Kilen	+	?	?	?	how, which (one)
Kili	?	?	?	?	?
Nanai	+	+	–	+	–
Samar	?	+	?	?	?
Ulcha	+	–	–	+	–
Uilta	+	+	–	+	–
U. Nanai	–	+	–	–	which (one)
Alchuka	?	?	?	?	?
Bala	?	?	?	?	?
Jurchen A	?	?	?	?	?
Jurchen B	?	?	?	?	?
Kyakala	?	?	?	?	?
sp. Manchu	–	+	–	–	–
sp. Sibe	–	+	–	–	–
w. Manchu	?–	+	–	–	–

The reconstruction of the Proto-Tungusic PNQ depends on the internal classification of Tungusic. If Jurchenic is considered the oldest branch of the language family (e.g., Kazama 2003), the presence of a second person possessive marker could well be a later innovation in the non-Jurchenic branch. But Jurchenic preserves some traces of the personal markers that must have been present earlier. For instance, Doerfer (1978: 7) observed that ordinal numerals in some Tungusic languages are ultimately derived from what appears to be a third person plural possessive marker (Table 9). The possessive form is preserved, for example, in Udihe, e.g. *neŋu-ti* 'their younger sibling' (Nikolaeva & Tolskaya 2001: 107). In Udihe, a case marker can occasionally precede the ordinal marker, which might be a relic of its origin as a possessive marker, e.g. *nada* 'seven', *nadä-ma-ti* 'seventh (ACC)' (Nikolaeva & Tolskaya 2001: 424). The syllable **ti* that is still recorded as such in Alchuka regularly changed to *ci* in Manchu (e.g., *nadan*, *nada-ci*).

Table 9: Ordinal markers in Alchuka (Mu 1986), and Manchu, Kilen (Zhang, Yanchang, Zhang Xi, et al. 1989), and Udihe (Nikolaeva & Tolskaya 2001)

Language	three	third	-3PL.POSS
Alchuka	ila(-n)	ila-t'i	-
Manchu	ila-n	ila-ci	-
Kilen	ila-n	ila-n-tin	-ti
Udihe	ila	ile-n-ti	-ti

This strongly speaks in favor of head-marking (e.g., head-marked possession) being present in Proto-Tungusic.

Given the presence of Type A constructions throughout the entire language family, Proto-Tungusic must have been of the same type (Table 10). Type B is restricted to few examples, most of which can be found in Jurchenic. For instance, as seen before, the typical Jurchenic question containing a speech act verb (Manchu *se-*) is clearly calqued from the Mongolian pattern (§4.6). Apart from the use of a personal interrogative, the construction is almost a perfect match.

- (124) Sibe (Zikmundová 2013: 138)²¹
šin *gəvə-f* *ai* *zə-mie?*
 2SG.OBL.GEN name-ACC what say-IPFV

²¹Sibe *šin* goes back to *si.n-i* '2SG.OBL-GEN'. Jurchenic also has sentence-final content question marking that is, however, not usually attested in the PNQs.

- (125) Mongolian (elicited in May 2019)
či.n-ii ner-iig xen ge-deg=ve?
2SG.OBL-GEN name-ACC who say-PTCP.HAB=CQ

Content question marking as in this Mongolian example is a feature absent from most Tungusic languages (Hölzl 2018a: 286–312). In those languages that have this feature, such as Jurchenic languages, Khamnigan Evenki, or Uilta, this is clearly an innovation. Consequently, Proto-Tungusic most likely did not have content question marking either. All Type B constructions can plausibly be explained with language contact.

In conclusion, the most likely reconstruction for the proto-Tungusic personal name question is perhaps the following Type A, more specifically Type A.1, construction with an optional pronoun and an optional genitive.²²

- (126) Proto-Tungusic
**(si(n-i)) gärbü-si ŋüi?*
2SG.OBL-GEN name-2SG.POSS who

All four subbranches of Tungusic have direct descendants of this construction, such as the following from Even (with optional pronoun *hi* ‘2SG’, *hin* ‘2SG.OBL(.GEN)’) and Manchu.

- (127) Even (Doerfer et al. 1980: 304, modified transcription)
gerbe-s ŋi?
name-2SG.POSS who
- (128) Manchu (Haenisch 1961: 73)
si.n-i gebu ai?
2SG.OBL-GEN name what

Some languages, such as Manchu, have introduced a new interrogative into the construction, replacing the original **ŋüi*. Jurchenic has generally lost the possessive marker **-si*, at the same time generalizing the genitive.

One can suspect that the Tungusic construction above was based on a more schematic construction that has the following form, X being a pronoun, Y a possessive ending, and Z a proper name or the interrogative **ŋüi*: **(X(n-i)) gärbü-Y Z*. The genitive might have been restricted to first and second person pronouns. Only Jurchenic has third person pronouns that can take a genitive (singular *i.n-i*, plural *ce.n-i* in Manchu) and it remains an open question whether this represents

²²Very similar constructions to this one reconstructed to Tungusic can be found in some surrounding languages. These cannot be addressed here for reasons of space (see, e.g., 39).

4 'What's your name?' in Tungusic and beyond

Table 10: The type of PNQs in Tungusic languages

Language	Type A (copula)	Type B (speech act verb)
Arman	?	?
Even	+	-
Evenki	+	-
Oroqen	+	-
Solon	+	-
Negidal	+	-
Oroch	+	-
Udihe	+	+
Kilen	+	+
Kili	?+	?
Nanai	+	-
Samar	+	-
Ulcha	+	-
Uilta	+	-
U. Nanai	+	-
Alchuka	?+	?
Bala	?	?
Jurchen A	?	?
Jurchen B	?	?
Kyakala	?	?
sp. Manchu	+	+
sp. Sibe	+	+
w. Manchu	+	+

a Proto-Tungusic pattern that was replaced everywhere else or is also an innovation in Jurchenic (e.g., Zikmundová 2022 [this volume]). The use of the genitive on elements other than the pronouns is probably a Jurchenic innovation that later spread to a few other Tungusic languages.

- (129) Manchu (Aixinjueluo 1987: 14)
te.re-i gebu yentugi.
that-GEN name PN

Another instantiation of the schematic construction can be observed in the following answer from Even.

- (130) Even (Doerfer et al. 1980: 304)
m̄.n gerbe-w Anna.
1SG.OBL(.GEN) name-1SG.POSS PN

The preposed interrogative as in the following Aoluguya Evenki example (Type A.1) appears to be restricted to Ewenic (found in Even, Evenki, Oroqen, and Solon in §4.3).

- (131) Aoluguya Evenki (Chaoke & Sirenbatu 2016: 1)
fi [ni] gərbi-tfi
2SG who name-POSS

This also illustrates another innovation in parts of Ewenic, which is the use of the comitative or possessive suffix (*gərbi-tfi* ‘with/having a name’), replacing the second person possessive marker in the PNQ (*gərbi-fi* ‘your name’, Chaoke & Sirenbatu 2016: 5).

Seemingly similar expressions in Jurchenic (see 17 and §4.6) cannot be based on the same construction because the interrogative (Manchu *ai*) functions as an attribute to the dummy noun (Manchu *gebu*) (Type A.2).

- (132) Manchu (Sanjiazi; Chaoke 2014d: 8)
shi [ayi gewe]?
2SG what name

The personal interrogative in Evenki cannot, however, stand attributively to a noun (Nedjalkov 1997: 215). The interrogative, therefore, must be interpreted as an argument of its own that stands in some sort of focus position that is specific to Ewenic. In Evenki, interrogatives often are sentence-initial, but there is another construction: “Much more rarely, they appear in the second position after the subject or the object of the question in cases when these components are stressed.” (Nedjalkov 1997: 7f.) This must be considered an early innovation of Ewenic languages.

5.2 Conclusion: Construction and frame

This study has investigated a potentially universal property of human language, the personal name question (PNQ, ‘What’s your name?’). While the focus was on Tungusic languages, several typological dimensions of variation were discussed from a global perspective. Cross-linguistically, there are two main types of PNQs that contain an equational copula (Type A) and a speech act verb (Type B), respectively. Tungusic languages show a tendency for Type A, although the Jurchenic subbranch due to language contact also has many instances of Type B. On the basis of the PNQ in the individual Tungusic languages, the PNQ in Proto-Tungusic has been reconstructed as an instance of Type A. This reconstruction lacks a copula but contains a personal interrogative **ŋüi* ‘who’, an optional personal pronoun **si* ‘you (SG)’ (oblique **si.n-*) with optional genitive **-i*, and a dummy noun **gärbü* ‘name’ that functions as a host for head-marked possessive affixes. The basis for the apparent split between head-marking on the one hand and double marking on the other remains unclear for now.

Generally, personal name questions can be said to be semantically based on what has been called the personal name frame (§3) that has several subevents, each with its individual roles. The Tungusic Type A construction highlights or profiles the subevents of **HAVING A NAME** and **ACQUAINTING**. The whole expression is the result of a complex interaction of the individual frames and constructions (Figure 1).

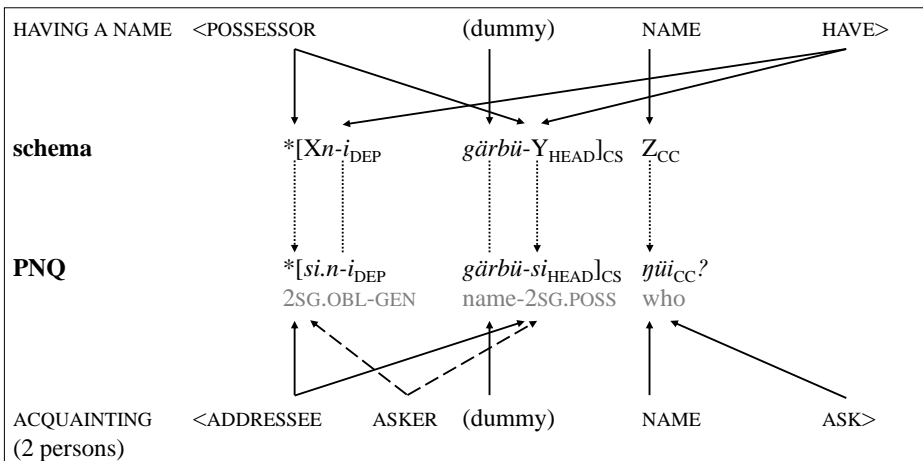


Figure 1: The interaction of frames and constructions in the Proto-Tungusic PNQ (figure created by the author)

In the schematic construction, X is an open slot for a pronoun, Y for a possessive ending corresponding to X, and Z for a proper name or the interrogative **ηüi*. CS and CC stand for copula subject and copula complement, respectively (Dixon 2010). The dummy noun **gärbü* ‘name’ is head and the personal pronoun **si* ‘you (SG)’ is the dependent. Dotted lines indicate that a given element is identical in the schematic and in the specific construction, e.g. the genitive remains **-i*. Dotted arrows show the filling of an open slot with a certain element, e.g. of X with the pronoun **si* ‘you (SG)’. Arrows from the frames to the constructions indicate the place of realization of roles and relations. In some cases, multiple realization is possible, e.g. of the possessor as both the personal pronoun and possessive affix. Finally, dashed arrows are used for roles and relations that are only indirectly coded in the construction. In this example, the role of the person asking is only indirectly represented by the second person elements. The interrogative force of the question, here tentatively indicated with the semantic relation ASK, has no overt morphosyntactic expression but is indirectly encoded in the interrogative and perhaps a special intonation contour that is difficult to reconstruct given the scarcity of data from modern languages.

Abbreviations

PNQ stands for *personal name question* and PNF for *personal name frame*. Abbreviations follow the general convention. Special grammatical abbreviations include:

ABST	abstract, non-concrete (cf. Munshi 2006)	GL	general topic (Lock 2011)
ANX	anxiousness (Prins 2017)	L	low tone form (Paperno 2014)
CQ	content question marker	SIM	simple (unmarked) tense (Coler 2014)

Acknowledgements

This paper is dedicated to the memory of Prof. Wolfgang Schulze (1953–2020).

I want to thank Veronika Zikmundová, Patryk Czerwinski, Elena Perekhvalskaya, Natalia Aralova, and Tom Payne for their valuable comments on Sibe, Uilta, Udihe, Negidal, and Panare, respectively.

References

- Aalto, Pentti. 1977. G. J. Ramstedts Onkor-Solonisches Wörterverzeichnis. *Rocznik Orientalistyczny* 39. 55–84.
- Abbi, Anvita. 2013. *A grammar of the Great Andamanese language: An ethnolinguistic study* (Brill's Studies in South and Southwest Asian Languages). Leiden: Brill.
- Aixinjueluo, Yingsheng. 1987. Tantan manyu de jingyu. *Manyu yanjiu* 1. 2–15, 73.
- Alonso de la Fuente, José Andrés. 2011. Venjukov's [1858] 1862/1868 Nanai materials. *Journal de la Société Finno-Ougrienne* 93. 11–35.
- Alonso de la Fuente, José Andrés. 2012/2014. Some thoughts on Onomastica Manjurica: Strange or amusing names in Manchu. *Faventia* 34/36. 293–301.
- An, Jun. 1986. *Hezheyu jianzhi*. Beijing: Minzu chubanshe.
- Angina, S. V. 1993. *Russko-ul'čskij razgovornik*. Bogorodskoe.
- Anthony, David W. 2007. *The horse, the wheel and language. How Bronze-Age riders from the Eurasian Steppe shaped the modern world*. Princeton: Princeton University Press.
- Aoi, Hayato. 2015. Tarama Miyako grammar. In Heinrich Patrick, Shinsho Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages. History, structure, and use* (Handbooks of Japanese Language and Linguistics 11), 405–421. Berlin: de Gruyter Mouton.
- Arsen'ev, Vladimir K. 1921. *Po ussurijskomu kraju*. Vladivostok: Tipografija "Èxo". <https://ru.wikisource.org/>.
- Arsen'ev, Vladimir K. 2016 [1921]. *Across the Ussuri Kray: Travels in the Sikhotealin mountains*. Indiana: Indiana University Press.
- Avrorin, Valentin A. 1959. *Grammatika nanajskogo jazyka*, vol. 1. Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Avrorin, Valentin A. 2000. *Grammatika man'čžurskogo pis'mennogo jazyka*. Saint Petersburg: Nauka.
- Avrorin, Valentin A. & Elena P. Lebedeva. 1978. *Oročskie teksty i slovar'*. Leningrad: Nauka.
- Balakrishnan, R. 1999. *Badaga: A Dravidian language*. Annamalai: Annamalai University.
- Barðdal, Jóhanna. 2013. Construction-based historical-comparative reconstruction. In Thomas Hoffmann & Graeme Trousdale (eds.), *Oxford handbook of Construction Grammar*, 438–457. Oxford: Oxford University Press.
- Benzing, Johannes. 1955. *Lamutische Grammatik* (Akademie der Wissenschaften der Literatur. Veröffentlichungen der orientalischen Kommission 6). Wiesbaden: Steiner.

- Benzing, Johannes. 1956. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik* (Akademie der Wissenschaften und der Literatur. Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 11). Mainz: Akademie der Wissenschaften und der Literatur.
- Black, Jeremy A. 2007. Sumerian. In J. Nicholas Postgate (ed.), *Languages of Iraq, ancient and modern*, 4–30. Cambridge: British School of Archaeology in Iraq.
- Blust, Robert A. 2013. *The Austronesian languages*. 2nd edn. Canberra: Research School of Pacific & Asian Studies, Australian National University.
- Boldyrev, Boris V. 2000. *Èvenkijsko-russkij slovar'*. 2 vols. Novosibirsk: Izdatel'stvo so ran filial "Geo".
- Brylkin, A. 1861. Zaměčanija o svojstvax" jazyka xodzenov" i xodzenskij slovar'. In R. Maak" (ed.), *Putešestvie po dolině rěki usuri*, vol. 1, 204–225. Saint Petersburg: V. Bezobrazova.
- Bulatova, Nadezhda & Stefano Cotrozzi. 2004. *Ewenkische Erzählungen aus der Insel Sachalin*. Bonn: Verlag für Kultur und Wissenschaft.
- Castrén, M. Alexander. 1856. *Grundzüge einer tungusischen Sprachlehre nebst kurzem Wörterverzeichnis*. St Petersburg: Buchdruckerei der kaiserlichen Akademie der Wissenschaften.
- Chaoke, D. O. 2006. *Xiandai xiboyu kouyu yanjiu*. Beijing: Renmin chubanshe.
- Chaoke, D. O. 2007. *Nanmu elunchunyu yanjiu*. Beijing: Minzu chubanshe.
- Chaoke, D. O. 2014a. *Elunchunyu 366 ju huihuaju*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. 2014b. *Hezheyu 366 ju huihuaju*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. 2014c. *Mantonggusi yuzu yuyan cihui bijiao*. Beijing: Zhongguo shehui kexue chubanshe.
- Chaoke, D. O. 2014d. *Manyu 366 ju huihuaju*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. 2014e. *Xiboyu 366 ju huihuaju*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. & Kajia. 2016. *Tonggusi ewenkeyu huihua*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O., Kajia, Aliya & Tami'er. 2014. *Ewenkeyu 366 ju huihuaju*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. & Kalina. 2017. *Arong ewenkeyu*. Beijing: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. & Sirenbatu. 2016. *Aoluguya ewenkeyu huihua*. Beijing: Shehui kexue wenxian chubanshe.

- Cincius, Vera I. 1975/77. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov. materialy k etimologičeskomu slovarju*. 2 vols. Leningrad: Nauka.
- Coler, Matt. 2014. *A grammar of Muylaq' Aymara: Aymara as spoken in Southern Peru*. Leiden: Brill.
- Coupe, Alexander Robertson. 2007. *A grammar of Mongsen Ao* (Mouton Grammar Library 39). Berlin: Mouton de Gruyter.
- Croft, William. 2003. *Typology and universals*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Dai, Guangyu. 2012. *Sanjiazi manyu yuyin yanjiu*. Beijing: Beijing daxue chubanshe.
- Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27(3). 635–655.
- Dixon, Robert M. W. 2010. *Basic linguistic theory, vol. 1*. Oxford: Oxford University Press.
- Doerfer, Gerhard. 1978. Classification problems of Tungus. In Gerhard Doerfer & Michael Weiers (eds.), *Tungusica, vol.1: Beiträge zur nordasiatischen Kulturgeschichte*, 1–26. Wiesbaden: Harrassowitz.
- Doerfer, Gerhard, Wolfram Hesse & Hartwig Scheinhardt. 1980. *Lamutisches Wörterbuch*. Wiesbaden: Harrassowitz.
- Doerfer, Gerhard & Michael Knüppel. 2004. *Etymologisch-ethnologisches Wörterbuch tungusischer Dialekte (vornehmlich der Mandschurei)*. Hildesheim: Olms. 932.
- Doerfer, Gerhard & Michael Knüppel. 2013. *Armanisches Wörterbuch*. Nordhausen: Verlag Traugott Bautz.
- Dong, Xingye. 2016. *Hezheyu*. Harbin: Heilongjiang renmin chubanshe.
- Dryer, Matthew S. 2013. Position of interrogative phrases in content questions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/93>.
- Elias, David. 2014. *The Tigre language of Ginda', Eritrea* (Studies in Semitic Languages and Linguistics). Leiden: Brill.
- Enhebatu, Merden. 1995. *Manyu kouyu yanjiu*. Hohhot: Neimonggu daxue chubanshe.
- Everett, D. L. & B. Kern. 2007. *Wari': The Pacaas Novos language of western Brazil* (Descriptive Grammars). London: Routledge.
- Faehndrich, Burgel R. M. 2007. *Sketch grammar of the Karlong variety of Mongghul, and dialectal survey of Mongghul*. Honolulu: University of Hawai'i. (Doctoral dissertation).

- Fillmore, Charles. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222–254.
- Georg, Stefan. 2004. Unreclassifying Tungusic. In Carsten Naehrer (ed.), *Proceedings of the First International Conference on Manchu-Tungus studies*, vol. 2 (Tunguso Sibirica 9), 45–57. Wiesbaden: Harrassowitz.
- Gil, David. 2018. *Who is your name?* Abstract for The Twenty-Second International Symposium on Malay/Indonesian linguistics (ISMIL 22), 2018.05.11-12, Los Angeles.
- Grube, Wilhelm. 1900. *Goldisch-deutsches Wörterverzeichnis mit vergleichender Berücksichtigung der übrigen tungusischen Dialekte* (Reisen und Forschungen im Amur-Lande 3(2.1)). Sankt Petersburg: Bei den Commissionaren der Kaiserlichen Akademie der Wissenschaften.
- Haenisch, Erich. 1961. *Mandschu-Grammatik mit Lesestücken und 23 Texttafeln*. Leipzig: VEB Verlag Enzyklopädie.
- Han, Youfeng & Shuxian Meng. 1993. *Elunchun hanyu duizhao duben*. Beijing: Zhongyang minzuyuan chubanshe.
- Hasibate'er. 2016. *Aoluguya fangyan yanjiu*. Beijing: Minzu chubanshe.
- Hauer, Erich. 2007. *Handwörterbuch der Mandschusprache*. 2nd edn. Wiesbaden: Harrassowitz.
- He, Rongwei. 2009. *Manyu 365 ju*. Shenyang: Liaoning minzu chubanshe.
- Hilpert, Martin & Jan-Ola Östman. 2014. Reflections on constructions across grammars. *Constructions and Frames* 6(2). 137–142.
- Hölzl, Andreas. 2014. "What's your name?" in Tungusic. Paper presented at The 8th International Conference on Construction Grammar, 2014.09.03-06, Osnabrück.
- Hölzl, Andreas. 2018a. *A typology of questions in Northeast Asia and beyond: An ecological perspective* (Studies in Diversity Linguistics 20). Berlin: Language Science Press.
- Hölzl, Andreas. 2018b. Constructionalization areas: The case of negation in Manchu. In Evie Coussé, Peter Andersson & Joel Olofsson (eds.), *Grammaticalization meets construction grammar* (Constructional Approaches to Language 21). Amsterdam: Benjamins.
- Hölzl, Andreas. 2018c. Udi, Udihe, and the language(s) of the Kyakala. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 15. 111–146.
- Hölzl, Andreas. to appear. The complexification of Tungusic interrogative systems. In Dankmar W. Enke, Larry M. Hyman, Johanna Nichols, Guido Seiler & Thilo Weber (eds.), *Language change for the worse* (Studies in Diversity Linguistics). Berlin: Language Science Press.

- Hölzl, Andreas & Yadi Hölzl. 2019. A wedding ceremony of the Kyakala in China: Language and ritual. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 16. 87–144.
- Hough, Carole (ed.). 2016. *The Oxford handbook of names and naming*. Oxford: Oxford University Press.
- Huang, Lillian M. 1996. Interrogative constructions in Mayrinax Atayal. *Bulletin of National Taiwan Normal University* 41. 263–296.
- Idiatov, Dmitry. 2007. *A typology of non-selective interrogative pronominals*. Antwerp: Universiteit Antwerpen. (Doctoral dissertation).
- Ikegami, Jirō. 1985. The category of person in Tungus: Its representation in the indicative forms of verbs. *Gengō Kenkyū* 88. 86–96.
- Ikegami, Jirō. 1997. *Uirutago jiten*. Sapporo: Hokkaidō University Press.
- Ivanovskiy, A. O. 1982 [1894]. *Mandjurica I. Specimens of the Solon and Dagur languages* (Debter, Deb-Ther, Debtelin: Materials for Central Asiatic and Altaic Studies 2). Budapest: Akadémiai Kiadó.
- Janhunen, Juha. 1991. *Material on Manchurian Khamnigan Evenki* (Castrenianumin toimitteia 40). Helsinki: University of Helsinki.
- Janhunen, Juha. 2012a. *Mongolian* (Oriental and African Language Library 19). Amsterdam: Benjamins.
- Janhunen, Juha. 2012b. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.
- Jin, Ning. 1993. *Sibe English conversations*. Wiesbaden: Harrassowitz.
- Johnson, Allen. 2003. *Families of the forest: The Matsigenka Indians of the Peruvian Amazon*. Berkeley: University of California Press.
- Kałużyński, Stanisław. 1977. *Die Sprache des mandschurischen Stammes Sibe aus der Gegend von Kuldscha*, vol. 1. Warsaw: Państwowe Wydawnictwo Naukowe.
- Kane, Daniel. 1989. *The Sino-Jurchen vocabulary of the Bureau of Interpreters*. Bloomington: Indiana University Press.
- Kazama, Shinjirō. 2002. *Negidarugo tekisuto to bunpō gaisetsu* (Publications on Tungus Languages and Cultures 19). Kyōto: Nakanishi.
- Kazama, Shinjirō. 2003. *Basic vocabulary (A) of Tungusic languages* (Endangered Languages of the Pacific Rim A2-037). Kyōto: Nakanishi.
- Khabtagaeva, Bayarma. 2022. On some shared and distinguishing features of Nercha and Khamnigan Ewenki dialects. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 149–197. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053367.
- Khasanova, Marina & Alexander M. Pevnov. 2003. *Myths and tales of the Negidals* (ELPR Publication Series A2-024). Osaka: Nakanishi.

- Kim, Juwon. 2011. *A grammar of Ewen*. Seoul: Seoul National University Press.
- Kim, Juwon, Dongho Ko, D. O. Chaoke, Youfeng Han, Lianyu Piao & Boris V. Boldyrev. 2008. *Materials of spoken Manchu*. Seoul: Seoul National University Press.
- King, J. R. P. 1987. An introduction to Soviet Korean. *Language Research* 23(2). 233–271.
- Kiyose, Gisaburo N. 1977. *A study of the Jurchen language and script: Reconstruction and decipherment*. Kyōto: Shōwado.
- Klamer, Marian. 2010. *A grammar of Teiwa* (Mouton Grammar Library 49). Berlin: Mouton de Gruyter.
- Kluge, Angela. 2017. *A grammar of Papuan Malay*. Berlin: Language Science Press.
- Ko, Dongho & Gyudong Yurn. 2011. *A description of Najkhin Nanai* (Altaic Languages Series 7). Seoul: Seoul National University Press.
- Kobayashi, Masato & Bablu Tirkey. 2017. *The Kurux language: Grammar, texts, and lexicon* (Brill's Studies in South and Southwest Asian Languages 8). Leiden: Brill.
- Konstantinova, Olga A. 1964. *Èvenkijskij jazyk: fonetika i morfologija*. Moskau: Nauka.
- Lakoff, George & Mark Johnson. 1999. *Philosophy in the flesh. The embodied mind and its challenges to Western thought*. New York: Basic Books.
- Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Lavrillier, Alexandra. 2006. S'orienter avec les rivières chez les Évenks du Sud-Est sibérien. Un système d'orientation spatial, identitaire et rituel. *Études mongoles et sibériennes, centrasiatiques et tibétaines* 36/37. 95–138.
- Lefebvre, Claire. 2015. *Relabeling in language genesis*. Oxford: Oxford University Press.
- Li, Paul Jen-kuei & Shigeru Tsuchida. 2001. *Pazih dictionary* (Language and Linguistics Monograph Series A2). Taipei: Institute of Linguistics, Academia Sinica.
- Li, Shulan. 1984. Xiboyu dongci chenshushi de qinzhi kouqi he fei qinzhi kouqi. *Minzu yuwen* 6. 26–32.
- Lichtenberk, Frantisek. 1983. *A grammar of Manam* (Oceanic Linguistics Special Publication 18). Honolulu: University of Hawaii Press.
- Lie, Hiu. 1978. Solonisches Material aus dem Huin-gol. In Gerhard Doerfer & Michael Weiers (eds.), *Tungusica, vol. 1: Beiträge zur nordasiatischen Kulturgeschichte*, 126–78. Wiesbaden: Otto Harrassowitz.
- Lock, Arnold. 2011. *Abau grammar*. Ukarumpa: SIL-PNG Academic Publications.

- Lopatin, Ivan A. 1957. Material on the Orochee language. *Micro-Bibliotheca Anthropos* 26. 1–110.
- Maak", R. 1859. *Putešestvie na Amur*". Saint Petersburg: K. Wulf.
- Majewicz, Alfred F. (ed.). 2011. *The collected works of Bronisław Piłsudski, vol 4: Materials for the study of Tungusic languages and folklore*. Berlin: Mouton de Gruyter.
- Makarova, K. I. 1999. *Èvenkijsko-russkij razgovornik: Načinajuščim izučat' èvenkijskij jazyk (na vostočnom dialekte)*. Yakutsk: Bičik.
- Moutu, Andrew. 2013. *Names are thicker than blood: Kinship and ownership amongst the Iatmul*. Oxford: Oxford University Press.
- Mu, Yejun. 1986. Alechuka manyu de shuci yu gezhuci. *Manyu yanjiu* 1986(1). 2–17.
- Mu, Yejun. 1987. Balayu. *Manyu yanjiu* 1987(2). 2–31, 128.
- Mu, Yejun. 1988. Lun balayu de yuyin bianhua. *Manyu yanjiu* 6. 1–26.
- Munshi, Sadaf. 2006. *Jammu and Kashmir Burushaski: Language, language contact and change*. Ann Arbor: UMI. (Doctoral dissertation).
- Mushin, Ilana. 1995. Epistememes in Australian languages. *Australian Journal of Linguistics* 15(1). 1–31.
- Nagaraja, K. S. 2014. *The Nihali language: Grammar, texts, vocabulary*. Mysore: Central Institute of Indian Languages.
- Nakanome, Akira. 1928. *Grammatik der Orokko-Sprache*. Osaka: The Osaka Asiatic Society.
- Nau, Nicole. 1999. Was schlägt der Kasus? Zu Paradigmen und Formengebrauch von Interrogativpronomen. *Sprachtypologie und Universalienforschung* 2. 130–150.
- Nedjalkov, Igor. 1997. *Evenki (Descriptive Grammars)*. London: Routledge.
- Nikolaeva, Irina & Maria Tolskaya. 2001. *A grammar of Udihe* (Mouton Grammar Library 22). Berlin: Mouton de Gruyter.
- Norman, Jerry. 2013. *A comprehensive Manchu-English dictionary*. Cambridge: Harvard University Asia Center.
- Ozolinja, L. V. 2001. *Oroksko-russkij slovar'*. Novosibirsk: Izdatel'stvo So Ran.
- Pallas, Peter Simon. 1786, 1789. *Linguarum totius orbis vocabularia comparativa*, vol. 1. Saint Petersburg: Typis Iohannis Caroli Schnoor.
- Paperno, Denis. 2014. Grammatical sketch of Beng. *Mandenkan* 51. 1–130.
- Payne, Thomas E. & Doris L. Payne. 2013. *A typological grammar of Panare: A Cariban language of Venezuela* (Brill's Studies in the Indigenous Languages of the Americas). Leiden: Brill.

- Perekhval'skaya, Elena. 2022. From consonant to tone: Laryngealized and pharyngealized vowels in Udihe. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 227–262. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053371.
- Prins, Maria Clazina. 2017. *A grammar of rGyalrong, Jiāomùzú (Kyom-kyo) dialects: A web of relations*. Leiden: Brill.
- Sa, Xirong. 1981. *Jianming hanyu elunchunyu duizhao duben*. Beijing: Minzu chunbanshe.
- Sameng, Yierhanzhi, Guo Xiangyang & Xie Wei. 2010. *Xiboyu tonglun*. Ürümqi: Xinjiang renmin chubanshe.
- Sandman, Erika. 2016. *A grammar of Wutun*. Helsinki: University of Helsinki. (Doctoral dissertation).
- Sarvasy, Hannah. 2017. *A grammar of Nungon: a Papuan language of Northeast New Guinea*. Leiden: Brill.
- Schebesta, Pater P. & C. O. Blagden. 1928. Grammatical sketch of the Jahai dialect, spoken by a Negrito tribe of Ulu Perak and Ulu Kelantan, Malay Peninsula. *Bulletin of the School of Oriental Studies, University of London* 4(4). 803–826.
- Schiefner, Anton. 1878. Alexander Czekanowski's tungusisches Wörterverzeichnis. *Bulletin de l'Academie imperiale des sciences de St.-Petersbourg* 24. 89–146.
- Schlüssel, Eric. 2014. *Two encounters on the riverbank*. Manchu Studies Group. <http://www.manchustudiesgroup.org/2014/04/23/two-encounters-on-the-riverbank/> (Accessed 2019-05-06.)
- Schmidt, Peter. 1923a. The language of the Negidals. *Acta Universitatis Latviensis* 5. 3–38.
- Schmidt, Peter. 1923b. The language of the Olchas. *Acta Universitatis Latviensis* 8. 229–288.
- Schmidt, Peter. 1928a. The language of the Oroches. *Acta Universitatis Latviensis* 17. 17–62.
- Schmidt, Peter. 1928b. The language of the Samagirs. *Acta Universitatis Latviensis* 19. 219–49.
- Schulze, Wolfgang. 2007. Communication or memory mismatch: Towards a cognitive typology of questions. In Günter Radden, Klaus-Michael Köpcke, Thomas Berg & Peter Siemund (eds.), *Aspects of meaning construction*, 247–264. Amsterdam: Benjamins.
- Sem, Lidija I. 1976. *Očerki dialektov nanaj'skogo jazyka: Bikinskij (ussurijskij) dialekt*. Leningrad: Nauka.
- Senft, Gunter. 1986. *Kilivila: The language of the Trobriand islanders* (Mouton Grammar Library 3). Berlin: Mouton de Gruyter.

- Sohn, Ho-Min. 1999. *The Korean language* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Song, Jae Jung. 2005. *The Korean language. Structure, use and context*. London: Routledge.
- Sunik, Orest. 1958. *Kur-urmijskij dialekt*. Leningrad: Nauka.
- Tournadre, Nicolas & Randy LaPolla. 2014. Towards a new approach to evidentiality. Issues and directions for research. *Linguistics of the Tibeto-Burman area* 37(2). 240–263.
- Trousdale, Graeme. 2014. On the relationship between grammaticalization and constructionalization. *Folia Linguistica* 48(2). 557–578.
- Tsumagari, Toshirō. 1992. A basic vocabulary of Khamnigan and Oluguya Ewenki in northern Inner Mongolia. *Bulletin of the Institute for the Study of North Eurasian Cultures* 21. 83–103.
- Tsumagari, Toshirō. 2009a. A sketch of Solon grammar. *Journal of the Center for Northern Humanities* 2. 1–21.
- Tsumagari, Toshirō. 2009b. Grammatical outline of Uilta (revised). *Journal of the Graduate School of Letters* 4. 1–21.
- Tsumagari, Toshirō. 2011. An Udihe folktale text: “Solomo and Tausima”. *Hoppō jinbun kenkyū* 4. 75–93.
- Tsumagari, Toshirō et al. 2006. *Udehe-go no tebikiki*. Sapporo: Hokkaidō daigaku bungaku kenkyūka.
- Van Langendonck, Willy & Mark Van de Velde. 2016. Names and grammar. In Carole Hough (ed.), *The Oxford handbook of names and naming*, 17–39. Oxford: Oxford University Press.
- Venukoff, M. 1862. Vokabular der Golde-Sprache. *Petermanns geographische Mittheilungen* 8. 109.
- von Klaproth, Julius. 1817. Mandshuisch. In Johann C. Adelung (ed.), *Mithridates oder allgemeine Sprachkunde mit dem Vater Unser als Sprachprobe in bey nahe fünfhundert Sprachen und Mundarten*, vol. 4, 194–224. Berlin: In der Vossischen Buchhandlung.
- von Goethe, Johann Wolfgang. 2018 [1808]. *Faust*. (Translation.) Minneapolis: Avenue Editions.
- Vovin, Alexander. 2006. Why Manchu and Jurchen look so non-Tungusic? In Alessandra Pozzi, Juha Janhunen & Michael Weiers (eds.), *Tumen jalafun jecen akū. Manchu studies in honour of Giovanni Stary*, 255–266. Wiesbaden: Harrassowitz.
- Wang, Qingfeng. 2005. *Manyu yanjiu*. Beijing: Minzu chubanshe.
- Witsen, Nicolaas. 1705. *Noord en Oost Tartarye*. Amsterdam: François Halma.

- Wuge, Shouping & Mingyuan Cheng. 1730. *Manju nikan hergen-i cing wen ki meng bithe*. *Manhan zi qingwen qimeng*. 4 vols.
- Wurm, Stephen A. & Peter Mühlhäusler. 1985. *Handbook of Tok Pisin (New Guinea Pidgin)*. Stefan A. Wurm & Peter M.ühlhäusler (eds.) (Pacific Linguistics: Series C 70). Canberra: The Australian National University.
- Wylie, Alexander. 1855. *Translation of the Ts'ing wan k'e mung, a Chinese grammar of the Manchu Tartar language; with introductory notes on Manchu literature*. Shanghai: London Mission Press.
- Yu, Wonsoo, Kwon Jae-il, Choi Moon-Jeong, Shin Yong-kwon, Bayarmend Borjigin & Luvsandorj[iin] Bold. 2008. *A study of the Tacheng dialect of the Dagur language* (Altaic Languages Series 2). Seoul: Seoul National University Press.
- Zhang, Yanchang, Li Bing & Zhang Xi. 1989. *The Qroqen language*. Jilin: Jilin daxue chubanshe.
- Zhang, Yanchang, Zhang Xi & Dai Shuyan. 1989. *The Hezhen language*. Jilin: Jilin daxue chubanshe.
- Zhao, Jie. 1989. *Xiandai manyu yanjiu*. Beijing: Minzu chubanshe.
- Zhao, Jie. 2000. Yanbian manyu duoshi. *Minzu yuwen* 3. 18–24.
- Zikmundová, Veronika. 2013. *Spoken Sibe: Morphology of the inflected parts of speech*. Prague: Karolinum.
- Zikmundová, Veronika. 2022. Historical language contact between Sibe and Khorchin. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 295–329. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053375.

Chapter 5

On some shared and distinguishing features of Nercha and Khamnigan Ewenki dialects

Bayarma Khabtagaeva

University of Naples L'Orientale, Department of Asian, African and Mediterranean Studies

The present paper is a brief addition to the author's recent monograph (Khabtagaeva 2017) which deals with Mongolic elements in Ewenki dialects (Barguzin, Nercha, Baunt and North-Baikal) spoken in the territory of Buryatia, Russia. Today Nercha Ewenki is no longer spoken. During the initial years of Soviet rule (ca. 1918–1932) some Nercha speakers crossed the border into Manchuria, China, and today their descendants are speakers of Manchurian Khamnigan Ewenki. The aim of this paper is to find out similarities and differences between the extinct Nercha Ewenki dialect and Manchurian Khamnigan Ewenki.

Keywords: Nercha Ewenki, Khamnigan Ewenki, etymology, Mongolic loanwords

1 Introduction

The present paper is a brief addition to the author's recent monograph (Khabtagaeva 2017) which focuses on Mongolic loanwords in Ewenki dialects spoken in the territory of Buryatia, Russia. The idea to write this paper was motivated by the author's trip in September 2017, to carry out fieldwork among Mongolic and Tungusic people in Manchuria.

The main goal of the published monograph (Khabtagaeva 2017) was to clarify the status of early Mongolic (i.e. non-Buryat) and later Mongolic or Buryat layers in Ewenki, with the main finding being that almost all phonetic characteristics of



Mongolic loanwords in Ewenki dialects coincide with Khamnigan Mongol. This implies that an “early” Mongolic language related to Modern Khamnigan Mongol¹ was spoken in the Transbaikalian territory before the Buryat tribes arrived here, and this language had a considerable effect on Ewenki dialects in the earlier stages of borrowing (Khabtagaeva 2017: 200–201).

The introductory part of Khabtagaeva (2017) provides a brief overview of the Ewenki dialects of Buryatia (Barguzin, Nercha, Baunt and North-Baikal), their language status, common phonetic and semantic features and differences among them. However, in contrast with other Ewenki dialects, Nercha Ewenki is not spoken any more (Khabtagaeva 2017: 34–35). In my published monograph I very briefly mentioned the Khamnigan Ewenki people, but I did not explicitly connect them with the Nercha Ewenki people. The fieldwork among the Khamnigan Ewenki people (September 2017, Hulunbuir, China) has proven my early assumptions to be correct.

The aim of this paper is to compare the lexical material of Nercha Ewenki published by Castrén (1856) with the Manchurian Khamnigan Ewenki data published by Janhunen (1991) and our fieldwork materials.

2 Ewenki dialects

The Ewenki language belongs to the Tungusic language family, traditionally believed to form the Altaic language family together with the Turkic and Mongolic languages. Although the classification of Tungusic languages is not definitive, Tungusic languages are traditionally divided into two branches (for more details, see Khabtagaeva 2017: 17–18). The northern branch includes 51 dialects and subdialects of Ewenki, Ewen or Lamut, Negidal, etc. The southern branch is divided into two groups. The Manchuric group consists of Jurchen or Old Manchu,

¹Nowadays Khamnigan Mongol has three dialects, which are close to each other linguistically but differ geographically. Khamnigan Mongol is spoken in three different countries. (1) The Trans-Baikalian or Onon Khamnigan dialect is spoken in the Chita Province of Russia, and in several Regions of the Buryat Aga National District. (2) The Khamnigan Mongol dialect of Mongolia is spoken in the northeastern region of Mongolia in Khentei Province and in Dornod Province. The Khamnigan dialect of Dadal sum of Khentei Province was investigated by Uray-Kõhalmi. (3) The Manchurian Khamnigan dialect is spoken in the northeastern region of China, in the Hulunbuir district in the Ewenki Autonomous Arrow of the Old Bargut Banner.

Khamnigan Mongol is an endangered Mongolic language, its speakers total approximately 2,600 persons: Onon Khamnigans number 600, Manchurian Khamnigans 1,500, while Khamnigan Mongols of Mongolia 530 speakers (for more details and references, see Khabtagaeva 2017: 49).

Manchu, and its sole living member Sibe ~ Sibó (Xibe ~ Xibo). The Amuric group includes Nanai, Ulcha, Oroch, Oroch, and Udihe.²

The Ewenki people live in Russia, China and Mongolia,³ scattered over a vast territory. Janhunén (1997: 130) suggests a differentiation of the Ewenki people into two groups: (1) the Siberian Ewenki in Russia and (2) the Manchurian Ewenki in China.

1. In Russia, the Ewenki people live in small groups of a few thousand people, far from each other. They number approximately 37,100 (Ewenki statistics in 2010).⁴ The Ewenki language in Russia has 51 dialects, which can be grouped into northern, southern and eastern branches (Atknine 1997: 115; Bulatova 2002: 270–271). Literary or Standard Ewenki is based on the Stony Tunguska dialect, which belongs to the southern branch (for more details, see Khabtagaeva 2017: 19–23);
2. In Northeastern China, along the rivers Mergel and Hailar, in the Ewenki Autonomous Banner of the Hulun Buir Region at least four historically and linguistically distinct groups live in close proximity:
 - a) the Solon Ewenkis – the largest group (25,000 or 90% of the Ewenki by nationality in official statistics);⁵

²A new classification of Tungusic languages was recently proposed by Janhunén (2012: 16), where the northern branch includes the Ewenic group as well as the Udegheic group, while the southern branch consists of the Nanaic and Jurchenic groups. Accordingly, the Tungusic languages are divided into two branches. The NORTHERN TUNGUSIC branch includes the Ewenic group: a) Siberian Ewenic (Ewen, Arman, Ewenki, Negidal, Orochen and Urulga dialect of Khamnigan Ewenki); and b) Manchurian Ewenic (Mankovo dialect of Khamnigan Ewenki, Nonni Solon, Hailar Solon and Ongkor Solon). The Udegheic group includes Udeghe and Oroch. The SOUTHERN TUNGUSIC branch contains two groups: a) the Nanaic (Nanai, Kili and Kilen) and Ulchaic (Ulcha and Oroch) group and b) the Jurchenic group (Jurchen, Manchu and Sibe).

³A group of Ewenkis of unknown size also lives near Lake Buir in Northeastern Mongolia.

⁴The Republic of Yakutia – 21,008; the Krasnoyarsk Region – 4,372; the Khabarovsk Region – 4,101; the Republic of Buryatia – 2,974; the Province of Amur – 1,481; the Zabaikalsk Region – 1,387; the Province of Irkutsk – 1,272; the Province of Sakhalin – 209 and other Provinces – 312. On the geographical position of the Ewenki dialects in Russia, see the appended map in Vasilevič's (1958) dictionary.

⁵Historically they are a satellite group of the Dagur. Just like the Dagurs, the Solon Ewenkis used to live in the Zeya basin north of the Middle Amur, from where the Qing government relocated them to other parts of Manchuria in 1654. Today they live in four different places. One place is the Zeya basin in Russia (Bulatova 1987), while the other three are in Manchuria, China: along the Nonni basin in Nehe country, in the Ewenki Autonomous Banner of Hulun Buir, and in Ili Region of Xinjiang (Janhunén 1997: 130–131).

- b) the Orochens – another major group of Ewenki speakers in China forming four geographical and dialectal subgroups, including approximately 7,000 people;⁶
- c) the Manchurian Reindeer Tungus or *Yakute*, officially classified together with the Solon as *Ewenke*, numbering about 200 individuals;⁷
- d) the Khamnigan Ewenkis – the last Ewenki group in China that numbers about 10,000 people (for more details, see below).

While they differentiate themselves from each other, most groups are erroneously called *Ewenke* by the administration and some Chinese linguists (Janhunnen 1997: 130–131).

3 Nercha Ewenki

The homeland of Nercha Ewenki people was the southeastern part of Transbaikalia. Today the territory is situated in the Aga Buryat National District of Chita Province. Politically and geographically it is not Buryatia, but it is the place where Buryat people have lived for a long time.

We do not have any current information about Nercha Ewenki speakers. The dialect is likely extinct, with only historical and ethnographic materials available. Uvarova (2006) focuses on historical facts, the social structure, and some cultural features of 18th to 20th century Nercha Ewenki peoples, with no examination of their language. The historical materials, including statistics, different government ordinances, and laws, were collected from various archives in Russia, originating mostly from the 19th century and author's fieldwork material collected in the 1970s (Uvarova 2006: 10–13). According to Uvarova, by the early 20th century the Nercha Ewenki people merged with the Buryat and Russian populations of Transbaikalia. In the 1970 census, only 32 persons indicated the Ewenki language as their mother tongue (Uvarova 2006: 9; 122). The total assimilation with the Buryats was completed by the 1980s.

Based on archive materials, Tugolukov (1975) characterizes the traditional culture and religion of Nercha *murčens*, i.e. 'horse breeders', provides statistical data on the Ewenki tribes in the 18th and 19th centuries, and describes various historical facts connected with the Gantimur dynasty. The ethnic history of the Nercha

⁶Their ancestors moved to China from the regions north of the Amur during the 18th century and nowadays settled the two Khingan Ranges (Janhunnen 1997: 131–132).

⁷They live in the region of the river Jiliuhe in the Hulun Buir Province, and are culturally close to the Orochen and different from the Solon (Janhunnen 1997: 132).

Ewenki people is closely linked with the name Gantimur, who was a leading representative of the Nercha *murčen* people. From the middle of 17th century he started to pay tribute to Russia. His oldest son Katana converted to Christianity and was presented to the court of Russian Tsar Petr Alekseevich. From the end of the 17th century Nercha *murčens* started to guard the Russian-Chinese border. The exact origin of Gantimur is unclear. According to various archive sources (for details, see Tugolukov 1975: 98–103), he was Tungus or Dagur. Tugolukov (1975: 101–102) concludes that Gantimur was of Tungusic origin but “Dagurified”, which may be confirmed with Gantimur being of Nercha Ewenki origin from the Dulikagir tribe, i.e. the original Ewenki tribe which is not Mongolic.

Here it also needs to be considered, as noted correctly by Janhunen (1991: 16), that the Transbaikalian aboriginal population represents a complex mixture of Mongolic and Tungusic elements. According to Janhunen (1991: 16), roughly one half of the Khamnigan Ewenki clans had Tungusic-speaking ancestors, while the ancestors of the other half were Mongolic-speaking. Therefore, it cannot be excluded that Gantimur was of either Tungusic or Mongolic origin. It is important to mention that the Nercha Ewenkis, like other Ewenki people, had no right to marry a person of the same tribe until the ninth generation, which may have led to their assimilation with Mongolic people. In Tugolukov’s opinion (1975: 109), before the 12th or 13th century, the Ewenki people were reindeer breeders and later assimilated to Mongolic people. Besides intermarriage, another reason for assimilation was the change of lifestyle from reindeer breeding to horse breeding.⁸

The full list of Nercha Ewenki tribes and their numbers in 1762 and 1823 are given by Tugolukov (1975: 93), whose data includes 14 tribes in the following order: *Balikagir* (87 persons), *Bajagir* (477 persons), *Wakasil* (42 persons), *Gunow* (258 persons), *Dolot* (126 persons), *Dulikagir* (436 persons), *Konur* (171 persons), *Lunikir* (304 persons), *Namyat* (512 persons), *Počegor* (248 persons), *Sortoc* (240 persons), *Uzon* (358 persons), *Ulyat* (140 persons), and *Čemčagir* (346 persons). The speakers of the Borzya dialect of Manchurian Khamnigan Ewenki comprise the *Balkiegid*, *Bayagiid*, *Čimčagiid*, *Duligaad* and *Marugiid* clans, while the speakers of the Urulyngui dialect include the *Namied*, *Altanganuud*, *Čibčiniüüd*, *Jal-tood*, *Koonud*, *Dulaad*, *Galjood*, *Ulied* and *Ūjeed* (Janhunen 1991: 14). The forms with the final consonant *-d* (*-gid/-gad*, *-Ad*, *-nuud*) in the clan names are possibly connected with the Mongolic plural forms. Thus, if we compare all the above-mentioned clan names, the common clans of Nercha and Khamnigan Ewenki people are *Bajagir*, *Dulikagir*, *Čemčagir*, *Konur*, *Ulyat* and *Namyat*.

⁸Ewenki legends tell us that when they came out with their reindeers to the steppe, they were forced to change their lifestyle because of the absence of reindeer moss (Tugolukov 1975: 106).

It seems that most of the Nercha Ewenkis were the ancestors of Manchurian Khamnigan Ewenkis, who crossed the border into Manchuria and China during the initial years of Soviet rule (ca. 1918–1932), moving with the ancestors of Khamnigan Mongols and Shinehen Buryats for “a better life” (Janhunen 1997: 130).

The first linguist who worked on the Nercha Ewenki dialect with native consultants was M. A. Castrén (1856). His work was translated into Russian by Ye. I. Titov and published as the appendix in his *Tungus–Russian dictionary* (Titov 1926). Titov was the second and possibly last researcher of Nercha Ewenki. Titov met with people from the Bultegir and Turuyagir clans (Titov 1926: ix) and claimed that the people spoke similar dialects. Today the people of the Turuyagir clan belong to the Baunt Ewenki people (Khabtagaeva 2017: 29), while the people of the Bultegir clan were mentioned among the Dagur people (Vasilevič 1969: 265). The lexical material in Vasilevič’s (1969) *Ewenki–Russian dictionary* was probably collected from Castrén’s and Titov’s works.⁹

Another important fact is that the histories of Nercha Ewenkis and Transbaikalian Khamnigan Mongols are closely related to each other. The two groups were likely often confused in Russian official documents and were considered to be *Tungus* and later *Ewenki*. For instance, the Russian anthropologist Tal’ko-Gryncevič (1904: 77) wrote that the number of *Ewenkis* who adopted Buddhism exceeded the number of Buryats. Or when we read that at the beginning of the 19th century there were six Buddhist monasteries in the Urulga territory built by *Ewenkis* (Galdanova et al. 1983: 41), we have to suppose that they mean the Nercha Ewenki and Transbaikalian (or Onon) Khamnigan Mongol people.

4 Khamnigan Ewenki

Speakers of Manchurian Khamnigan Ewenki of China use two separate Ewenki dialects, both distinct from all other known Ewenki dialects and also relatively different from each other. The first one is the Borzya dialect, referring to the Upper Borzya river on the Russian side, while the second one is the Urulyungui dialect, referring to the river Urulyungui also on the Russian side (Janhunen 1991: 11–12; 1997: 132–133). Nowadays, geographically both rivers are situated in Transbaikalia, where the Nercha Ewenki people formerly lived.

⁹During my comparison of the materials of Castrén’s and Vasilevič’s dictionaries, I noticed that Castrén’s transcription does not always coincide with Vasilevič’s. For instance, the consonant *c* in Vasilevič’s dictionary is incorrectly indicated as *č* and I had to correct it in my own materials (Khabtagaeva 2017).

As Janhunen (1997: 132–133) states, this group is ethnolinguistically the most atypical one, in that it is more or less congruous with the population speaking the Khamnigan Mongol language. They are bilingual in Ewenki and Khamnigan Mongol. Mongol is the dominant community language, while Ewenki is mainly used as an additional means of communication within many families. The main languages of interethnic communication between Khamnigan Ewenkis and Khamnigan Mongols is Khamnigan Mongol,¹⁰ so the Khamnigan Mongols do not speak Ewenki. As Janhunen noted in the 1990s (1991: 11–15), Khamnigan Mongol is a stable and homogeneous variety, showing no essential variation within the community. Ewenki is destined eventually to lose its remaining role as a family language. By contrast, Khamnigan Mongol may well further strengthen its position as the principal community language in the Mergel region (for details, see Janhunen 1997: 130).

When interviewing the Khamnigan Ewenki people during our fieldwork, we observed a slight shift in self-identification compared to Janhunen’s description. Our Khamnigan informants mostly emphasized their Ewenki affiliation, stating that Ewenki is probably the original language of the community while Khamnigan Mongol (termed [*evenkilig mongol üge*] by them) was adopted later “somewhere in Russia”. At the same time, the speakers supposed that the *Boorji* variety existed earlier and was the original language of the Khamnigan Ewenki community while *Namieetii* was a Mongolized variety adopted by them in Manchuria. *Boorji* is a Borzya dialect, while *Namieetii* is connected to the clan *Namied*, which is listed among the Khamnigan Ewenki tribes of Mongolic origin and as speakers of the Urulyungui dialect (Janhunen 1991: 16, 14). As has been mentioned earlier, besides the *Namied*, the speakers of the Urulyungui dialect include the *Altan-ganuud*, *Čibčinüüd*, *Koonuud*, *Dulaad*, *Galjood*, *Ulited* and *Ūjeed* clans, while the speakers of Borzya dialect comprise the *Balkiegid*, *Bayagiid*, *Čimčagiid*, *Duligaad* and *Marugiid* clans (Janhunen 1991: 14).

The slight change of Khamnigan Ewenkis in “self-classification” as *Ewenki* people in recent years may be connected to the recent promotion of Ewenki culture in China and governmental support for the endangered Ewenki culture, which enables the Khamnigans to profit from their Ewenki identity. The idea of the unity of the three Ewenki branches of China has been promoted in various spheres. For instance, in Hailar we had an opportunity to meet with the Solon Ewenki scholar Do Dorji, who is the chief-editor of *Ewenki-Chinese* (1998) and *Ewenki-Mongol* (2013) dictionaries where he treats the “three Ewenki branches”

¹⁰As other Mongolian speakers (e.g. Buryat, Dagur, Ordos, Khorchin, Kharchin, etc.) in Inner Mongolia, Khamnigan Mongols and Khamnigan Ewenkis speak Standard Mongolian.

together, taking Solon as a base. Also, our 58-year-old informant told us that her mother tongue is very close to the Solon, Orochen and Yakut Ewenki varieties. When we asked about their connection with the Nercha Ewenki people, our informant replied that she did not hear about them recently but knows that they came from the Russian side and her parents were fluent in Russian.¹¹ It is an interesting fact that our Khamnigan Ewenki informants (a 59 year-old man and a 58 year-old woman) are fluent in both Ewenki (*Boorji* and *Namieetii*) varieties and Khamnigan Mongol. As they told us, both Ewenki varieties are very close to each other, but *Boorji* is the ‘original’ Ewenki, while *Namieetii* is “mixed and primitive”, i.e. *Namieetii* has more Mongolic and Russian elements.

It is important to mention the religion of the Khamnigan Ewenki people. As our informants told us, the ‘original’ religion was Christianity, after migration from Russia in the village where the Khamnigan Ewenki people lived there was one church with a priest, but during the Cultural Revolution in China in 1960-s and 70-s they were forced to ‘give up’ their religion and became atheists. Now the Khamnigan Ewenki people believe in shamanism, regularly visit shamans, and perform shamanistic rites.

A brief grammatical sketch of Khamnigan Ewenki was provided by Janhunén (1991).

5 Comparative analysis of Nercha Ewenki and Khamnigan Ewenki materials

Linguistically, the Nercha dialect belongs in the southern sibilant group, representing the hissing type (*s-*, *VsV*)¹² (Atkine 1997: 115; Bulatova 2002: 270–271; Khabtagaeva 2017: 19–20). The Khamnigan Ewenki variety also shares this phonetic feature.¹³

¹¹The personal names of our informant’s parents were unique: the mother’s name was Darima (a typical Buryat or Khamnigan Mongol name, it does not exist among other Mongolic people such as Khalkha, Oirat, Dariganga, etc.), the father’s name was Prank (cf. Russian *Frank*) and the uncle’s name was Mark.

¹²The main criterion used in the classification of the dialects is the fate of the Common Tungusic consonant *s in initial and intervocalic positions. In the three branches there appear the representations *h*, *s* and *š*. E.g. Common Tungusic ‘ear’ and ‘woman’ in northern group (spirant *h-*, *VhV*) are *hēn* and *ahī*, in southern (sibilant *s-*, *VsV* and *š-*, *VšV*) are *sēn* / *šēn* and *asī* / *ašī*, in eastern (sibilant and spirant *s-*, *VhV*) are *sēn* and *ahī*, respectively (Khabtagaeva 2017: 20).

¹³According to Castrén’s (1856) and Janhunén’s (1991) materials, Common Ewenki *sele* ‘iron’: Nercha Ewenki, Khamnigan Ewenki *sele*; Common Ewenki *asī* ‘woman’: Nercha *āśī*, Khamnigan Ewenki *asī*, Common Ewenki *ēsa* ‘eye’: Nercha Ewenki *īsa* ~ *ēsa*, Khamnigan Ewenki (Urulga) *iesā*, (Borzya) *isa*, etc.

The following provides a list of common Tungusic words in the Nercha Ewenki and Manchurian Khamnigan Ewenki dialects from Castrén's (1856) and Janhunen's (1991) works, and from our fieldwork material. The Nercha Ewenki dialect includes the Urulga and Mankovo subdialects, while Khamnigan Ewenki includes Borzya and Urulyngui. The extinct Mankovo subdialect of Nercha Ewenki corresponds to Borzya, while the extinct Urulga subdialect of Nercha Ewenki has a close relation with Urulyngui in Manchuria (Janhunen 1991: 12). Since the influence from Solon Ewenki language is assumed, the comparative data from the Hulunbuir Solon dialect is added (Dorji & Banzhibomi 1998; 2013; Chaoke 2014b). Additionally, the data of Orochen or Oroqen of Hulunbuir (Chaoke 2014a), Siberian Ewenki dialects (Vasilevič 1958) and other Tungusic languages are added (Cincius 1975/77; Hauer 1952-1955; Stary 1990; Zikmundová 2013).

5.1 Shared lexicon

In most cases, Nercha Ewenki and Khamnigan Ewenki have the common Tungusic vocabulary, which is also present in other Ewenki dialects.

- (1) A list of potential cognates shared by Nercha and Khamnigan Ewenki:
 - a. Kinship terms:
 - i. 'elder brother': Nercha Ewenki, Khamnigan Ewenki **akin**;
cf. Solon Ewenki *ahij*; Orochen *akin*; Siberian Common Ewenki *akīn*;
other Northern Tungusic: Lamut *akan*; Negidal *ahin*; *Southern Tungusic*: Nanai, Ulcha, Udihe *aga*; Oroch *aki*; Orok *aka*; Manchu *agu*; Sibe *avūⁿ*
(Castrén 1856: 71a; Janhunen 1991: 73; Dorji & Banzhibomi 1998: 14; Chaoke 2014a: 160; Vasilevič 1958: 21a; Cincius 1975/77 1: 23; Hauer 1952-1955 1: 14; Zikmundová 2013: 204);
 - ii. 'younger brother': Nercha Ewenki, Khamnigan Ewenki **nekün**;
cf. Solon Ewenki *nəhuŋ*; Orochen *nekun*; Siberian Ewenki: Sakhalin *nekün*; Podkamennyj, May, Tokko, Tommot, Urmi, Uchur, Chulman *nekē*;
other Northern Tungusic: Lamut *nu*; Negidal *nekun* ~ *nehun*;
Southern Tungusic: Nanai, Ulcha, Orok *neu*; Udihe *neŋu*; Oroch *neku*; Manchu *non*
(Castrén 1856: 85; Janhunen 1991: 24; Dorji & Banzhibomi 1998: 480a; Chaoke 2014a: 161; Vasilevič 1958: 302a; Cincius 1975/77 1: 617b-618; Hauer 1952-1955 3: 720);

- iii. ‘daughter-in-law’: Nercha Ewenki, Khamnigan Ewenki **kūkin**;
cf. Solon Ewenki *hūhiŋ*; Siberian Ewenki: Podkamennyj,
Yerbogochen, Barguzin, Zeya, Ilimpeya, May, Tokko, Tommot,
Uchur *kukīn*;
other Northern Tungusic: Lamut *kōken*; Negidal *kukin* ~ *kuhin*;
Remaining lgs. *n.a.*¹⁴
(Castrén 1856: 81; Janhunen 1991: 23; Dorji & Banzhibomi 1998:
303a; Vasilevič 1958: 217a; Cincius 1975/77 1: 425b);
- iv. ‘mother’: Nercha Ewenki (Urulga), Khamnigan Ewenki
(Urulyungui) **enin**;
cf. Solon Ewenki *əniŋ*; Orochen *enin*; Siberian Common Ewenki
eñin;
other Northern Tungusic: Lamut *eñin*; Negidal *enin*; *Southern
Tungusic*: Nanai, Ulcha, Udihe, Oroch, Orok *eni*; Manchu *eniyeŋ*
‘female moose’; Sibe *əñi*
(Castrén 1856: 73; Janhunen 1991: 23; Dorji & Banzhibomi 1998:
179a; Chaoke 2014a: 160; Vasilevič 1958: 562a; Cincius 1975/77 2:
456; Hauer 1952-1955 1: 253; Zikmundová 2013: 210);
- b. Names of body parts:
- i. ‘mouth’: Nercha Ewenki, Khamnigan Ewenki **amŋa**;
cf. Solon Ewenki *amma*; Orochen *amŋa*; Siberian Common
Ewenki *amŋa*;
other Northern Tungusic: Lamut *amŋa* ~ *amga*; Negidal *amŋa*;
Southern Tungusic: Nanai, Ulcha, Udihe *aŋma*; Oroch *amma*; Orok
amŋa ~ *aŋma*; Manchu *aŋga*; Sibe *an*
(Castrén 1856: 72; Janhunen 1991: 51; Dorji & Banzhibomi 1998:
27b; Chaoke 2014a: 159; Vasilevič 1958: 28b; Cincius 1975/77 1: 38;
Hauer 1952-1955 1: 52; Zikmundová 2013: 205);
- ii. ‘beard’: Nercha Ewenki, Khamnigan Ewenki **gurgakta**;
cf. Solon Ewenki *guggatt*; Siberian Common Ewenki *gurgakta*;
other Northern Tungusic: Lamut *gurgat*; Negidal *goigakta*;
Southern Tungusic: Nanai *gogakta*; Udihe *gugakta*; Oroch *gugahta*;
Orok *gudahta*; Remaining lgs. *n.a.*
(Castrén 1856: 82; Janhunen 1991: 40; Dorji & Banzhibomi 1998:
230a; Vasilevič 1958: 97b; Cincius 1975/77: 173b);

¹⁴The abbreviation *n.a.* means that the form is not available, it may be present but not found in the considered dictionaries.

- iii. 'skin': Nercha Ewenki, Khamnigan Ewenki **nanda**;
 cf. Solon Ewenki *nanda*; Orochen *nana*; Siberian Ewenki:
 Podkamennyj, Barguzin, Yerbogochén, Zeya, Nepa, Sakhalin,
 Tokko, Urmi, Chumikan *nanna*; Ilimpeya, May, North-Baikal,
 Uchur *nanda*; Sym *nandra*;
other Northern Tungusic: Lamut *nanra* ~ *nanda*; Negidal *nana*;
Southern Tungusic: Nanai, Ulcha *nanta*; Udihe *néhe*; Oroch *naha* ~
nasa ~ *niha*; Orok *natta* ~ *nata*; Manchu, Sibe *n.a.*
 (Castrén 1856: 85; Janhunen 1991: 49; Dorji & Banzhibomi 1998:
 473b; Chaoke 2014a: 158; Vasilevič 1958: 276b; Cincius 1975/77 1:
 583b);
- iv. 'neck': Nercha Ewenki **nikimna** ~ **nikinma**; Khamnigan Ewenki
nikimne;
 cf. Solon Ewenki *niham*; Orochen *nikimna*; Siberian Ewenki:
 Podkamennyj, Nepa, Upper Lena, North-Baikal, Tungir, Zeya,
 Aldan, Urmi, Ayan, Sakhalin *nikinma*; Yerbogochén *nikinma* ~
nikimña; Nepa *nikinmña*; Barguzin *nikin*; Ilimpeya, North-Baikal,
 Uchur *nikimda*;
other Northern Tungusic: Lamut *níken*; Negidal *nihma*; *Southern*
Tungusic: Ulcha *níkin*; Orok *nikimña*; Remaining lgs. *n.a.*
 (Castrén 1856: 85; Janhunen 1991: 49; Dorji & Banzhibomi 1998:
 487b; Chaoke 2014a: 159; Vasilevič 1958: 291b; Cincius 1975/77 1:
 591);
- c. Names of animals:
- i. 'fish': Nercha Ewenki, Khamnigan Ewenki **oldo**;
 cf. Solon Ewenki *n.a.*; Orochen *olo*; Siberian Ewenki:
 Podkamennyj, Nepa, Yerbogochén, Tungir, Zeya, Aldan, Urmi,
 Chumikan, Sakhalin *ollo*; Ilimpeya, North-Baikal, Uchur, Upper
 Lena *oldo*; Sym *oldro*;
other Northern Tungusic: Lamut *olra*; Negidal *olo*; *Southern*
Tungusic: Nanai *olo*; Ulcha, Orok *holto* 'cooked fish'; Udihe *oloho*
 'cooked fish'; Oroch *okto* < **olto*; Manchu, Sibe *n.a.*
 (Castrén 1856: 75; Janhunen 1991: 23; Chaoke 2014a: 156; Vasilevič
 1958: 320a; Cincius 1975/77 2: 14);
- ii. 'kind of duck': Nercha Ewenki, Khamnigan Ewenki **tarmi**;
 cf. Solon Ewenki *n.a.*; Siberian Common Ewenki *tarmi* 'drake'

other Northern Tungusic: Lamut ; Negidal ; *Southern Tungusic*: Nanai, Ulcha *tarmi*; Udihe *tanmi*; Oroch *tajmi*; Orok *n.a.*; Manchu *tarmin*; Sibe *n.a.*

(Castrén 1856: 86; Janhunen 1991: 48; Vasilevič 1958: 388b; Cincius 1975/77 2: 169a; Hauer 1952-1955 3: 891);

- iii. ‘goose’: Nercha Ewenki, Khamnigan Ewenki ***nuŋnakī***;
cf. Solon Ewenki *nunnahi*; Orochen *niunnaki*; Siberian Ewenki: Podkamennyj, Ayan, Aldan, Barguzin, Yerbogochen, Ilimpeya, May, Sakhalin, Tommot, Tungir, Urmi, Uchur, Chumikan *nuŋnakī*; Nepa, Tokmin *numnakī*;

other Northern Tungusic: Lamut *n.a.*; Negidal *noŋŋahi*; *Southern Tungusic*: Nanai *noŋŋa*; Ulcha *nuŋŋa*; Udihe *nuŋŋai*; Oroch *nuŋŋahi*; Orok *nuŋna* ~ *nuŋŋa*; Manchu *nioŋniyaha*; Sibe *n.a.*

(Castrén 1856: 86; Janhunen 1991: 51; Dorji & Banzhibomi 1998: 497b; Chaoke 2014a: 156; Vasilevič 1958: 307b; Cincius 1975/77 1: 646b; Hauer 1952-1955 2: 707);

- iv. ‘squirrel’: Nercha Ewenki (Mankovo), Khamnigan Ewenki (Borzya) ***ülükī***;

cf. Solon Ewenki *uluhi*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Sakhalin *ulukī*; Ayan *olokī*;

other Northern Tungusic: Lamut *uliki*; Negidal *oluki*; *Southern Tungusic*: Nanai *hulu*; Orok, Ulcha *holo*; Udihe *olohi*; Oroch *oloki*; Manchu, Sibe *ulhu*

(Castrén 1856: 78; Janhunen 1991: 52; Chaoke 2014b: 42; Vasilevič 1958: 440a; Cincius 1975/77 2: 263b; Hauer 1952-1955 3: 957; Stary 1990: 92);

d. Inanimate nature:

- i. ‘river’: Nercha Ewenki, Khamnigan Ewenki ***birā***;

cf. Solon Ewenki *bera*; Orochen *bira*; Siberian Common Ewenki *bira*;

other Northern Tungusic: Lamut *bira*; Negidal *bija*; *Southern Tungusic*: Nanai, Ulcha *bira*; Udihe *b^leæsa*; Oroch *biaka*; Orok *n.a.*; Manchu, Sibe *bira*

(Castrén 1856: 95; Janhunen 1991: 23; Dorji & Banzhibomi 1998: 71b; Chaoke 2014a: 153; Vasilevič 1958: 56a; Cincius 1975/77 1: 84; Hauer 1952-1955 1: 96; Zikmundová 2013: 206);

e. Names of metals:

- i. 'iron': Nercha Ewenki, Khamnigan Ewenki *sele*;
 cf. Solon Ewenki *səl*; Orochen *sele*; Siberian Ewenki:
 Podkamennyj, Nepa, Barguzin, Tungir, Zeya, Uchur, Urmi, Aldan,
 Chumikan, Ayan, Sakhalin *sele*; Yerbogochon, Ilimpeya,
 VilyuiVilyuy *hele*; Sym, North-Baikal *šele*;
other Northern Tungusic: Lamut *hel*; Negidal *sele*; *Southern
 Tungusic*: Nanai, Ulcha, Udihe, Oroch, Orok, Manchu *sele*; Sibe
selei [ʃugūn] 'railroad'
 (Castrén 1856: 91; Janhunen 1991: 23; Dorji & Banzhibomi 1998:
 592b; Chaoke 2014a: 163; Vasilevič 1958: 376a; SSTMJa 2: 140;
 Hauer 3: 778; Stary 1990: 76);

f. Names of plants:

- i. 'tree': Nercha Ewenki, Khamnigan Ewenki *mō*;
 cf. Solon Ewenki *mo*; Orochen *mō*; Siberian Common Ewenki *mō*;
other Northern Tungusic: Lamut, Negidal *mō*; *Southern Tungusic*:
 Nanai, Ulcha, Udihe, Oroch, Orok, Manchu *mō*; Sibe *mo*
 (Castrén 1856: 96; Janhunen 1991: 30; Dorji & Banzhibomi 1998:
 444b; Chaoke 2014a: 153; Vasilevič 1958: 254b; Cincius 1975/77 1:
 540; Zikmundová 2013: 218);
- ii. 'pine tree': Nercha Ewenki, Khamnigan Ewenki *ǰagda*;
 cf. Solon Ewenki *ǰadda* ~ *ǰagda*; Siberian Common Ewenki *ǰagda*;
other Northern Tungusic: Lamut, Negidal *ǰagda*; *Southern Tungusic*:
 Nanai *ǰāgda*; Ulcha *ǰagda*; Udihe, Oroch, Orok *n.a.*; Manchu
ǰaqdan; Sibe *n.a.*
 (Castrén 1856: 93; Janhunen 1991: 23; Dorji & Banzhibomi 1998:
 350a; Vasilevič 1958: 146b; Cincius 1975/77 1: 242);
- iii. 'steppe lily': Nercha Ewenki, Khamnigan Ewenki: Borzya *ǰōgta*;
 cf. Solon Ewenki *ǰōtt*; Siberian Ewenki: Barguzin *ǰokta*;
Southern Tungusic: Manchu *ǰoqtoda* ~ *ǰoqtonda*; Remaining lgs. *n.a.*
 (Castrén 1856: 93; Janhunen 1991: 42; Dorji & Banzhibomi 1998:
 379b; Vasilevič 1958: 112b; Cincius 1975/77 1: 262);

g. Time measures:

- i. 'tomorrow': Nercha Ewenki, Khamnigan Ewenki *timī*;
 cf. Solon Ewenki *timašij*; Orochen *timāna*; Siberian Common
 Ewenki *timānī*;

other Northern Tungusic: Lamut *t'em'en*; Negidal *t'emana*;
Southern Tungusic: Nanai *čimaj*, Ulcha, Udihe *tīmani*; Oroch
timaki; Orok *čimani*; Manchu *čimari*; Sibe *čimar*
(Castrén 1856: 87; Janhunen 1991: 29; Dorji & Banzhibomi 1998:
689b; Chaoke 2014a: 155; Vasilevič 1958: 410b; Cincius 1975/77 2:
181; Zikmundová 2013: 207);

- ii. 'day': Nercha Ewenki: Urulga *inaŋ*, Man'kovo *ineŋi*; Khamnigan Ewenki *ineŋi*;
cf. Solon Ewenki *inaŋ* 'afternoon'; Orochen *iniyi*; Siberian Ewenki: Podkamennyj, Nepa, Ilimpeya *ineŋ* 'noon; south'; Sym, Barguzin, Nercha, Tungir, Zeya, Urmi, Chumikan, Ayan, Sakhalin *ineŋi* 'day';
'day, in the daytime': *other Northern Tungusic*: Lamut *ineŋ*; Negidal *ineŋi*; *Southern Tungusic*: Nanai *inie*; Udihe, Oroch *ineŋi*; Orok *inuŋi*; Ulcha *ineŋni*; Manchu *ineŋgi*; Sibe *ənəŋ* 'today'
(Castrén 1856: 74; Janhunen 1991: 57; Dorji & Banzhibomi 1998: 333b; Chaoke 2014a: 154; Vasilevič 1958: 175a; Cincius 1975/77 1: 318; Hauer 1952-1955 2: 499; Zikmundová 2013: 210);

h. Buildings and their parts:

- i. 'door': Nercha Ewenki, Khamnigan Ewenki *ürke*;
cf. Solon Ewenki *ükkə*; Orochen *urke*; Siberian Common Ewenki *urke*;
other Northern Tungusic: Lamut *urke*; Negidal *ujke*; *Southern Tungusic*: Nanai *ujke*; Ulcha *uče*, Udihe *uke ~ uče*; Oroch *ukke*; Orok *ute*; Manchu *uče*; Sibe *uči*
(Castrén 1856: 78; Janhunen 1991: 23; Dorji & Banzhibomi 1998: 727a; Chaoke 2014a: 162; Vasilevič 1958: 453a; Cincius 1975/77 2: 286; Hauer 1952-1955 3: 942; Zikmundová 2013: 223);

i. Names of foods:

- i. 'meat': Nercha Ewenki, Khamnigan Ewenki *ülde*;
cf. Solon Ewenki *üldə*; Orochen *ule*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochon, Tokmin, Upper Lena, Barguzin, Vitim, Tungir, Zeya, Aldan, Urmi, Ayan, Sakhalin *ulle*; Sym *uldre*; Ilimpeya, North-Baikal, Uchur *ulde*; Tokma *unle*; Ayan *ulre*;
other Northern Tungusic: Lamut *ulre*; Negidal *ule*; *Southern Tungusic*: Nanai *ulikse*; Ulcha *ulse*; Udihe *ulehe*; Oroch *ulese*; Orok *ulise*; Manchu, Sibe *n.a.*

(Castrén 1856: 78; Janhunen 1991: 41; Dorji & Banzhibomi 1998: 728a; Chaoke 2014a: 160; Vasilevič 1958: 439b; Cincius 1975/77 2: 262);

j. Clothes, utensils:

- i. ‘footwear, shoes’: Nercha Ewenki, Khamnigan Ewenki **unta**;
cf. Solon Ewenki *unta*; Siberian Common Ewenki *unta*;
other Northern Tungusic: Lamut *unta*; Negidal *onta*; *Southern Tungusic*: Nanai *ota*; Ulcha, Udihe, Oroch *unta*; Orok *utta*;
Manchu, Sibe *n.a.*

(Castrén 1856: 77; Janhunen 1991: 49; Dorji & Banzhibomi 1998: 530b; Vasilevič 1958: 448b; Cincius 1975/77 2: 275);

- ii. ‘knife’: Nercha Ewenki: Man’kovo **üfi** ~ **üči**; Khamnigan Ewenki **üči**;

cf. Solon Ewenki *n.a.*; Siberian Ewenki: Nepa, Uchur, Urmi, Sakhalin, Chumikan *ut-* ‘to wind, twist, twirl’;

‘to fix, to repair, to mend’: *other Northern Tungusic*: Lamut *ut-* ~ *uč-*; Negidal *ute-*; *Southern Tungusic*: Nanai *ute-* ‘to quilt clothes, blanket’; Ulcha *uteče* ‘seam’; Remaining lgs. *n.a.*

(Castrén 1856: 78; Janhunen 1991: 43; Vasilevič 1958: 456b; Cincius 1975/77 2: 293);

k. Abstract nouns connected with human life:

- i. ‘name’: Nercha Ewenki, Khamnigan Ewenki **gerbī**;
cf. Solon Ewenki *gərbi* ~ *gəbbi*; Siberian Common Ewenki *gerbī*;
other Northern Tungusic: Lamut *gerbe*; Negidal *gelbi*; *Southern Tungusic*: Nanai *gebu* ← Manchu; Ulcha, Orok *gelbu*; Udihe *gegbi*;
Oroch *gebbi*; Manchu *gebu*; Sibe *gəf*

(Castrén 1856: 81; Janhunen 1991: 40; Dorji & Banzhibomi 1998: 205; Vasilevič 1958: 100b; Cincius 1975/77 1: 180; Hauer 1952-1955 1: 339; Zikmundová 2013: 211);

l. Numbers:

- i. ‘one’: Nercha Ewenki, Khamnigan Ewenki **umun**;
cf. Solon Ewenki *əmuŋ*; Orochen *emun*; Siberian Ewenki:
North-Baikal, Tokma, Tungir *emūn*; Remaining dialects *umūn*;
other Northern Tungusic: Lamut *umen*; Negidal *omon*; *Southern Tungusic*: Nanai *em*; Ulcha, Udihe, Oroch *omo*; Orok *umūke*;
Manchu *emu*; Sibe *əm*

- (Castrén 1856: 77; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 174b; Chaoke 2014a: 169; Vasilevič 1958: 444b; Cincius 1975/77 2: 270; Hauer 1952-1955 1: 247; Zikmundová 2013: 209);
- ii. ‘two’: Nercha Ewenki, Khamnigan Ewenki *jūr*;
cf. Solon Ewenki, Orochen *jūr*; Siberian Common Ewenki *jūr*;
other Northern Tungusic: Lamut *jur*; Negidal *jul*; *Southern Tungusic*: Nanai *juer*; Ulcha *juel*; Udihe, Oroch *ju*; Orok *dū*;
Manchu *juwe*; Sibe *ju*
(Castrén 1856: 94; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 391b; Chaoke 2014a: 169; Vasilevič 1958: 143a; Cincius 1975/77 1: 276; Hauer 1952-1955 2: 563; Zikmundová 2013: 216);
- iii. ‘three’: Nercha Ewenki, Khamnigan Ewenki *ilan*;
cf. Solon Ewenki *ilan*; Orochen *ilan*; Siberian Ewenki: Aldan, Zeya, Ilimpeya *yelan*; Remaining dial. *ilan*;
other Northern Tungusic: Lamut *yelan*; Negidal *yelan* ~ *ilan*;
Southern Tungusic: Nanai *yelan*; Ulcha, Udihe, Oroch, Orok *ila*;
Manchu *ilan*; Sibe *yila*ⁿ
(Castrén 1856: 74; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 326a; Chaoke 2014a: 169; Vasilevič 1958: 162b; Cincius 1975/77 1: 305; Hauer 1952-1955 2: 490; Zikmundová 2013: 113);
- iv. ‘four’: Nercha Ewenki, Khamnigan Ewenki *digin*;
cf. Solon Ewenki *diyij*; Orochen *diyjin*; Siberian Common Ewenki *digin*;
other Northern Tungusic: Lamut, Negidal *digin*; *Southern Tungusic*: Nanai, Ulcha *duin*; Udihe *dī*; Oroch *dī*; Orok *jīn*; Manchu *duin*;
Sibe *duyi*ⁿ
(Castrén 1856: 90; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 130a; Chaoke 2014a: 170; Vasilevič 1958: 127b; Cincius 1975/77 1: 204; Hauer 1952-1955 1: 217; Zikmundová 2013: 113);
- v. ‘six’: Nercha Ewenki *nūñün* ~ *nūñün*; Khamnigan Ewenki *nūñün*;
cf. Solon Ewenki *niñuŋ*; Orochen *niñuñun*; Siberian Ewenki: North-Baikal *ñugun*; Remaining dial. *ñuñun*;
other Northern Tungusic: Lamut *ñuñi*; Negidal *ñuñi*; *Southern Tungusic*: Nanai, Ulcha *ñuñgu(n)*; Udihe, Oroch *ñuñu*; Orok *ñuñg’ē*; Manchu *niñgun*; Sibe *ñiñu*ⁿ
(Castrén 1856: 86; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 490b; Chaoke 2014a: 170; Vasilevič 1958: 308a; Cincius 1975/77 1: 647; Hauer 1952-1955 2: 703; Zikmundová 2013: 113);

- vi. ‘seven’: Nercha Ewenki, Khamnigan Ewenki **nadan**;
 cf. Solon Ewenki *nadan*; Orochen *nadan*; Siberian Common
 Ewenki *nadan*;
other Northern Tungusic: Lamut, Negidal *nadan*; *Southern
 Tungusic*: Nanai, Ulcha, Udihe, Oroch, Orok *nada*; Manchu *nadan*;
 Sibe *nadaⁿ*
 (Castrén 1856: 85; Janhunen 1991: 76; Dorji & Banzhibomi 1998:
 469b; Chaoke 2014a: 170; Vasilevič 1958: 273b; Cincius 1975/77 1:
 576; Hauer 1952-1955 2: 684; Zikmundová 2013: 113);

m. Qualitative adjectives:

- i. ‘good’: Nercha Ewenki, Khamnigan Ewenki **aja**;
 cf. Solon Ewenki *aja*; Orochen *aji*; Siberian Common Ewenki *aja*;
other Northern Tungusic: Lamut *aj*; Negidal *aja*; *Southern Tungusic*:
 Nanai, Ulcha, Udihe, Oroch, Orok *aja*; Manchu, Sibe *n.a.*
 (Castrén 1856: 71; Janhunen 1991: 95; Dorji & Banzhibomi 1998:
 49a; Chaoke 2014a: 165; Vasilevič 1958: 44a; Cincius 1975/77 1: 18);
- ii. ‘bad’: Nercha Ewenki, Khamnigan Ewenki **erū**;
 cf. Solon Ewenki *əru*; Orochen *eru*; Siberian Ewenki:
 Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Tokma, Tungir,
 Zeya, Aldan, Uchir *erū*; Podkamennyj, Yerbogochen, Ilimpeya
urū;
other Northern Tungusic: *n.a.*; *Southern Tungusic*: Nanai *eru*; Ulcha
orkin; Oroch, Orok *orke*; Manchu, Sibe *n.a.*
 (Castrén 1856: 73; Janhunen 1991: 52; Dorji & Banzhibomi 1998:
 185a; Chaoke 2014a: 165; Vasilevič 1958: 566a; Cincius 1975/77 2:
 465–466);
- iii. ‘warm’: Nercha Ewenki, Khamnigan Ewenki **nama**;
 cf. Solon Ewenki *namaddi* (< **namagdi* < *ńama*+*gdi* Solon
 denominal noun/adjective suffix); Orochen *niama*; Siberian
 Common Ewenki *ńama*;
other Northern Tungusic: Lamut *ńam*; Negidal *ńamagdi*; *Southern
 Tungusic*: Nanai, Ulcha, Oroch *ńama*; Udihe *ńamahi*; Orok
ńamauli; Manchu, Sibe *n.a.*
 (Castrén 1856: 85; Janhunen 1991: 57; Dorji & Banzhibomi 1998:
 471a; Chaoke 2014a: 166; Vasilevič 1958: 310b; Cincius 1975/77 1:
 630–631);

- iv. ‘other’: Nercha Ewenki, Khamnigan Ewenki **hüntü**;
cf. Solon Ewenki *əntu*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Chumikan, Sakhalin *huŋtu*; Sym *uŋtu*;
other Northern Tungusic: Lamut *hunte*; Negidal *heŋte* ~ *huŋtu*;
Southern Tungusic: Udihe *hoŋto*; Oroch *hoŋto* ~ *honto*; Nanai, Ulcha, Orok *n.a.*; Manchu *enču*; Sibe *unc^u* ‘different’
(Castrén 1856: 83; Janhunen 1991: 49; Dorji & Banzhibomi 1998: 180b; Vasilevič 1958: 496a; Cincius 1975/77 2: 349–350; Hauer 1952-1955 1: 249; Zikmundová 2013: 224);
- n. Demonstratives:
- i. ‘this’: Nercha Ewenki **er**, Khamnigan Ewenki: Borzya **eri** ~ **er**;
cf. Solon Ewenki *əri*; Orochen *eri*; Siberian Common Ewenki *er* ~ *eri*;
other Northern Tungusic: Lamut *er*; Negidal *ej*; *Southern Tungusic*: Nanai, Ulcha *ej*; Udihe, Oroch *ej*; Orok *er* ~ *eri*; Manchu *ere*; Sibe *er*
(Castrén 1856: 73; Janhunen 1991: 69; Dorji & Banzhibomi 1998: 183b; Chaoke 2014a: 164; Vasilevič 1958: 564a; Cincius 1975/77 2: 460–461; Hauer 1952-1955 1: 255; Zikmundová 2013: 210);
- ii. ‘that’: Nercha Ewenki: Mankovo **tar**, Urulga **tari** ~ **tara**;
Khamnigan Ewenki: Borzya **tari** ~ **tar**;
cf. Solon Ewenki, Orochen *tari*; Siberian Common Ewenki *tar* ~ *tarā* ~ *tari*;
other Northern Tungusic: Lamut *tar*; Negidal *taj*; *Southern Tungusic*: Nanai *tej*; Ulcha *tī*; Udihe *tei*; Oroch *tī*; Orok *tari*; Manchu *tere*; Sibe *tər*
(Castrén 1856: 86; Janhunen 1991: 69; Dorji & Banzhibomi 1998: 666a; Chaoke 2014a: 164; Vasilevič 1958: 387b; Cincius 1975/77 2: 164–165; Hauer 1952-1955 3: 903; Zikmundová 2013: 222);
- o. Interrogatives and quantifiers:
- i. ‘which’: Nercha Ewenki, Khamnigan Ewenki **abgū**;
cf. Solon Ewenki *awu* ‘who’; Siberian Ewenki: Tokma, Zeya, Aldan *abgū*; Podkamennyj, Nepa, North-Baikal, Barguzin, Zeya, Uchir, Urmi, Chumikan, Sakhalin *awgū*; Nepa *awawū*; Urmi *awagū*;

other Northern Tungusic: Lamut *awgida*; Negidal *awwu ~ awgu ~ au*; *Southern Tungusic*: Nanai, Ulcha *hawuj*; Udihe, Oroch *n.a.*; Orok *hāwu*; Manchu, Sibe *ai*

(Castrén 1856: 72; Janhunen 1991: 71; Dorji & Banzhibomi 1998: 48b; Vasilevič 1958: 13b; Cincius 1975/77 1: 4; Hauer 1952-1955 1: 15; Zikmundová 2013: 204; see also Hölzl 2018a: 315–330);

- ii. ‘how many’: Nercha Ewenki: Urulga **adi**, Man’kovo **adi**; Khamnigan Ewenki **adi**;
 cf. Solon Ewenki, Orochen *adi*; Siberian Ewenki: Nepa, Yerbogochen, Upper Lena, North-Baikal, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Sakhalin *adi*;
other Northern Tungusic: Lamut, Negidal *adi*; *Southern Tungusic*: Nanai, Ulcha *hadu*; Udihe, Oroch *adi*; Orok *n.a.*; Manchu *udu*; Sibe *ut*
 (Castrén 1856: 72; Janhunen 1991: 71; Dorji & Banzhibomi 1998: 10a; Chaoke 2014a: 165; Vasilevič 1958: 18a; Cincius 1975/77 1: 14–15; Hauer 1952-1955 3: 944; Zikmundová 2013: 224);
- iii. ‘many’: Nercha Ewenki: Man’kovo, Khamnigan Ewenki **kete**;
 cf. Solon Ewenki *hətə* ‘extremely’; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Tokma, Upper Lena, Tungir, Aldan, Uchur, Chumikan *kete*;
other Tungusic: Nanai *ketu*; Ulcha *kete ~ ketu ~ ket*; Udihe *ketu*; Oroch *ketu*; Orok *ketette* ‘a little bit’; other Tungusic *n.a.*
 (Castrén 1856: 79; Janhunen 1991: 42; Chaoke 2014b: 510; Vasilevič 1958: 231b; Cincius 1975/77 1: 455–456);

p. Verbs:

- i. ‘to find’: Nercha Ewenki, Khamnigan Ewenki **baka-**;
 cf. Solon Ewenki *baha-*; Siberian Common Ewenki *baka-*;
other Northern Tungusic: Lamut *baq-*; Negidal *baha-*; *Southern Tungusic*: Nanai *bā-*; Ulcha *bā- ~ baqa-*; Udihe *b’a-*; Orok *bā- ~ baqqa-*; Oroch *bā-*; Manchu *baha-*; Sibe *n.a.*
 (Castrén 1856: 94; Janhunen 1991: 82; Dorji & Banzhibomi 1998: 53b; Vasilevič 1958: 48a; Hauer 1952-1955 1: 66; Cincius 1975/77 1: 66–67);
- ii. ‘to come’: Nercha Ewenki, Khamnigan Ewenki **eme-**;
 cf. Solon Ewenki *əmə-*; Siberian Common Ewenki *eme-*;

other Northern Tungusic: Lamut *em-*; Negidal *eme-*; *Southern Tungusic*: Nanai *eme-*; Udihe *eme-*; Oroch *emegi-* ‘to return’; Ulcha, Orok, Manchu, Sibe *n.a.*

(Castrén 1856: 73; Janhunen 1991: 66; Dorji & Banzhibomi 1998: 173b; Vasilevič 1958: 558a; Cincius 1975/77 2: 452);

- iii. ‘to step’: Nercha Ewenki, Khamnigan Ewenki *girkü-*; cf. Solon Ewenki *n.a.*; Siberian Ewenki: Podkamennyj, North-Baikal, Barguzin *gerku-*; Remainig dial. *girku-*; *other Northern Tungusic*: Lamut *gerka-*; Negidal *gīhu-*; *Southern Tungusic*: Nanai, Ulcha, Orok *giran-*; Udihe *geæna-*; Oroch *gia-*; Manchu *giru-*; Sibe *guri-* ‘to move’
(Castrén 1856: 82; Janhunen 1991: 25; Vasilevič 1958: 89b; Cincius 1975/77 1: 154–155; Hauer 1952-1955 1: 360; Zikmundová 2013: 212);
- iv. ‘to read’: Nercha Ewenki: Urulga; Khamnigan Ewenki: Borzya *tañir-*; cf. Solon Ewenki *n.a.*; Siberian Common Ewenki *tañ-*; *other Northern Tungusic*: Lamut, Negidal *tañ-*; *Southern Tungusic*: Nanai *taon-*; Ulcha, Orok *taun-*; Udihe, Oroch *taji-*; Manchu, Sibe *n.a.*
(Castrén 1856: 86; Janhunen 1991: 25; Vasilevič 1958: 386a; Cincius 1975/77 2: 161);
- v. ‘to milk’: Nercha Ewenki, Khamnigan Ewenki: Borzya *širi-*; cf. Solon Ewenki *sirə-* ‘to wring out (a towel)’; Siberian Ewenki: Podkamennyj, Nepa, Tokma, Barguzin, Tungir, Aldan, Uchur, Urmi, Chumikan, Ayan, Sakhalin *sir-* ‘to reap’; Yerbogochon, Ilimpeya *hir-*; North-Baikal *širi-*; *other Northern Tungusic*: Lamut *her-*; Negidal *sij-*; *Southern Tungusic*: Nanai *sere-*; Ulcha *siri-*; Udihe, Oroch *sī-*; Orok *sere-*; Manchu *siri-*; Sibe *n.a.*
(Castrén 1856: 84; Janhunen 1991: 80; Chaoke 2017: 126; Vasilevič 1958: 356b; Cincius 1975/77 2: 93; Hauer 1952-1955 3: 806);
- vi. ‘to know’: Nercha Ewenki, Khamnigan Ewenki *sā-*; cf. Solon Ewenki, Orochen *sā-*; Siberian Ewenki: Podkamennyj, Nepa, Tokma, Barguzin, Nercha, Tungir, Zeya, Aldan, Uchur, Urmi, Chumikan, Ayan, Sakhalin *sā-*; Yerbogochon, Ilimpeya *hā-*; Sym, North-Baikal *šā-*; *other Northern Tungusic*: Lamut *ha-*; Negidal *sa-*; *Southern Tungusic*: Nanai, Ulcha, Udihe, Oroch, Orok *sā-*; Manchu, Sibe *sa-*

(Castrén 1856: 84; Janhunen 1991: 79; Dorji & Banzhibomi 1998: 573b; Chaoke 2014a: 169; Vasilevič 1958: 340a; Cincius 1975/77 2: 49–51; Hauer 1952-1955 3: 764; Zikmundová 2013: 220);

- vii. ‘to return’: Nercha Ewenki; Khamnigan Ewenki **mučū-**;
cf. Solon Ewenki *musū-*; Siberian Common Ewenki *mučū-*;
other Northern Tungusic: Lamut *muču-*; Negidal *močo-*; *Southern Tungusic*: Nanai *močogo-*; Ulcha *mučuju-* ~ *mučuji-*; Orok *moto-*, *mutudu-*; Udihe, Oroch, Manchu, Sibe *n.a.*

(Castrén 1856: 96; Janhunen 1991: 30; Dorji & Banzhibomi 1998: 455a; Vasilevič 1958: 266b; Cincius 1975/77 1: 562);

- viii. ‘to wash’: Nercha Ewenki **silki-** ~ *šilki-*; Khamnigan Ewenki **silki-**;
cf. Solon Ewenki *silgi-*; Orochen *šilki-*; Siberian Ewenki: Podkamennyj, Nepa, Tungir, Zeya, Aldan, Uchur, Urmi, Sakhalin *silki-*; Yerbogochon, Ilimpeya *hilki-*;
other Northern Tungusic: Lamut *helka-*; Negidal *silki-*; *Southern Tungusic*: Nanai *silko-*; Ulcha *silču-*; Udihe *sik-*; Oroch *sikki-*; Orok *siltu-*; Manchu *silgi-*

(Castrén 1856: 84; Janhunen 1991: 92; Dorji & Banzhibomi 1998: 615; Chaoke 2014a: 169; Vasilevič 1958: 352b; Cincius 1975/77 2: 84b; Hauer 1952-1955 3: 794);

- ix. ‘to forget’: Nercha Ewenki: Urulga, Khamnigan Ewenki **omŋo-**;
cf. Solon Ewenki *ommo-*; Orochen *omŋo-*; Siberian Ewenki: Nepa, Tokma *ommo-*; Yerbogochon, Ilimpeya, Sym *omgo-*; Remaining dial. *omŋo-*;
other Northern Tungusic: Lamut *omŋa-*; Negidal *omŋo-*; *Southern Tungusic*: Nanai, Orok *omgo-*; Ulcha *oŋbo-*; Udihe *oŋmo-*; Oroch *ommo-*; Manchu *oŋgo-*; Sibe *onŋ-*

(Castrén 1856: 76; Janhunen 1991: 51; Dorji & Banzhibomi 1998: 507a; Chaoke 2014a: 169; Vasilevič 1958: 322b; Cincius 1975/77 2: 17; Hauer 1952-1955 3: 738; Zikmundová 2013: 219).

- (2) The next group includes some Nercha Ewenki words without long vowels in the last syllable, which probably can be explained by the peculiarities of Castrén’s transcriptions from 1856. The Khamnigan Ewenki data are noted down with long vowels as other Ewenki dialects:

- a. ‘son’: Nercha Ewenki **omolgi**; Khamnigan Ewenki **omolgī**;

cf. Solon Ewenki *omol* ‘grandson’ (← Manchu *omolo*); Orochen *omolie*; Siberian Ewenki: Podkamennyj, Yerbogochen, Ilimpeya, North-Baikal, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Chumikan, Ayan, Sakhalin *omolgī*;

other Northern Tungusic: Lamut *omolgo*; Negidal *omolgi*; *Southern Tungusic*: Udihe *omolo*; Oroch *omolī*; Nanai, Ulcha, Orok *n.a.*;
Manchu *omolo*

(Castrén 1856: 76; Janhunen 1991: 40; Dorji & Banzhibomi 1998: 507a; Chaoke 2014a: 161; Vasilevič 1958: 322b; Cincius 1975/77 2: 17b; Hauer 1952-1955 3: 736);

- b. ‘worm’: Nercha Ewenki ***kulikan***; Khamnigan Ewenki ***kulikān***;
cf. Solon Ewenki *kuliharj*; Orochen *kulikan* ‘insect’; Siberian Common Ewenki *kulikān* < *kulin* ‘snake’ + *tkĀn* Ewenki diminutive suffix;
other Northern Tungusic: Lamut *qulin* ~ *quličān* ~ *qolisān* ~ *kuličān* ~ *quličān* ‘mosquito’; Negidal *kolixān* ~ *kulikān* ‘worm, bug’; *Southern Tungusic*: Oroch *kulæ* ‘worm (*common name for worms, snakes, caterpillars*)’; Udihe *kuliga* ‘id.’; Ulcha *qoli* ‘kind of aquatic insect’, *qula* ‘worm’; Orok *qola* ~ *qolia* ~ *qoliya* ‘insect, worm’; Nanai *qolā* ‘worm; caterpillar; insect’; Remaining lgs. *n.a.*
(Castrén 1856: 81; Janhunen 1991: 42; Dorji & Banzhibomi 1998: 402a; Chaoke 2014a: 157; Vasilevič 1958: 218a; Cincius 1975/77 1: 428b);
- c. ‘elk’: Nercha Ewenki ***tōki***; Khamnigan Ewenki: Borzya ***tōki*** ~ ***tōkī***;
cf. Solon Ewenki *tōhi*; Siberian Ewenki: Ilimpeya, Sym, Tokma, North-Baikal Tungir, Zeya, Aldan, Uchur, Urmi, Ayan, Chumikan, Sakhalin *tōkī*; Khingan *towekī*;
other Northern Tungusic: Lamut, Negidal *tōki*; *Southern Tungusic*: Nanai, Ulcha *tō*; Orok *tō* ~ *toyo*; Udihe *n.a.*; Oroch *tōki*; Manchu *toho*; Sibe *n.a.*
(Castrén 1856: 87; Janhunen 1991: 103; Dorji & Banzhibomi 1998: 695b; Vasilevič 1958: 391a; Cincius 1975/77 2: 191–192; Hauer 1952-1955 3: 909);
- d. ‘year’: Nercha Ewenki: Urulga ***aŋaŋi***, Man’kovo ***aŋani***; Khamnigan Ewenki: Borzya, Urulyungui ***aŋjanī***;
cf. Solon Ewenki *annani*; Orochen *aŋjani*; Siberian Common Ewenki *aŋjanī*;
other Northern Tungusic: Lamut *aŋjan*; Negidal *aŋgani*; *Southern Tungusic*: Nanai *ajjaŋa*; Ulcha *aŋan*; Udihe *aŋan(i)*; Oroch *aŋjani*; Orok *anani*; Manchu *aniya*; Sibe *aŋ*

- (Castrén 1856: 71; Janhunen 1991: 51; Dorji & Banzhibomi 1998: 34b; Chaoke 2014a: 154; Vasilevič 1958: 32a; Cincius 1975/77 1: 43–44; Hauer 1952-1955 1: 53; Zikmundová 2013: 204);
- e. ‘autumn’: Nercha Ewenki ***bolóni***; Khamnigan Ewenki ***bolonī***;
 cf. Solon Ewenki *bolonn*; Orochen *bolo*; Siberian Ewenki:
 Podkamennyj, Nepa *bolonī*; Sakhalin, Chumikan *borolī*; Yerbogochon,
 Tungir *bolo*;
other Northern Tungusic: Lamut *bolani*; Negidal *bolonī*; *Southern
 Tungusic*: Nanai, Ulcha, Udihe, Oroch, Orok *bolo*; Manchu, Sibe *boroli*
 (Castrén 1856: 95; Janhunen 1991: 40; Dorji & Banzhibomi 1998: 78;
 Chaoke 2014a: 154; Vasilevič 1958: 60a; Cincius 1975/77 1: 92; Hauer
 1952-1955 1: 110; Zikmundová 2013: 207);
- f. ‘summer’: Nercha Ewenki ***jugani***; Khamnigan Ewenki ***juganī***;
 cf. Solon Ewenki *juγann*; Orochen *jugā*; Siberian Ewenki *juyanī*;
other Northern Tungusic: Lamut *jugani*; Negidal *ḵowani*; *Southern
 Tungusic*: Nanai *joa*; Ulcha, Udihe, Oroch *jua*; Orok *juwa*; Manchu
juwari
 (Castrén 1856: 93; Janhunen 1991: 40; Dorji & Banzhibomi 1998: 380a;
 Chaoke 2014a: 154; Vasilevič 1958: 138b; Cincius 1975/77 1: 268; Hauer
 1952-1955 2: 563);
- g. ‘winter’: Nercha Ewenki ***tügeni***; Khamnigan Ewenki ***tügenī***;
 cf. Solon Ewenki *tüγunn*; Orochen *tuwe*; Siberian Common Ewenki
tügenī;
other Northern Tungusic: Lamut, Negidal *tuweni*; *Southern Tungusic*:
 Nanai, Ulcha, Udihe, Oroch *tue*; Orok *tuwe*; Manchu, Sibe *tuweri*
 (Castrén 1856: 89; Dorji & Banzhibomi 1998: 708b; Chaoke 2014a: 154;
 Vasilevič 1958: 397a; Cincius 1975/77 2: 204b; Hauer 1952-1955 3: 939;
 Stary 1990: 91);
- h. ‘wine’: Nercha Ewenki ***araki***; Khamnigan Ewenki ***arakī***;
 cf. Solon Ewenki *arki*; Orochen *araki*; Siberian Common Ewenki
arakī;
other Northern Tungusic: Lamut *arigi* ← Yakut; Negidal *ayahī*;
Southern Tungusic: Udihe *ayi*; Nanai, Ulcha, Oroch, Orok *araki*;
 Manchu *arki*; Sibe *erk*;
 Tungusic ← Mongolic ← Turkic ← Arabic
 (Castrén 1856: 71; Janhunen 1991: 31; Dorji & Banzhibomi 1998: 43b;
 Chaoke 2014a: 163; Vasilevič 1958: 34a; Cincius 1975/77 1: 48; Hauer
 1952-1955 1: 58; Zikmundová 2013: 209);

- (3) The following Nercha Ewenki example can also be explained by Castrén's transcription, the Khamnigan Ewenki of Manchuria as other Ewenki dialects have a long vowel in the last syllable of word:
- a. 'woman': Nercha Ewenki *āsī*; Khamnigan Ewenki *asī*;
cf. Solon Ewenki *ase*; Orochen *aši*; Siberian Ewenki: Podkamennyj, Nepa, Vitim *asī*; Yerbogochon, Ilimpeya, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Chumikan, Ayan, Sakhalin *ahī*; Sym, North-Baikal, Baunt *aši*;
other Northern Tungusic: Lamut, Negidal *asi*; *Southern Tungusic*: Nanai, Ulcha, Orok *asi*; Udihe *a^hanta*; Oroch *asa*; Manchu *aša* 'elder brother's wife'
(Castrén 1856: 72; Janhunen 1991: 45; Dorji & Banzhibomi 1998: 45a; Chaoke 2014a: 160; Vasilevič 1958: 38a; Cincius 1975/77 1: 55; Hauer 1952-1955 1: 61);
- (4) The change of Tungusic vowel **u* > *o* in Nercha Ewenki:
- a. 'five': Nercha Ewenki *toŋa*, Khamnigan Ewenki *tunŋa* ~ *tunna*;
cf. Solon Ewenki *tunŋa*; Orochen *tunŋa*; Siberian Ewenki: Ilimpeya, Tokma, Nercha, Zeya, Aldan *tonŋa*; Remaining dial. *tunŋa*;
other Northern Tungusic: Lamut *tuńŋan*; Negidal *tońŋa*; *Southern Tungusic*: Nanai *tojŋa*; Ulcha *tunŋa*; Udihe, Oroch *tunŋa*; Orok *tunda*; Manchu, Sibe *sunŋa*
(Castrén 1856: 88; Janhunen 1991: 76; Dorji & Banzhibomi 1998: 703; Chaoke 2014a: 170; Vasilevič 1958: 401b; Cincius 1975/77 2: 214; Hauer 1952-1955 3: 830; Zikmundová 2013: 221);
- (5) The palatalization of vowels in Nercha and Khamnigan Ewenki:
- a. 'milk': Nercha Ewenki *ükümńi*; Khamnigan Ewenki *ükün*;
cf. Solon Ewenki *uhun* ~ *akun*; Orochen *ukun*; Siberian Ewenki: Podkamennyj, Ilimpeya, Tokma *ukunmī*; Nepa, Yerbogochon, Upper Lena, North-Baikal, Barguzin, Tungir, Aldan, Ayan *ukumnī*;
other Northern Tungusic: Lamut *ukeń* ~ *ukuń*; Negidal *ukuńi*; *Southern Tungusic*: Nanai *ukuń*; Ulcha *kuen*; Udihe *košo*; Oroch *okon*; Orok *kō(n)* ~ *kū(n)*; Manchu, Sibe *n.a.*
(Castrén 1856: 77; Janhunen 1991: 24; Chaoke 2014b: 184; Chaoke 2014a: 163; Vasilevič 1958: 435a; Cincius 1975/77 2: 255);

- b. ‘to die’: Nercha Ewenki *bü-*; Khamnigan Ewenki *büde-*;
 cf. Solon Ewenki, Orochen, Siberian Common Ewenki *bu-*;
other Northern Tungusic: Lamut ; Negidal *bu-*; *Southern Tungusic*:
 Nanai *bu-* ~ *buj-*; Orok, Ulcha *bu-* ~ *bul-*; Udihe *bude-*; Oroch *bude-* ~
bukki-; Manchu *bude-* ~ *buče-*; Sibe *beš-* ~ *bəč(ə)-*
 (Castrén 1856: 96; Janhunen 1991: 79; Dorji & Banzhibomi 1998: 95;
 Chaoke 2014a: 169; Vasilevič 1958: 62b; Cincius 1975/77 1: 98–99;
 Hauer 1952-1955 1: 117; Zikmundová 2013: 206);
- (6) The change of original Tungusic nasal consonant **ŋ*- > *n*- in Nercha and
 Khamnigan Ewenki as in Solon Ewenki:
- a. ‘hand’: Nercha Ewenki *nāla* ~ *nala*; Khamnigan Ewenki *nāla*;
 cf. Solon Ewenki *nāl*; Orochen *ŋāla*; Siberian Ewenki: Yerbogochen,
 Ayan *nāle*; Remaining all dialects *ŋāle*;
other Northern Tungusic: Lamut *ŋāl*; Negidal *ŋāla* ~ *ŋala*; *Southern*
Tungusic: Nanai, Ulcha, Udihe *ŋala*; Orok *ŋāla*; Oroch *ŋāla* ~ *ŋala*;
 Manchu *gala*; Sibe *gal*
 (Castrén 1856: 85, 83; Janhunen 1991: 49; Dorji & Banzhibomi 1998:
 467b; Chaoke 2014a: 159; Vasilevič 1958: 278b; Cincius 1975/77 1: 656;
 Hauer 1952-1955 1: 331; Zikmundová 2013: 213);
- b. ‘who?’: Nercha Ewenki *nī* ~ *nīi*; Khamnigan Ewenki *nī*;
 cf. Solon Ewenki *ni* ‘which one’; Orochen *ni*; Siberian Ewenki:
 Podkamennyj, Yerbogochen, Ilimpeya, Sym, Tungir, Zeya, Uchur,
 Urmi, Chumikan, Sakhalin *ŋī*; Upper Lena, North-Baikal, Barguzin,
 Chumikan *nī*;
other Northern Tungusic: Lamut, Negidal *nī* ~ *ŋī*; *Southern Tungusic*:
 Nanai *uj* ~ *ui*; Ulcha *ŋui* ~ *uj*; Orok *ŋuji* ~ *ŋuj* ~ *ŋui*; Udihe *nī*; Oroch *nīi*;
 Manchu, Sibe *n.a.*
 (Castrén 1856: 85; Janhunen 1991: 29; Chaoke 2014b: 340; Chaoke
 2014a: 165; Vasilevič 1958: 280a; Cincius 1975/77 1: 660).

5.2 Phonetic differences between Nercha and Khamnigan Ewenki

There are some phonetic differences between Nercha and Khamnigan Ewenki dialects, which are possibly results of change in Khamnigan Ewenki in the 20th and 21st centuries.

- (7) The deletion of Tungusic initial consonant **h*- (< Proto Tungusic **p*-) in
 Khamnigan Ewenki. In some cases *h*- is sporadically preserved in the

Borzya subdialect of Khamnigan Ewenki (Janhunen 1991: 46). The consonant *h-* was already deleted in the extinct Mankovo subdialect, while we find it in the extinct Urulga and do not see it in Urulyungui, the related subdialect of Manchuria. The initial *h-* disappeared in Solon Ewenki and Orochen too:

- a. ‘daughter, young girl’: Nercha Ewenki: Urulga **hunāt**; Man’kovo **unāt**; Khamnigan Ewenki: Borzya **hunāji**,¹⁵ Urulyungui **unād**; cf. Solon Ewenki **unāji**; Orochen **unāji**; Siberian Ewenki: Podkamennyj, Nepa, Ilimpeya, Barguzin, Zeya, Uchur, Urmi, Chumikan, Sakhalin **hunāt**; Tokma **honāt**; Yerbogochen, Tungir, Aldan **sunāt**;
other Northern Tungusic: Lamut **hunāj** ~ **hunāt**; Negidal **honāt**;
Southern Tungusic: Nanai **pondažo**; Ulcha **pundažu**; Udihe, Oroch **hunāji**; Orok **pundado** ~ **pundadu**; Manchu, Sibe *n.a.*
(Castrén 1856: 77, 83; Janhunen 1991: 43; Dorji & Banzhibomi 1998: 529b; Chaoko 2014a: 161; Vasilevič 1958: 495b; Cincius 1975/77 2: 347);
- b. ‘blanket’: Nercha Ewenki: Urulga **hulda**, Man’kovo **ulda**; Khamnigan Ewenki: Borzya **(h)ulda**; Urulyungui **ulda**;
cf. Solon Ewenki **ulda**; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Tokma, Barguzin, Tungir, Zeya, Aldan, Urmi, Chumikan, Sakhalin **hulla**; Ilimpeya, Uchur, Chumikan **hulda**; Ilimpeya, North-Baikal, Ayan **ulda**;
other Northern Tungusic: Lamut **hulra**; Negidal **hola**; *Southern Tungusic*: Nanai **polta**; Orok, Ulcha **pulta**; Udihe **hulaha**; Oroch **hukta**; Manchu, Sibe *n.a.*
(Castrén 1856: 76; Janhunen 1991: 52; Dorji & Banzhibomi 1998: 526a; Vasilevič 1958: 493b; Cincius 1975/77 2: 345);
- c. ‘road’: Nercha Ewenki: Urulga **hokto**, Man’kovo **okto**; Khamnigan Ewenki: Borzya **(h)ogto**;
cf. Solon Ewenki **otto**; Orochen **okto**; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Sym, Tokma, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Ayan, Sakhalin **hokto**; Aldan **sokto**; Upper Lena, North-Baikal, Chumikan **okto**;
other Northern Tungusic: Lamut **hōt**; Negidal **hokto**; *Southern Tungusic*: Nanai, Ulcha, Orok **pokto**, Udihe, Oroch **hokto**; Manchu **oktoron** ‘hare tracks’

¹⁵The Borzya form *hunāji* ‘daughter, young girl’ is a hybrid word: the preservation of initial *h-* is a heritage, while the rest is close to Solon Ewenki *unāji*.

(Castrén 1856: 83; Janhunen 1991: 47; Dorji & Banzhibomi 1998: 515a; Chaoke 2014a: 163; Vasilevič 1958: 484a; Cincius 1975/77 2: 331; Hauer 1952-1955 3: 732);

- d. ‘red’: Nercha Ewenki **ularin**; Khamnigan Ewenki: Borzya **hularin**, Urulyungui **ularin**;

cf. Solon Ewenki *ularin*; Orochen *ularin*; Siberian Ewenki: Podkamennyj, Tokma, Tungir, Vitim, Zeya, Uchur, Urmi, Chumikan, Ayan, Sakhalin *hularin*; Upper Lena, North-Baikal, Vitim *ularin*; *other Northern Tungusic*: Lamut *hulaña*; Negidal *holajin*; *Southern Tungusic*: Nanai *folgæ ~ forgæ*; Ulcha, Udihe *hulaligi*; Oroch, Orok *n.a.*; Manchu *fulgian*; Sibe *fulavũⁿ*

(Castrén 1856: 76; Janhunen 1991: 46; Dorji & Banzhibomi 1998: 525; Chaoke 2014a: 167; Vasilevič 1958: 493a; Cincius 1975/77 2: 343; Hauer 1952-1955 1: 314; Zikmundová 2013: 211);

- e. ‘big’: Nercha Ewenki: Urulga **hegdi**, Man’kovo **egdi**; Khamnigan Ewenki: Borzya **hegdi**, Urulyungui **egdi**;

cf. Solon Ewenki *əgdə ~ ədduy*; Orochen *egdeŋe*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochon, Ilimpeya, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Chumikan, Sakhalin *hegdi*; *other Northern Tungusic*: Lamut *egde*; Negidal *egdi*; *Southern Tungusic*: Nanai, Orok *egji*; Ulcha, Udihe, Oroch *egdi*; Manchu, Sibe *n.a.*

(Castrén 1856: 82, 72; Janhunen 1991: 47; Dorji & Banzhibomi 1998: 150b; Chaoke 2014a: 165; Vasilevič 1958: 503b; Cincius 1975/77 2: 359);

- (8) The fluctuation of initial vowel *u- ~ i-* in Nercha and Khamnigan Ewenki dialects:

- a. ‘to drink’: Nercha Ewenki: Urulga **umi-**, Man’kovo **imi-**; Khamnigan Ewenki: Borzya **imi-**, Urulyungui **um-**;

cf. Solon Ewenki *omi- ~ imo-*; Orochen *imo-*; Siberian Common Ewenki *um-*; *other Northern Tungusic*: Lamut *n.a.*; Negidal *om-*; *Southern Tungusic*: Nanai *omi-*; Orok, Ulcha, Udihe *umi-*; Oroch *imi-*; Manchu *omi-*; Sibe *emi-*

(Castrén 1856: 77; Janhunen 1991: 98; Chaoke 2014b: 392; Chaoke 2014a: 167; Vasilevič 1958: 441b; Cincius 1975/77 2: 266; Hauer 1952-1955 3: 735; Zikmundová 2013: 209);

- (9) The diphthongization of long vowel $\bar{i} > ie$ in the Urulyungui subdialect of Khamnigan
- a. ‘eye’: Nercha Ewenki: Urulga *yēsa*, Man’kovo *yīsa*; Khamnigan Ewenki: Borzya *yīsa*, Urulyungui *iesa*;
cf. Solon Ewenki *īsa*; Orochen *yiesa*; Siberian Ewenki: Podkamennyj, Nepa *ēsa*; Yerbogochen, Ilimpeya, Upper Lena, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Ayan, Chumikan, Sakhalin *ēha*; Sym, North-Baikal *ēša*;
other Northern Tungusic: Lamut *yasal*; Negidal *ēsa*; *Southern Tungusic*: Nanai *nasal*; Ulcha *isal*; Udihe *yeha*; Oroch *isa*; Orok *īsa*; Manchu *yasa*; Sibe *yas*
(Castrén 1856: 75; Janhunén 1991: 34; Dorji & Banzhibomi 1998: 320a; Chaoke 2014a: 159; Vasilevič 1958: 567b; Cincius 1975/77 1: 291b-292b; Hauer 1952-1955 3: 1015; Zikmundová 2013: 225);
- b. ‘ear’: Nercha Ewenki: Urulga *šen*; Khamnigan Ewenki: Borzya *šin*, Urulyungui *sien*;
cf. Solon Ewenki *səŋ*; Orochen *šien*; Siberian Ewenki: Podkamennyj, Ayan, Aldan, Barguzin, Upper Lena, Zeya, Nepa, Sakhalin, Tokma, Tungir, Urmi, Uchur, Chumikan *sēn*; Yerbogochen, Ilimpeya *hēn*; Sym, North-Baikal *šēn*;
other Northern Tungusic: Lamut, Negidal *sen*; *Southern Tungusic*: Nanai *siā*; Ulcha, Orok *sēn*; Udihe *n.a.*; Oroch *sē*; Manchu *šan*; Sibe *saⁿ*
(Castrén 1856: 84; Janhunén 1991: 34; Dorji & Banzhibomi 1998: 588a; Chaoke 2014a: 159; Vasilevič 1958: 347b; Cincius 1975/77 2: 70b-71b; Hauer 1952-1955 3: 843; Zikmundová 2013: 220);
- c. ‘moon’: Nercha Ewenki *bēga*; Khamnigan Ewenki: Borzya *bīga*, Urulyungui *biega*;
cf. Solon Ewenki *bēya*; Orochen *biēga*; Siberian Ewenki: Nepa, Tokma *bēwa*; Remaining dial. *bēga*;
other Northern Tungusic: Lamut *beg*; Negidal *bega*; *Southern Tungusic*: Nanai *bia*; Ulcha, Orok *bē*; Udihe *beæ*; Oroch *bæ*; Manchu *biya*
(Castrén 1856: 95; Janhunén 1991: 34; Dorji & Banzhibomi 1998: 70a; Chaoke 2014a: 152; Vasilevič 1958: 52b; Cincius 1975/77 1: 78; Hauer 1952-1955 1: 100);
- d. ‘what’: Nercha Ewenki *ēkun*, Mankovo *īkun*; Khamnigan Ewenki: Borzya *ikun*, Urulyungui *iekun*;

cf. Solon Ewenki *aγu*; Orochen *ikun*; Siberian Ewenki: Yerbogochen, Ilimpeya *ikūn*; Upper Lena *yākūn*; Remaining dial. *ēkūn*;
other Northern Tungusic: Lamut *ǣk*; Negidal *ēhun*; *Southern Tungusic*:
 Udihe *y'eu*; Oroch *yaw*; Orok, Nanai, Ulcha *n.a.*; Manchu *ya*
 (Castrén 1856: 73; Janhunen 1991: 34; Dorji & Banzhibomi 1998: 154b;
 Chaoke 2014a: 165; Vasilevič 1958: 551b; Cincius 1975/77 1: 286–287;
 Hauer 1952-1955 3: 1002);

- (10) The development of a secondary long vowel from *VgV* in Urulyungui subdialect of Khamnigan Ewenki and in Orochen:
- a. ‘to sit’: Nercha Ewenki *tege*-, Khamnigan Ewenki: Urulyungui *tē*-, Borzja *tege*-;
 cf. Solon Ewenki *təγ*-; Orochen *tē*-; Siberian Ewenki: Tokma, North-Baikal *tē*-; Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Sym, Barguzin, Tungir, Zeya, Aldan, Uchur, Urmi, Ayan, Sakhalin *teye*-;
other Northern Tungusic: Lamut *teg*-; Negidal *tege*-; *Southern Tungusic*:
 Nanai, Ulcha, Udihe, Oroch, Orok, Manchu, Sibe *te*-
 (Castrén 1856: 87; Janhunen 1991: 99; Dorji & Banzhibomi 1998: 674a;
 Chaoke 2014a: 168; Vasilevič 1958: 418a; Cincius 1975/77 2: 226–228;
 Hauer 1952-1955 3: 899; Zikmundová 2013: 222);
- (11) The voicing of consonant clusters **rk*, **ks*, **kt* in Khamnigan Ewenki:
- a. ‘fist’: Nercha Ewenki *nurka*; Khamnigan Ewenki *nurga*;
 cf. Solon Ewenki, Siberian Ewenki *n.a.*;
other Northern Tungusic: Negidal *nelga* ~ *nojga*; *Southern Tungusic*:
 Ulcha *ńugja*, Oroch *nugga*; Manchu *nujan*; Remaining lgs. *n.a.*
 (Castrén 1856: 86; Janhunen 1991: 23; Cincius 1975/77 1: 590a; Hauer 1952-1955 3: 722);
 Tungusic ← Mongolic: Middle Mongol: MNT *nodurqa*; HY, Muq. *nudurqa*; Literary Mongolian *nidurya*; Onon Khamnigan Mongol *nidurga*, cf. *nyudarga* (← Buryat); Dadal-sum Khamnigan *nidurya*;
 Dagur *nyɔdruγw*; Buryat *nyudarga*;
 Mongolic ← Turkic **ńidru-*: cf. Old Turkic *yīðruq* ‘fist’ (Khabtagaeva 2017: 120);
- b. ‘evening’: Nercha Ewenki *śikśe*; Khamnigan Ewenki *sigsenī*;
 Siberian Ewenki: Uchur, Urmi, Ayan, Chumikan, Sakhalin *sikse* ‘in the evening’;

cf. Solon Ewenki *dolbon*; Orochen *dolbo*; Nercha Ewenki *dolboni* ‘night’;

other Northern Tungusic: Lamut *hisečin*; Negidal *sikse*; *Southern Tungusic*: Udihe *sikie*; Nanai, Ulcha, Oroch *sikse*; Orok *šekše*; Manchu *sikse* ‘yesterday’; Sibe *čaksə*

(Castrén 1856: 84; Janhunen 1991: 45; Dorji & Banzhibomi 1998: 133a; Chaoke 2014a: 155; Vasilevič 1958: 351a; Cincius 1975/77 2: 81; Hauer 1952-1955 3: 793; Zikmundová 2013: 207);

- c. ‘to meet’: Nercha Ewenki *uktu-*; Khamnigan Ewenki ***ugtu-***;

cf. Solon Ewenki *otto-*; Siberian Ewenki *n.a.*;

other Northern Tungusic: Negidal *oktul-*; *Southern Tungusic*: Ulcha *oktoli-* ~ *uktuli-*, Oroch *uktu-*; Orok *uktulli-*; Remaining lgs. *n.a.*

(Castrén 1856: 76; Janhunen 1991: 26; Dorji & Banzhibomi 1998: 515b; Cincius 1975/77 2: 254b)

← Mongolic: Middle Mongol: MNT *uqtu-* ~ *uqdu-*; HY *uqtu-* ~ *uytu-*; Literary Mongolian *uytu-* ‘to greet, to meet, to welcome’; Onon Khamnigan Mongol *ugta-*

(← Buryat); Dagur *ort-*; Buryat *ugta-* (Khabtagaeva 2017: 139);

- (12) The additional sound at the end of Borzya subdialect of Khamnigan Ewenki as in Orochen or Solon, which proves their influence:

- a. ‘head’: Nercha Ewenki *dil*; Khamnigan Ewenki: Urulyungui *dil*, Borzya ***dili***;

cf. Solon Ewenki *del*; Orochen *dili*; Siberian Common Ewenki *dil*;

other Northern Tungusic: Lamut, Negidal *dil*; *Southern Tungusic*: Nanai, Orok *jili*; Ulcha, Udihe, Oroch *dili*; Manchu, Sibe *n.a.*

(Castrén 1856: 90; Janhunen 1991: 27; Dorji & Banzhibomi 1998: 127b; Chaoke 2014a: 158; Vasilevič 1958: 128b; Cincius 1975/77 1: 205b-206a, see also Hölzl 2018b: 129 for some discussion);

- b. ‘house’: Nercha Ewenki *jū*, Khamnigan Ewenki: Urulyungui *jū*, Borzya ***jūg***;

cf. Solon Ewenki *ju*, cf. *jūγ* (Ivanovskij); Orochen *jū*; Siberian Common Ewenki *jū*;

other Northern Tungusic: Lamut *jū*; Negidal *jō*; *Southern Tungusic*: Nanai *jōg*; Ulcha *jūg*; Udihe *jugdi*; Oroch *jug*; Orok *duku*; Manchu, Sibe *n.a.*

(Castrén 1856: 94; Janhunen 1991: 41; Dorji & Banzhibomi 1998: 385a; Chaoke 2014a: 162; Vasilevič 1958: 138a; Cincius 1975/77 1: 266);

- (13) In some cases the initial **ŋ*- through *n*- is further deleted in Nercha and Khamnigan Ewenki as in Southern Tungusic languages:
- a. ‘dog’: Nercha Ewenki: Man’kovo ***inakin***; Khamnigan Ewenki: Urulyungui ***inakin***; Borzya *ninakin* (← Solon);
 cf. Solon Ewenki *ninihiŋ*; Orochen *ŋanakin*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochon, Sym, Tungir, Urmi, Sakhalin *ŋinakin*; Upper Lena *ninakin*;
other Northern Tungusic: Lamut *ŋin* ~ *ŋen*; Negidal *ninahin* ~ *ŋinahin*;
Southern Tungusic: Nanai *inda*; Ulcha *iŋda*; Udihe *inai*; Oroch *inaki*;
 Orok *nina* ~ *ŋina*; Manchu *indahûn*; Sibe *yindahûn*
 (Castrén 1856: 74; Janhunen 1991: 49; Dorji & Banzhibomi 1998: 491a; Chaoke 2014a: 158; Vasilevič 1958: 280b; Cincius 1975/77 1: 661; Hauer 1952-1955 2: 498; Zikmundová 2013: 225);
- b. ‘to go’: Nercha Ewenki: Man’kovo *nene-* ~ *ŋene-*; Khamnigan Ewenki: Borzya *nene-*, Urulyungui ***ene-***;
 cf. Solon Ewenki *nənə-*; Siberian Ewenki: Tokma, Tungir *gene-*;
 Remaining dial. *ŋene-*;
other Northern Tungusic: Lamut *ŋen-*; Negidal *ŋene-* ~ *gene-*; *Southern Tungusic*: Nanai *ene-*; Ulcha, Udihe, Oroch, Orok *ŋene-*; Manchu *gene-*;
 Sibe *gən(ə)-*
 (Castrén 1856: 85; Janhunen 1991: 85; Dorji & Banzhibomi 1998: 483b; Vasilevič 1958: 284a; Cincius 1975/77 1: 669–671; Hauer 1952-1955 1: 344; Zikmundová 2013: 212);
- (14) The change of the consonant *VŋV* > *VgV* in Khamnigan Ewenki:
- a. ‘he/she’: Nercha Ewenki *nunən*; Khamnigan Ewenki ***nugan***;
 cf. Solon Ewenki *nuganŋ*; Orochen *nugan*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochon, Ilimpeya, Tungir, Zeya, Aldan, Uchur, Urmi, Sakhalin *nunən*; Tokma, Upper Lena *noan*; North-Baikal *noan* ~ *nuan*; Sym *nugan*;
other Northern Tungusic: Lamut *noŋen*; Negidal *noŋan*; *Southern Tungusic*: Nanai *noani*; Ulcha *nān*, *nāni* ~ *nōni*; Udihe *nuan*; Oroch *nuañ* ~ *nuañi* ~ *nunəñi*; Orok *nōni*; Manchu, Sibe *n.a.*
 (Castrén 1856: 86; Janhunen 1991: 49; Chaoke 2014b: 338; Chaoke 2014a: 164; Vasilevič 1958: 299b; Cincius 1975/77 1: 611);

- (15) The change of $VwV > VbV$ in Khamnigan Ewenki:
- a. ‘to catch’: Nercha Ewenki *ǰawa-*; Khamnigan Ewenki **ǰaba-**;
cf. Solon Ewenki, Orochen *ǰawa-*; Siberian Ewenki: Ilimpeya, Uchur
ǰaba-; Remaining dial. *ǰawa-*;
other Northern Tungusic: Lamut *ǰaw-*; Negidal *ǰawa-*; *Southern
Tungusic*: Nanai, Ulcha *ǰapa-*; Udihe, Oroch *ǰawa-*; Orok *dapa-* ~
dappa-; Manchu *ǰafa-*; Sibe *ǰaf-* ~ *ǰavə-*
(Castrén 1856: 93; Janhunén 1991: 90; Dorji & Banzhibomi 1998: 361b;
Chaoke 2014a: 168; Vasilevič 1958: 145b; Cincius 1975/77 1: 240–241;
Hauer 1952-1955 2: 510; Zikmundová 2013: 215);
- (16) The reduplication of the consonant $-ŋ-$ in Khamnigan Ewenki:
- a. ‘tongue’: Nercha Ewenki *iŋi*; Khamnigan Ewenki ***iŋŋi***;
cf. Solon Ewenki *iŋi*; Siberian Ewenki: Podkamennyj, Tokma,
North-Baikal, Barguzin, Tungir, Zeya, Sakhalin *inni*; Barguzin,
North-Baikal, Chumikan, Zeya, Aldan *iŋi*; Chumikan, Ayan *iŋi*;
other Northern Tungusic: Lamut ⁱ*enŋe*; Negidal *iŋni* ~ *iŋŋi*; *Southern
Tungusic*: Nanai *siŋmu* ~ *sirmu*; Ulcha *siŋu*; Orok *sinu*; Udihe *iŋi*;
Oroch *iŋi* ~ *iŋŋi*; Manchu *ilŋgu*
(Castrén 1856: 74; Janhunén 1991: 52; Janhunén 1991: ; Dorji &
Banzhibomi 1998: 334b; Vasilevič 1958: 174a; Cincius 1975/77 1: 316;
Hauer 1952-1955 2: 492).

5.3 Solon Ewenki influence on Khamnigan Ewenki

There are many words in Khamnigan Ewenki which were borrowed in Manchuria from Solon Ewenki. In most cases the Solon influence is observable in the Borzya subdialect.

- (17) The change of affricate $*č > t$ as in Solon Ewenki:
- a. ‘Russian’: Nercha Ewenki, Khamnigan Ewenki (Urulyungui) *lūča*; cf.
Khamnigan Ewenki (Borzya) ***lūta***
← Solon Ewenki *lūt*; cf. Siberian Ewenki: Zeya *lōča*; Ayan, Aldan,
May, Tommot, Uchur *nūča*; Remaining dial. *lūča*;
other Northern Tungusic: Lamut *nūči*; Negidal *lōča*; *Southern Tungusic*:
Nanai *loča*; Ulcha *luča* ~ *nuča*; Udihe *lusa*; Oroch *luča*; Orok *lūt’a* ~
luča; Manchu *loča* ‘demon, devil’; Sibe *n.a.*
(Castrén 1856: 84; Janhunén 1991: 98; Dorji & Banzhibomi 1998: 417a;
Vasilevič 1958: 242a; Cincius 1975/77 1: 513b; Hauer 1952-1955 2: 626);

- (18) The appearance of words with a final unstable consonant $-n^{16}$ as in Solon Ewenki and Orochen:
- a. ‘father’: Nercha Ewenki: Urulga *ama*, Man’kovo *amā*; cf. Khamnigan Ewenki ***amin***
 ← Solon Ewenki *amiŋ*; Orochen *amin*; cf. Siberian Common Ewenki *amā*;
other Northern Tungusic: Lamut *amā*; Negidal *amaj*; *Southern Tungusic*: Nanai, Ulcha, Udihe *amin*; Oroch *ama*; Orok *ama ~ amma*; Manchu *ama*; Sibe *amə*
 (Castrén 1856: 72; Janhunen 1991: 23; Dorji & Banzhibomi 1998: 27a; Chaoko 2014a: 160; Vasilevič 1958: 26a; Cincius 1975/77 1: 34b; Hauer 1952-1955 1: 39; Zikmundová 2013: 204);
- b. ‘carriage’: Nercha Ewenki *terge*; cf. Khamnigan Ewenki ***tergēn***
 ← Solon Ewenki *təyunn*; Orochen *tergen*; cf. Siberian Ewenki: Barguzin, Ayan *terge*; Manchu, Sibe *sejen*,¹⁷ *other Tungusic*: *n.a.*
 (Castrén 1856: 87; Janhunen 1991: 100; Dorji & Banzhibomi 1998: 676b; Chaoko 2014a: 163; Vasilevič 1958: 424a; Cincius 1975/77 2: 238; Hauer 1952-1955 3: 776; Stary 1990: 76);
Tungusic ← Mongolic: Middle Mongol: MNT *terge(n)*; HY, ‘Phags-pa, Muq. *tergen*; Literary Mongolian *terge* ‘vehicle; cart, wagon, carriage; car; rook (*in chess*)’; Manchurian and Onon Khamnigan Mongol *terge*; Dagur *təɾəy*; Buryat *terge* (Khabtagaeva 2017: 134; see also Doerfer 1985: 104);
- (19) The change of vowel $*o > u$ as in Solon Ewenki:
- a. ‘fingernail’: Nercha Ewenki *osikta*; Khamnigan Ewenki (Urulyungui) *osigta*; cf. Khamnigan Ewenki (Borzya) ***usigta***
 ← $*usikta$: Solon Ewenki *usitt*; cf. Siberian Ewenki: Podkamennyj, Nepa *osikta*; Yerbogochon, Tungir, Zeya, Aldan, Uchir, Urmi, Sakhalin *ohikta*; Sym, North-Baikal *ošikta*; Ayan *ōtta*;

¹⁶Vasilevič and Cincius indicate the final $-n$ as a possessive suffix (Vasilevič 1958: 26a; Cincius 1975/77 1: 34b).

¹⁷Hözl (p.c., 2020) drew my attention to the relationship of Manchu and Sibe forms with the Mongolic word. According to Norman (1977), a cognate $-rg-$ regularly yields to $-j-$ in Manchu, e.g. Proto-Tungusic $*bargilā$, cf. Ewenki *bargilā* ~ Manchu *bajila* ‘on the other side’, Proto-Tungusic *herga(kta)*, cf. Ewenki *irgakta* ~ Manchu *ija* ‘gadfly’; Proto-Tungusic $*tunja$. There are also some Mongolic loanwords as *songgo-* ‘to choose’ → Manchu *sonjo-* ‘id.’; Mongolic *tergen* ‘cart, vehicle’ → Manchu *sejen*; Mongolic *toryan* ‘silk’ → Manchu *suje* (the last two Mongolic loanwords are examined in this paper).

other Northern Tungusic: Lamut *oste*; Negidal *ōtta*; *Southern Tungusic*: Ulcha *husta*; Udihe *waikta*; Oroch *hosi-* ‘to scratch’; Nanai, Orok *hosikta*; Manchu *usiha*; Sibe *ušiḡa ~ ušḡa*
(Castrén 1856: 76; Janhunen 1991: 98; Dorji & Banzhibomi 1998: 535a; Vasilevič 1958: 328b; Cincius 1975/77 2: 26b; Hauer 1952-1955 3: 973; Zikmundová 2013: 224);

- (20) The insertion of long vowels in the first syllable of words as in Solon Ewenki:

- a. ‘when’: Nercha Ewenki *ālī*; cf. Khamnigan Ewenki *āli*
← Solon Ewenki *ālī*; cf. Siberian Ewenki: Vitim, Aldan, Uchir *ālī*;
other Northern Tungusic: Lamut *ālik* ‘once upon a time’; Negidal *ālī*;
Southern Tungusic: Nanai, Ulcha, Orok *hālī*; Udihe *ali*; Oroch *ālī*;
Manchu, Sibe *n.a.*
(Castrén 1856: 71; Janhunen 1991: 32; Dorji & Banzhibomi 1998: 2b; Vasilevič 1958: 25a; Cincius 1975/77 1: 32);

The loanwords from Khamnigan Mongol also belong here, the long vowel is possibly inserted under Solon Ewenki influence (Janhunen 1991: 100):

- (21) ‘medicine, drug’: Nercha Ewenki *n.a.*; Khamnigan Ewenki: Urulyungui *n.a.*; Borzya *ēm*;
cf. Solon Ewenki *əŋ*; other Tungusic lgs. *n.a.* (Janhunen 1991: 100; Dorji & Banzhibomi 1998: 156a);
Tungusic ← Mongolic: Middle Mongol: HY, Muq. *em*; Literary Mongolian *em*; Manchurian and Onon Khamnigan Mongol, Buryat *em*; Dagur *ə m*;
Mongolic ← Turkic: cf. Old Turkic *ām* ‘remedy’;
- (22) ‘silk’: Nercha Ewenki *n.a.*; Khamnigan Ewenki *tōrga*;
cf. Solon Ewenki *tōyo*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimpeya, Tokma, Upper Lena, Zeya, Uchur, Urmi, Chumikan, Sakhalin *tōrgā*; Manchu *suje*; other Tungusic lgs.: *n.a.*
(Janhunen 1991: 100; Dorji & Banzhibomi 1998: 695a; Cincius 1975/77 2: 199b; Hauer 1952-1955 3: 824);
← Mongolic: Middle Mongol: MNT *torqan*; ZY *turya*; Muq. *torqa*;
Literary Mongolian *toryan*; Manchurian Khamnigan Mongol *torgo*; Onon Khamnigan Mongol *torgo(n)*; Dagur *tory^w*; Buryat *torgo(n)*
(Khabtagaeva 2017: 135; see also Doerfer 1985: 94);

- (23) ‘marten’: Nercha Ewenki *n.a.*; Khamnigan Ewenki *sōlugī*
 ← Solon Ewenki *sōlgə ~ sōlhə*; cf. Siberian Ewenki: Podkamennyj, Baunt, Barguzin, North-Baikal, Sakhalin, Tokko, Tommot, Tungir, Urmi *solongō*; Sakhalin *solga*; Yerbogochen *honongo*;
other Tungusic: Nanai *sol’u*; Udihe *selue*; Manchu *solohi*; Remaining lgs. *n.a.*
 (Janhunen 1991: 100; Dorji & Banzhibomi 1998: 632a; Vasilevič 1958: 362b; Cincius 1975/77 2: 109a; Hauer 1952-1955 3: 813);
 ← Mongolic: Middle Mongol: MNT *solangqa*; Literary Mongolian *solongya*; Onon Khamnigan Mongol *solongo*; Dagur *n.a.*; Buryat *holongo* (Khabtagaeva 2017: 128; see also Doerfer 1985: 39–40; Rozycki 1994: 187);
- (24) Solon Ewenki lexical items:
- a. ‘sun’:
- i. Khamnigan Ewenki (Borzya) *sigün*
 ← Solon Ewenki *sigun*; cf. Siberian Ewenki: Zeya, Aldan, Khingan, Uchur *sigun*; *other Northern Tungusic*: Negidal *siwun ~ sigun*; *Southern Tungusic*: Nanai *siu*; Ulcha *siun ~ sun*; Udihe *sün*; Oroch *seun*; Orok *šun*; Manchu *šun*; Sibe *šuⁿ*;
 - ii. Nercha Ewenki *dilacā*; Khamnigan Ewenki (Urulyungui) *dilacā ~ gilacā*;
 cf. Orochen *diliča*; Siberian Ewenki: Podkamennyj, Nepa, Yerbogochen, Ilimepya, Upper Lena, North-Baikal, Tungir, Aldan, Uchur, Urmi, Chumikan, Sakhalin *dilačā*; *other Tungusic lgs*: Lamut *dilača*
 (Castrén 1856: 90; Janhunen 1991: 98; Dorji & Banzhibomi 1998: 607b; Chaoke 2014a: 152; Vasilevič 1958: 350b, 128a; Cincius 1975/77 2: 78; 1: 206a; Hauer 1952-1955 3: 867; Zikmundová 2013: 222);
- b. ‘new’: Nercha Ewenki *n.a.*; Khamnigan Ewenki *irkekīn*
 ← Solon Ewenki *irkəhin ~ ikkiŋ*; Orochen *irkin*; cf. Siberian Ewenki: Sakhalin *irkekīn* ‘new, fresh’;
other Northern Tungusic: Lamut ; Negidal *ihihīn ~ ihin ~ ihēhin*;
Southern Tungusic: Nanai *sikū*; Ulcha *sičeun*; Udihe *sike*; Oroch *ikken*;
 Orok *sitew ~ siteu*; Manchu *iče*; Sibe *ičə*
 (Janhunen 1991: 52; Dorji & Banzhibomi 1998: 339b; Chaoke 2014a: 166; Vasilevič 1958: 178a; Cincius 1975/77 1: 328a; Hauer 1952-1955 2: 483; Zikmundová 2013: 215)

- (25) There are some Russian loanwords in Khamnigan Ewenki which were investigated during our fieldwork. These were possibly borrowed before the migration from Russia,¹⁸ but are not found in Castrén's monograph:
- a. 'book': Khamnigan Ewenki *kinīska*
← Russian *knížka* 'small book' < *kniga* + *ka* Russian diminutive suffix;
 - b. 'pig': Khamnigan Ewenki *čūske*
← Russian *čúška* 'piglet';
 - c. 'bread': Khamnigan Ewenki *būlke*
← Russian *búlka* 'roll, white loaf';
 - d. 'candy, sweet': Khamnigan Ewenki *hampyētke*
← Russian *konfétka* < *konfeta* + *ka* Russian diminutive suffix ←
German ← Latin;
 - e. 'sugar': Khamnigan Ewenki *sākar*
← Russian *sáhar* ← Arabic; etc.

5.4 Mongolic loanwords

The Mongolic loanwords in Nercha and Khamnigan Ewenki can be divided into two groups. The first one includes the Mongolic loanwords peculiar to both dialects, while the second group contains loanwords borrowed at a different time.

5.4.1 Mongolic loanwords peculiar to both dialects

- (26) Most Mongolic loanwords were borrowed into other Tungusic languages too:
- a. 'person, man': Nercha Ewenki, Khamnigan Ewenki *beye*;
cf. Solon Ewenki *bəyə*; Orochen *beye*; Siberian Common Ewenki *beye*;
other Northern Tungusic: Lamut *bey*; Negidal *beye*; *Southern Tungusic*:
Nanai, Ulcha, Udihe, Oroch, Orok *beye*; Manchu *beye*; Sibe *bey*
(Castrén 1856: 94; Janhunen 1991: 59; Dorji & Banzhibomi 1998: 67b;
Chaoke 2014a: 158; Vasilevič 1958: 73b; Cincius 1975/77 1: 122a; Hauer
1952-1955 1: 89; Zikmundová 2013: 206)
← Mongolic 'body, organism': Middle Mongol: MNT *beye* ~ *be'e*; HY,
'Phags-pa, Muq., Rasulid *beye*;

¹⁸Russian loanwords in Khamnigan Mongol were analyzed in detail by Gruntov & Mazo (2015). These loanwords are possibly also present in Khamnigan Ewenki due to the fact that Russians lived together with Khamnigan Mongol and Khamnigan Ewenki people until the 1960s as one of our Khamnigan Ewenki informants told us.

Literary Mongolian *beye*; Manchurian, Onon and Mongolian Khamnigan Mongol *beye*; Dadal-sum Khamnigan *biye*; Dagur *bəy*; Buryat *beye*

(Khabtagaeva 2017: 64; see also Doerfer 1985: 20; Rozycki 1994: 29);

- b. ‘maral deer’: Nercha Ewenki, Khamnigan Ewenki ***bugu***;
cf. Solon Ewenki *buyu*; Siberian Ewenki: Barguzin, Upper Lena, Vitim *buyu*; Zeya *buy*; Ayan *buye*; Sakhalin, Urmi, Uchur *buyu*;
other Northern Tungusic: Lamut *n.a.*; Negidal *bočan*; *Southern Tungusic*: Nanai *bočā*; Ulcha *buča* ~ *boča*, Oroch *buča*; Manchu *buhû*; Sibe *bukun ihan* ‘mountain antelope’
(Castrén 1856: 95; Janhunen 1991: 24; Dorji & Banzhibomi 1998: 84a; Vasilevič 1958: 64a; Cincius 1975/77 1: 101b; Hauer 1952-1955 1: 119; Stary 1990: 9);
← Mongolic: Middle Mongol: MNT, HY *buqu*; ZY, Muq. *buyu*; Rasulid *buğa*; Literary Mongolian *buyu* ‘a male deer, stag’; Onon Khamnigan Mongol *bugu*; Dagur *bəy^w*; Buryat *buga*
(Khabtagaeva 2017: 67; see also Doerfer 1985: 78; Rozycki 1994: 37);
- c. ‘type of duck’: Nercha Ewenki, Khamnigan Ewenki ***angir***;
cf. Solon Ewenki *angir*; *other Northern Tungusic*: Negidal *ani*; *Southern Tungusic*: Nanai *āngi*; Manchu *angir*; Remaining lgs. *n.a.*
(Castrén 1856: 71; Janhunen 1991: 51; Chaoke 2014b: 52; Vasilevič 1958: 32b; Cincius 1975/77 1: 43b);
Tungusic ← Mongolic: Middle Mongol: MNT, ZY, HY *anggir*; Literary Mongolian *anggir*; Onon Khamnigan Mongol, Buryat *angir*; Dagur *n.a.*;
Mongolic ← Turkic:¹⁹ cf. Old Turkic *anjit* ‘a rather large bird predominantly red; the ruddy goose (*Anas casarca*)’
(Khabtagaeva 2017: 59; see also Doerfer 1985: 68; Rozycki 1994: 19);
- d. ‘to roar’: Nercha Ewenki, Khamnigan Ewenki ***barkirā-***;
cf. Solon Ewenki *baggera-*; Siberian Ewenki: *n.a.*; Remaining Tungusic lgs. *n.a.*
(Castrén 1856: 94; Janhunen 1991: 31; Dorji & Banzhibomi 1998: 53a; Cincius 1975/77 1: 75b);
← Mongolic: Middle Mongol *n.a.*; Literary Mongolian *barkira-*; Onon Khamnigan Mongol *barkir-*; Dagur *n.a.*; Buryat *barxir-* (Khabtagaeva 2017: 63; see also Doerfer 1985: 101);

¹⁹Turkic: Yakut *andī* ~ *anni* ‘scoter, pochard; black duck’ → Siberian Ewenki: Ayan, Uchur *anni* ~ *andi* ‘black duck’.

- e. ‘to think’: Nercha Ewenki, Khamnigan Ewenki **bodo-**;
cf. Solon Ewenki, Orochen *bodo-*; Siberian Ewenki *n.a.*;
other Tungusic: Nanai, Ulcha, Udihe *bodo-*; Oroch *budu-*; Orok *boddo-*
~ *bodo-*; Manchu *bodo-*; Sibe *bot-* ~ *bod(ə)-*; Remaining lgs. *n.a.*
(Castrén 1856: 95; Janhunen 1991: 101; Dorji & Banzhibomi 1998: 75b;
Chaoke 2014a: 168; Cincius 1975/77 1: 88a; Hauer 1952-1955 1: 104;
Zikmundová 2013: 207);
← Mongolic: Middle Mongol: *n.a.*; Literary Mongolian *bodo-*;
Manchurian Khamnigan Mongol *bod-*; Onon Khamnigan Mongol
bodo-; Dagur *bod-*; Buryat *bodo-*
(Khabtagaeva 2017: 65; see also Doerfer 1985: 78; Rozycki 1994: 33);
- (27) The Mongolic loanwords in Nercha and Khamnigan Ewenki with
Khamnigan Mongol features
- a. The disappearance of the Mongolic consonant *q-* through * χ -, which
points to an early period of borrowing:
- i. ‘twenty’: Nercha Ewenki, Khamnigan Ewenki **orin**;
cf. Solon Ewenki *urin*; Orochen *urin*; Siberian Ewenki: Barguzin
orin;
other Northern Tungusic: Negidal *ojin*; *Southern Tungusic*: Nanai
hori; Ulcha, Orok *hori(n)*; Udihe *waji* ~ *uai*; Oroch *oi*; Manchu
orin; Sibe *ori*ⁿ
(Castrén 1856: 75; Janhunen 1991: 23; Dorji & Banzhibomi 1998:
534a; Chaoke 2014a: 170; Vasilevič 1958: 326b; Cincius 1975/77 2:
24; Hauer 1952-1955 3: 740; Zikmundová 2013: 219);
Tungusic ← Mongolic: Middle Mongol: MNT, HY, ‘Phags-pa,
Leiden, Muq. *qorin*; Literary Mongolian *qorin*; Manchurian
Khamnigan Mongol *kori(n)*; Onon Khamnigan Mongol *xori(n)*;
Dadal-sum Khamnigan χ *ori*; Mongolian Khamnigan *orin*; Buryat
xori(n)
(Khabtagaeva 2017: 123; see also Doerfer 1985: 81; Rozycki 1994:
169);
- ii. ‘thumb’: Nercha Ewenki, Khamnigan Ewenki **ürügün**;
cf. Solon Ewenki *arugun*; Siberian Ewenki: Upper Lena,
Chumikan *urugun*; Barguzin *huruwūn*; Aldan, Sakhalin, Urmi,
Uchur *hurugun*;
other Northern Tungusic: Lamut *huregen*; Negidal *hojeŋen*;
Southern Tungusic: Udihe *hue*; Oroch *hōŋo(n)*; Manchu *urhun*

(Castrén 1856: 78; Janhunen 1991: 46; Chaoke 2014b: 153; Cincius 1975/77 2: 354b; Hauer 1952-1955 3: 969);

← Mongolic **χurugun*: Middle Mongol: MNT *quru'u(n)*; HY *quru'un*; Leiden *qurūn*; Muq. *qurūn* ~ *χurūn*; Rasulid *qurūn*; Literary Mongolian *quruyun* 'finger, toe; finger-like'; Manchurian Khamnigan Mongol *kurū(n)*; Onon Khamnigan Mongol *xurū*; Mongolian Khamnigan *xurguon*; Buryat *xurgan*; Dagur *xoᠷᠰ* (Khabtagaeva 2017: 90);

b. The preservation of original Mongolic **ti* which later became *či*:

- i. 'thirty': Nercha Ewenki: Urulga *gučín*, Man'kovo ***gutin***; Khamnigan Ewenki: Borzya ***gutin***, Urulyungui *gučín*; cf. Solon Ewenki *gutín*; Siberian Ewenki: *n.a.* instead *ilan jār*; *other Tungusic*: Nanai *goči*; Ulcha, Oroch *guti(n)*; Manchu *gúsin*; Sibe *gožiⁿ*; Remaining lgs. *n.a.*

(Castrén 1856: 82; Janhunen 1991: 23, 76; Dorji & Banzhibomi 1998: 235a; Cincius 1975/77 1: 175b; Hauer 1952-1955 2: 397; Zikmundová 2013: 213);

← Mongolic: Middle Mongol: MNT, 'Phags-pa *qučín*; Muq. *yučín*; Rasulid *qučín*; Literary Mongolian *yučín*; Manchurian and Onon Khamnigan Mongol *guči(n)*; Dadal-sum Khamnigan Mongol *yutš'i*; Dagur *gǝč* ~ *gǝčín*; Buryat *guša(n)*

(Khabtagaeva 2017: 87; see also Doerfer 1985: 79; Rozycki 1994: 95);

- ii. 'to write': Nercha Ewenki ***biti-*** ~ *biči-*; Khamnigan Ewenki *biči-*; cf. Solon Ewenki *biti-*; Siberian Ewenki: Upper Amur *biči-*;
- iii. 'letter, script': *other Northern Tungusic*: Negidal *bitehe*; *Southern Tungusic*: Nanai *bičhe*; Ulcha *bithe*; Udihe *n.a.*; Oroch *bitige* ~ *bithe*; Orok *bičihe*; Manchu *bithe*; Sibe *bitə*

(Castrén 1856: 95; Janhunen 1991: 101; Dorji & Banzhibomi 1998: 74b; Cincius 1975/77 1: 86; Hauer 1952-1955 1: 98; Zikmundová 2013: 206);

Tungusic ← Mongolic **biti-*: Middle Mongol: MNT, ZY, HY, Muq., Rasulid *bičik* 'script'; Literary Mongolian *biči-* 'to write'; Manchurian and Onon Khamnigan Mongol *biči-*; Dagur *n.a.*; Buryat *beše-*;

Mongolic ← Turkic: Old Turkic *biti-* 'to write' (Khabtagaeva 2017: 65; see also Doerfer 1985: 76; Rozycki 1994: 31);

- c. The preservation of the vowels *-u-* and *-ü-* in the last syllable:
- i. ‘butter, oil’: Nercha Ewenki (Man’kovo) *tosun*, (Urulga) *tohun*; Khamnigan Ewenki *tosun*;
cf. Solon Ewenki *n.a.*; Siberian Ewenki: North-Baikal *tosun*; Aldan, Barguzin *tohun*; *other Tungusic: n.a.*
(Castrén 1856: 88; Janhunen 1991: 24; Vasilevič 1958: 395b; Cincius 1975/77 2: 201a);
← Mongolic: Middle Mongol: MNT, HY, ‘Phags-pa, Leiden, Muq. *tosun*; Rasulid *tosun*; Literary Mongolian *tosun*; Manchurian Khamnigan Mongol *tohun*; Onon Khamnigan Mongol *tosu(n) ~ toso(n)*; Dadal-sum Khamnigan Mongol *t’osu*; Dagur *tōs*; Buryat *toho(n)* (Khabtagaeva 2017: 134);
 - ii. ‘age’: Nercha Ewenki *n.a.*; Khamnigan Ewenki *nasun*;
cf. Solon Ewenki *n.a.*; Siberian Ewenki: Barguzin *nahun*; *other Tungusic: n.a.*
(Janhunen 1991: 24; Cincius 1975/77 1: 587a);
← Mongolic: MNT *nasu*; ‘Phags-pa *nasu ~ nasun*; Leiden, Muq. *nasun*; Literary Mongolian *nasun*; Manchurian Khamnigan Mongol *nahun*; Onon Khamnigan Mongol *nasu(n) ~ nasa(n)*; Mongolian Khamnigan *nasu*; Dagur *nas*; Buryat *naha(n)* (Khabtagaeva 2017: 119; see also Doerfer 1985: 127);
 - iii. ‘bovine’: Nercha Ewenki *ükür*; Khamnigan Ewenki: Urulyungui *ükür*; Borzya *hükür*;
cf. Solon Ewenki *n.a.*; Orochen *ukur*; Siberian Ewenki: Barguzin, Zeya, Aldan, Khingan, Uchur *hukur*; *other Tungusic: n.a.*
(Castrén 1856: 83; Janhunen 1991: 46; Chaoke 2014a: 157; Vasilevič 1958: 491b; Cincius 1975/77 2: 341);
Tungusic ← Mongolic: Middle Mongol: MNT *hüker*; ZY *üger*; HY *hüger*; ‘Phags-pa, Leiden, Muq. *hüker*; Rasulid *üker*; Literary Mongolian *ükür*; Manchurian Khamnigan Mongol *üker*; Onon Khamnigan Mongol *üker ~ ökör* (← Khalkha); Dadal-sum Khamnigan Mongol *ük’ür*; Dagur *xukur*; Buryat *üxer*;
Mongolic **ükür* ‘bovine animal, ox, cow’ < *hükür* ← Bulgar Turkic **hökür*: cf. Old Turkic *öküz* ‘ox’ ← Tokharian (Khabtagaeva 2017: 89; see also Doerfer 1985: 67);
- d. The preservation of Middle Mongol intervocalic *q* which later is voiced:

- i. ‘hedgehog’: Nercha Ewenki, Khamnigan Ewenki **dorokon**;
 cf. Solon Ewenki *n.a.*; Siberian Ewenki: Podkamennyj, Barguzin
dorokon; Zeya *dorogun*; Chumikan *toraku*; Urmi *torokōn*;
other Northern Tungusic: Lamut *n.a.*; Negidal *doho*; *Southern
 Tungusic*: Nanai *doro*; Ulcha *doro(n)*, Oroch *dogo*; Udihe, Orok *n.a.*;
 Manchu *dorgon*; Sibe *dorhon*
 (Castrén 1856: 90; Janhunen 1991: 40; Vasilevič 1958: 122a; Cincius
 1975/77 1: 217a; Hauer 1952-1955 1: 210; Stary 1990: 19);
 ← Mongolic: Middle Mongol: Leiden *dorkon*; Literary Mongolian
doroyon, Mongolian Khamnigan Mongol *dorokon*; Onon
 Khamnigan Mongol *dorogon*; Dagur *n.a.*; Buryat *dorgon*
 (Khabtagaeva 2017: 80; see also Doerfer 1985: 39; Rozycki 1994:
 62);
- e. The preservation of Mongolic **ni*:
- i. ‘homeland’: Nercha Ewenki *n.a.*; Khamnigan Ewenki **nitug**;
 cf. Solon Ewenki, Siberian Ewenki *n.a.*; *other Tungusic*: *n.a.*
 (Janhunen 1991: 24);
 ← Mongolic: Middle Mongol: MNT *nuntuq* ~ *nutuq*; HY *nuntuq*;
 Muq. *nutuq*; Literary Mongolian *nituy*; Manchurian Khamnigan
 Mongol *nitug* ~ *nutug*; Onon Khamnigan Mongol *nitug*; Dagur
notog; Buryat *nyutag*;
- ii. ‘sin’: Nercha Ewenki, Khamnigan Ewenki **nigul**;
 cf. Solon Ewenki *niwul*; Siberian Ewenki: Upper Lena *niñul*; *other
 Tungusic*: *n.a.* (Castrén 1856: 85; Janhunen 1991: 24; Chaoke 2014b:
 316; Cincius 1975/77 1: 589a);
 ← Mongolic: Middle Mongol: ‘Phags-pa *ni’ül*; Literary
 Mongolian *niyul*; Manchurian Khamnigan Mongol *nigül*; Onon
 Khamnigan Mongol *nügel* (← Buryat); Dagur *nugul*; Buryat
nügel (Khabtagaeva 2017: 120);
- f. The development of the Mongolic sequence *egü* > *ē* as in Manchurian
 Khamnigan Mongol:²⁰
- i. ‘mare’: Nercha Ewenki **gēk**; Khamnigan Ewenki: Borzya **gēg**,
 Urulyungui **gē**;

²⁰E.g. Manchurian Khamnigan Mongol *kēgen* ‘child’ ~ Mongolic ‘girl’: Literary Mongolian *keūken*; Buryat *xūxen*; Manchurian Khamnigan Mongol *tēke* ‘history’ ~ Mongolic: Literary Mongolian *teūke*; Buryat *tūxe*; Manchurian Khamnigan Mongol *dē* ‘younger brother’ ~ Mongolic: Literary Mongolian *degūū*; Buryat *dū*, etc. (Janhunen 1990: 28).

cf. Solon Ewenki *gə*; Siberian Ewenki: Vitim *gēγ*; Barguzin *gog*;
Upper Lena *gēn*;

other Tungusic: Manchu *geo*

(Castrén 1856: 81; Janhunen 1991: 41; Dorji & Banzhibomi 1998:
207a; Vasilevič 1958: 84a; Cincius 1975/77 1: 145; Hauer 1952-1955 1:
345);

← Mongolic: Middle Mongol: MNT, HY, Muq. *ge'ün*; Literary
Mongolian *gegüü*; Manchurian Khamnigan Mongol *gē*;
Mongolian Khamnigan *gökü*; Onon Khamnigan Mongol, Buryat
gū; Dagur *gəu* (Khabtagaeva 2017: 52; see also Doerfer 1985: 102;
Rozycki 1994: 88);

- g. The preservation of Mongolic *VgV*, which became a secondary long
vowel in Modern Mongol:

- i. 'goat': Nercha Ewenki, Khamnigan Ewenki ***imagan***;

cf. Solon Ewenki *imayᠠᠭ*; Siberian Ewenki *imagan*: Barguzin
'goat'; Uchur, Urmi, Sakhalin 'bastard calf';

other Northern Tungusic: Lamut *n.a.*; Negidal *imaja*; *Southern
Tungusic*: Nanai, Ulcha, Udihe, Oroch *ima*; Orok *n.a.*; Manchu
imahū 'Capricorn'; Sibe *n.a.*

(Castrén 1856: 75; Janhunen 1991: 100; Dorji & Banzhibomi 1998:
330b; Vasilevič 1958: 167b; Cincius 1975/77 1: 312b; Hauer
1952-1955: 497);

← Mongolic: Middle Mongol: MNT *ima'a*; HY *ima'an*; Muq.
ima'an ~ *imān*; Rasulid *imān*; Literary Mongolian *imayᠠᠭ*;
Manchurian Khamnigan Mongol *imā(n)*; Onon Khamnigan
Mongol *yamā(n)* (← Buryat); Mongolian Khamnigan *imagān*;
Dagur *imā*; Buryat *yamā(n)* (Khabtagaeva 2017: 90; see also
Doerfer 1985: 37; Rozycki 1994: 116);

- ii. 'antelope': Nercha Ewenki *n.a.*; Khamnigan Ewenki ***jeđerēn***;
cf. Solon Ewenki *dʒəgəɾəŋ* 'Mongolian gazelle'; Siberian Ewenki
n.a.;

other Tungusic 'roe deer, wild goat': Udihe *jeļi*; Manchu *jeřen*;
Remaining lgs. *n.a.*

(Janhunen 1991: 100; Chaoke 2014b: 40; Cincius 1975/77 1: 282b;
Hauer 1952-1955 2: 530);

Tungusic ← Mongolic: Middle Mongol: HY, Muq. *jeřen*; Literary
Mongolian *jeđer-e(n)*; Manchurian Khamnigan Mongol *jeře(n)*;

Onon Khamnigan Mongol *dzēr* (← Khalkha); Dagur *ǰrən*; Buryat *zēren*;

Mongolic ← Turkic: cf. Old Turkic *yägrän* ‘gazelle’ (Doerfer 1985: 136; Rozycki 1994: 122);

- iii. ‘camel’: Nercha Ewenki *n.a.*; Khamnigan Ewenki ***temegēn***; Solon Ewenki *təmægəŋ*; Orochen *temegen*; Siberian Ewenki: Barguzin *temegēn*;
other Tungusic: Nanai, Oroch *teme*; Manchu *temen*; Remaining lgs. *n.a.*

(Janhunen 1991: 100; Chaoke 2014b: 56; Chaoke 2014a: 157; Cincius 1975/77 2: 235a; Hauer 1952-1955 3: 899);

Tungusic ← Mongolic: Middle Mongol: MNT *teme’en*; ZY *te[m]mē*; HY *teme’en*; Muq. *temēn*; Rasulid *temēn*; Literary Mongolian *temegen*; Manchurian and Onon Khamnigan Mongol *temē(n)*; Dadal-sum Khamnigan Mongol *t’emē*; Mongolian Khamnigan *temegēn*; Dagur *təmə*; Buryat *temē(n)* (Khabtagaeva 2017: 133; see also Doerfer 1985: 77–78; Rozycki 1994: 206);
Mongolic ← Turkic: cf. Old Turkic *täβäy* ‘camel’;

5.4.2 Mongolic loanwords borrowed at a different time

Castrén’s Nercha Ewenki material includes the Mongolic “early stage” loanwords which are present in other Tungusic languages as well, while Manchurian Khamnigan Mongol has Mongolic loanwords borrowed more recently.

- (28) The final Mongolic consonant *-l is presented as -n in all Ewenki dialects:

- a. ‘saddle’: Nercha Ewenki: Urulga ***emegen***, Borzya *emēl*;
cf. Khamnigan Ewenki *emegēl* ← Solon Ewenki *əmægəl*;
cf. Siberian Ewenki *emegen*: Yerbogochon, Upper Lena, Barguzin, Tungir, Ayan ‘saddle’; Zeya, Aldan, Uchur, Urmi, Ayan, Sakhalin ‘pack saddle’; Podkamennyj, Ilimpeya ‘men’s hunting saddle’;
other Tungusic: Lamut *emgun*; Orok *emē(n) ~ emegē(n)*; Manchu *engemu*; Sibe *əməŋ*; Remaining lgs. *n.a.*

(Castrén 1856: 73; Janhunen 1991: 100; Dorji & Banzhibomi 1998: 172; Vasilevič 1958: 558b; Cincius 1975/77 2: 452b; Hauer 1952-1955 1: 252; Zikmundová 2013: 210);

Tungusic ← Mongolic: Middle Mongol: MNT *eme’el*; ZY, Muq., Rasulid *emēl*; Literary Mongolian *emegel*; Manchurian Khamnigan

Mongol *emēl*; Onon Khamnigan Mongol *emēl* ~ *ömōl*; Dagur, Buryat *emēl* (Khabtagaeva 2017: 83; see also Doerfer 1985: 21; Rozycki 1994: 70);

- (29) The secondary long vowels have not developed yet in Nercha Ewenki and other Ewenki dialects from sequence VgV, while in Khamnigan Ewenki they have, as in Modern Mongol:
- a. ‘young’: Nercha Ewenki *ǰalaf*; cf. Khamnigan Ewenki: Urulyungui *ǰalō*, Borzya *ǰalau*
← Solon Ewenki *ǰalu*; cf. Siberian Ewenki: Aldan, Barguzin, Upper Lena, Zeya, Tungir, Uchur *ǰalaw*; *other Tungusic*: *n.a.* (Castrén 1856: 93; Janhunen 1991: 30, 54; Dorji & Banzhibomi 1998: 356b; Vasilevič 1958: 147a; Cincius 1975/77 1: 245a); ← Mongolic: Middle Mongol: MNT *ǰala’ui*; Muq. *ǰala’ū* ~ *ǰalū*; Ist. *ǰalau*; Rasulid *ǰalawu*; Literary Mongolian *ǰalayu*; Onon Khamnigan Mongol *dzalū*; Dadal-sum Khamnigan Mongol *džalalǰan* ‘boy’; Dagur *ǰalō*; Buryat *zalū* (Khabtagaeva 2017: 93; see also Doerfer 1985: 127);
 - (30) The original Tungusic words with initial and intervocalic *s in Nercha Ewenki and Manchurian Khamnigan Ewenki were preserved, while in Mongolic loanwords this sound was pharyngealized in the Urulga sub-dialect and preserved in the Mankovo sub-dialect of Nercha Ewenki (Khabtagaeva 2017: 175). Thus, the pharyngealization must have happened in a later period. More likely these loanwords were borrowed from the ancestors of the Urulyungui subdialect of Khamnigan Mongol:
 - a. ‘rope, loop, lasso’: Nercha Ewenki (Mankovo) *desün*; (Urulga) *dehün*; cf. Khamnigan Ewenki, Solon Ewenki, Siberian Ewenki *n.a.*; *other Tungusic lgs.* *n.a.* (Castrén 1856: 89);
← Mongolic: Middle Mongol: ZY *dēsü*; HY *de’esün*; Muq. *dēsün*; Rasulid *dēsün*; Literary Mongolian *degesün*; Onon Khamnigan Mongol *dēsün*; Dagur *dəs*; Buryat *dēhe(n)* (Khabtagaeva 2017: 80);
 - b. ‘fly’: Nercha Ewenki (Mankovo) *ilāsun*; (Urulga) *ilāhun*; Khamnigan Ewenki *ilāsun*;
cf. Solon Ewenki *ilā*; Siberian Ewenki *n.a.*; *other Tungusic lgs.* *n.a.* (Castrén 1856: 74; Janhunen 1991: 105; Dorji & Banzhibomi 1998: 324b; Cincius 1975/77 1: 306b);
← Mongolic: Literary Mongolian *ilayasun*; Manchurian Khamnigan Mongol *ilāhun*; Onon Khamnigan Mongol *ilā*; Dadal-sum Khamnigan

Mongol *ilā*; Dagur *xilā* ‘horsefly’; Buryat *ilāha(n)* (Khabtagaeva 2017: 90);

cf. Orochen *dilkān*, other Ewenki dial. *dilkēn*;

- (31) There is one Mongolic loanword which is absent in other Tungusic languages, yet must have been borrowed at an early stage:
- a. ‘forty’: Nercha Ewenki, Khamnigan Ewenki *düčün*;
in other Ewenki dialects *digin jār*; Solon Ewenki *n.a.*²¹
(Castrén 1856: 90; Janhunen 1991: 76);
← Mongolic: Middle Mongol: MNT *döčün*; ZY *düčün*; HY, Muq., Ist.,
Rasulid *döčün*; Literary Mongolian *döči(n)*; Manchurian Khamnigan
Mongol *düči(n)*; Onon Khamnigan Mongol *düči(n)*; Dadal-sum
Khamnigan Mongol *döts’i*; Dagur *duč*; Buryat *düše(n)* (Khabtagaeva
2017: 81).

6 Conclusion

As expected, the material examined shows a close connection between Nercha and Khamnigan Ewenki. Most of the vocabulary coincides, yet in several cases the other Ewenki dialects have differing forms (e.g. *nekün* ‘younger brother’, *nuḡnakī* ‘goose’, *timī* ‘tomorrow’, *nama* ‘warm’, etc.) or even lack some words (e.g. *düčün* ‘forty’). These facts argue strongly for a common linguistic background between the two varieties. A separate group of vocabulary items includes the Nercha Ewenki words which have some phonetic differences from Khamnigan Ewenki, though these variants can possibly be explained by Castrén’s transcriptions from 1856. However, there are some Khamnigan Ewenki words that changed phonetically recently, under the influence of Solon Ewenki – another neighboring Tungusic language of Manchuria. These include words with the unstable consonant *-n* (e.g. *amin* ‘father’, *tergen* ‘carriage’) and secondary long vowels in the first syllable of words (e.g. *āli* ‘when’). These vowels also appear in the Mongolic loanwords (e.g. *ēm* ‘medicine, drug’, *tōrga* ‘silk’). In addition, some Solon Ewenki words were borrowed by Khamnigan Ewenki after the migration from Russia (e.g. *sigün* ‘sun’, *irkekīn* ‘new’). More likely, the Russian loanwords in Khamnigan Ewenki were borrowed at an early time but were not noted by Castrén in his Nercha Ewenki material. According to phonetic criteria (e.g. the preservations of the vowel **u* in the last syllable, Middle Mongol *VqV*, and sequences **ni* and **si*), most of the Mongolic loanwords in Nercha and Khamnigan

²¹Solon *dāhi* (Chaoke 2014b: 346) ← Manchu *dehi*.

Ewenki dialects match each other. These facts also support a common linguistic heritage of the two varieties. Due to our short and preliminary fieldwork among the Khamnigan Ewenki people in 2017, further fieldwork is necessary for a morphological and syntactic analysis, which I was not able to present here.

Citation of data

Dadal-sum Khamnigan Mongol	= Uray-Kõhalmi (1959)
HY	= Mostaert (1977)
Leiden	= Poppe (1927/1928)
Literary Mongolian	= Lessing (1960)
Manchurian Khamnigan Mongol	= Janhunnen (1990)
MNT	= Haenisch (1939)
Mongolian Khamnigan	= Rinčen (1968)
Muq.	= Poppe (1938)
Onon Khamnigan Mongol	= Damdinov & Sundueva (2015)
'Phags-pa	= Tumurtogoo (2010)
Rasulid	= Golden (2000)
ZY	= Kara (1990)

Acknowledgements

This paper is supported by the Alexander von Humboldt Foundation. I would like to thank my colleagues and friends from Charles University Prague, Dr. Veronika Zikmundová and Dr. Veronika Kapišovská, with whom I conducted the fieldwork. Also, I would like to thank all our consultants and contacts in Hulunbuir, China. I am also grateful to Dr. Andreas Hölzl (University of Zurich) for his helpful and invaluable remarks.

References

- Atknine, Victor. 1997. The Evenki language from the Yenisei to Sakhalin. *Senri Ethnological Studies* 44. 109–121.
- Bulatova, N. Ja. 1987. *Govory évenkov amurskoj oblasti*. Leningrad: Nauka.
- Bulatova, N. Ja. 2002. Évenkijskij jazyk. In V. P. Neroznak (ed.), *Ĵazyki narodov Rossii. Krasnaja kniga. Enciklopedičeskij slovar'-spravočnik*, 267–272. Moscow: Academia.

- Castrén, M. Alexander. 1856. *Grundzüge einer tungusischen Sprachlehre nebst kurzem Wörterverzeichnis*. St Petersburg: Buchdruckerei der kaiserlichen Akademie der Wissenschaften.
- Chaoke, D. O. 2014a. *Elunchunyu 366 ju huihuaaju [366 conversational phrases in Orochen]*. Peking: Shehui kexue wenxian chubanshe.
- Chaoke, D. O. 2014b. *Mantonggusi yuzu yuyan cihui bijiao [A comparative dictionary of Manchu-Tungusic languages]*. Peking: Zhongguo shehui kexue chubanshe.
- Chaoke, D. O. 2017. *Ewenke zu san da fangyan cihui bijiao [A comparative dictionary of the three big Ewenke dialects]*. Peking: Shehui kexue wenxian chubanshe.
- Cincius, Vera I. et al. 1975/77. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov*. 2 vols. Leningrad: Nauka.
- Damdinov, D. G. & E. V. Sundueva. 2015. *Xamnigansko-russkij slovar'*. Irkutsk: Ottisk.
- Doerfer, Gerhard. 1985. *Mongolo-Tungusica (Tungusica 3)*. Wiesbaden: Otto Harrassowitz.
- Dorji, Do. 2013. *Ewengki Mønggøl bilehu biteḡe/Emeng cidian [Ewenki-Mongolian dictionary]*. Peking: Minzu chubanshe.
- Dorji, Do & Banzhibomi. 1998. *Ewengki nihang bilehu biteḡ/Ehan cidian [Ewenki-Chinese dictionary]*. Hailar: Neimenggu wenhua chubanshe.
- Galdanova, G. R., K. M. Gerasimova, D. B. Dašiev & G. C. Mitupov. 1983. *Lamaizm v Burjatii XVII – načala XX veka. Struktura i social'naja rol' kul'tovoj sistemy*. Novosibirsk: Nauka.
- Golden, Peter et al. (ed.). 2000. *The King's dictionary: The Rasulid Hexaglot: Fourteenth century vocabularies in Arabic, Persian, Turkic, Greek, Armenian and Mongol*. Leiden: Brill.
- Gruntov, Ilya A. & Olga M. Mazo. 2015. Russkie zaimstvovanija v mongol'skom hamniganskom. In Grigorij C. Pjurbeev (ed.), *Rossijsko-mongol'skij lingvističeskij sbornik*, 165–185. Moscow: Institut jazykoznanija.
- Haenisch, Erich. 1939. *Wörterbuch zu Mangḡol-un Niuca Tobca'an. (Yüan-ch'ao pi-shi): Geheime Geschichte der Mongolen*. Leipzig: Harrassowitz.
- Hauer, Erich. 1952-1955. *Handwörterbuch der Mandschusprache*. 3 vols. Wiesbaden: Otto Harrassowitz.
- Hözl, Andreas. 2018a. *A typology of questions in Northeast Asia and beyond: An ecological perspective (Studies in Diversity Linguistics 20)*. Berlin: Language Science Press.

- Hölzl, Andreas. 2018b. Udi, Udihe, and the language(s) of the Kyakala. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 15. 111–146.
- Janhunen, Juha. 1990. *Material on Manchurian Khamnigan Mongol* (Castrenianumin toimitteia 37). Helsinki: University of Helsinki.
- Janhunen, Juha. 1991. *Material on Manchurian Khamnigan Evenki* (Castrenianumin toimitteia 40). Helsinki: University of Helsinki.
- Janhunen, Juha. 1997. The languages of Manchuria in today's China. *Senri Ethnological Studies* 44. 123–146.
- Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.
- Kara, György. 1990. Zhiyuan Yiyu. Index alphabétique des mots mongols. *Acta Orientalia Academiae Scientiarum Hungaricae* 44. 279–344.
- Khabtagaeva, Bayarma. 2017. *The Ewenki dialects of Buryatia and their relationship to Khamnigan Mongol* (Tunguso Sibirica 41). Wiesbaden: Harrassowitz.
- Lessing, Ferdinand D. 1960. *Mongolian-English dictionary*. 3rd edn. Berkeley: University of California Press.
- Mostaert, Antoine. 1977. *Le matériel mongol du Houa I I Iu de Houng-ou (1389)*. Brussels: Institut Belge des hautes études Chinoises.
- Norman, Jerry. 1977. The evolution of Proto-Tungusic *t to Manchu s. *Central Asiatic Journal* 21(3-4). 229–233.
- Poppe, Nicholas. 1938. *Mongol'skij slovar' Mukaddimat al-Adab*. vols. 1-2. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Poppe, Nicholas. 1927/1928. Das mongolische Sprachmaterial einer Leidener Handschrift. *Izvestija Akademii Nauk SSSR* 1927(12-14). 1009-1040; 1927(15-17). 1251-1274; 1928. 55-79.
- Rinčen, Byambin. 1968. *Mongol ard ulsiin xamnigan ayalguu*. Ulaanbaatar: Šinžlex Uxaanii Akademiin xewlel.
- Rozycki, William. 1994. *Mongol elements in Manchu*. Bloomington: Indiana University.
- Stary, Giovanni. 1990. *Taschenwörterbuch Sibemandschurisch-Deutsch*. Wiesbaden: Otto Harrassowitz.
- Tal'ko-Gryncevič, Yu. D. 1904. *K antropologii tungusov. Irojskie xamneganj antropologo-ětnologičeskij očerk*, vol. 7. 66–199.
- Titov, E. 1926. *Tungussko-russkij slovar' s priloženijem knigi M.A. Kastrena „Osnovy izučenija tungusskogo jazyka”*. Irkutsk: Izdatel'stvo Čitinskogo gosudarstvennogo muzeja imeni A. K. Kuznecova.

- Tugolukov, V. A. 1975. Konnye tungusy (ètničeskaja istorija i ètnogenez). In I. S. Gurvič (ed.), *Ètnogenez i ètničeskaja istorija narodov Severa*, 78–110. Moscow: Nauka.
- Tumurtogoo, Domiin. 2010. *Mongolian monuments in 'Phags-pa script. Introduction, transliteration, transcription and bibliography*. Taipei: Institute of Linguistics, Academia Sinica.
- Uray-Kõhalmi, Katalin. 1959. Der mongolisch-kamniganische Dialekt von Dadal Sum und die Frage der Mongolisierung der Tungusen in der Nordmongolei und Transbaikalien. *Acta Orientalia Scientiarum Hungaricae* 9. 163–204.
- Uvarova, T. B. 2006. *Nerčinskije èvenki v 18-20 vekax*. Moscow: INIOH RAN.
- Vasilevič, Glafira M. 1958. *Èvenkijsko-ruskij slovar'*. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovar'ej.
- Vasilevič, Glafira M. 1969. *Èvenki: Istoriko-etnografičeskie očerki (XVIII – načalo XX v.)* Leningrad: Nauka.
- Zikmundová, Veronika. 2013. *Spoken Sibe: Morphology of the inflected parts of speech*. Prague: Karolinum.

Chapter 6

Functions of placeholder words in Evenki

Elena Klyachko

Higher School of Economics & Institute of Linguistics, Russian Academy of Sciences

Placeholders are used to fill in the pause when the speaker has forgotten the exact word. They have the syntactic properties of the word the speaker cannot recall (the target word). Studying placeholders is thus important for understanding how discourse works. However, the area has been much understudied, especially for low-resource languages, due to the lack of oral corpora. This paper fills in this lacuna for the Evenki language. It describes the functions of placeholders and their grammatical properties, drawing on data from oral corpora and elicitation. More specifically, it looks into the transfer of grammatical features from the target word to the placeholder. Dialectal distribution of placeholders and their correlates in other Tungusic languages are also discussed.

1 Introduction

1.1 Placeholder words

In conversation, speakers can employ a number of devices in case they hesitate or have forgotten the exact word. “Non-silence devices” which are used to fill in the pause are called fillers. More specifically, fillers “fulfilling the syntactic projection” of a phrase (in contrast to interjections) are called placeholders (see Fox 2010 for a discussion of the terms). For many languages, corpus-based placeholder studies may be difficult. Firstly, placeholder words were usually omitted in older published materials, which are not accompanied by audio. Secondly, the very technique of writing down texts without speech recorders (such as asking



the speaker to dictate) may have forced the speakers to use fewer placeholder words.

1.2 The Evenki language

In this work, a study of placeholder words in the Evenki language is performed. Evenki is an endangered Tungusic language spoken in Russia, China and Mongolia. In Russia, there are fewer than 5000 speakers (*Russian census* 2010). For China, a number of 11 000 is given in *Ethnologue* (2019) but we should take into account that the traditional Chinese classification counts Solon, Aoluguya and Khamnigan Evenki as Evenki dialects whereas Oroqen is considered a separate language (Tsumagari 1992). However, Oroqen is actually closer to the Russian Evenki dialects (as well as Aoluguya and Khamnigan Evenki) than Solon. Therefore, if we count only dialects of Evenki proper in China, there are roughly 2500 Oroqen speakers (Whaley & Li 2000), fewer than 200 Aoluguya speakers (Tsumagari 1992), and fewer than 1000 Khamnigan Evenki speakers (Whaley 1998). As regards Mongolia, the Khamnigan Evenki language seems to be extinct. This paper addresses the Evenki dialects of Russia due to the lack of oral speech corpora from China or Mongolia.

The Evenki language is spread over a huge territory and comprises numerous dialects, which are quite different from each other. Vasilevich (1948) provides a classification of the Evenki dialects spoken in the former USSR, dividing them into three groups: Northern, Southern, and Eastern (see Figure 1 for a map).

1.3 Notes on Evenki morphology

In this paper, placeholders are analyzed from the morphological point of view. Therefore, a brief introduction into the Evenki morphology will be given here.

Evenki is an agglutinating language with rich derivational and inflectional morphology. Nedjalkov lists the following morphological classes in Evenki (Nedjalkov 1997: 139–140):

- nouns
- pronouns
- verbs
- adjectives

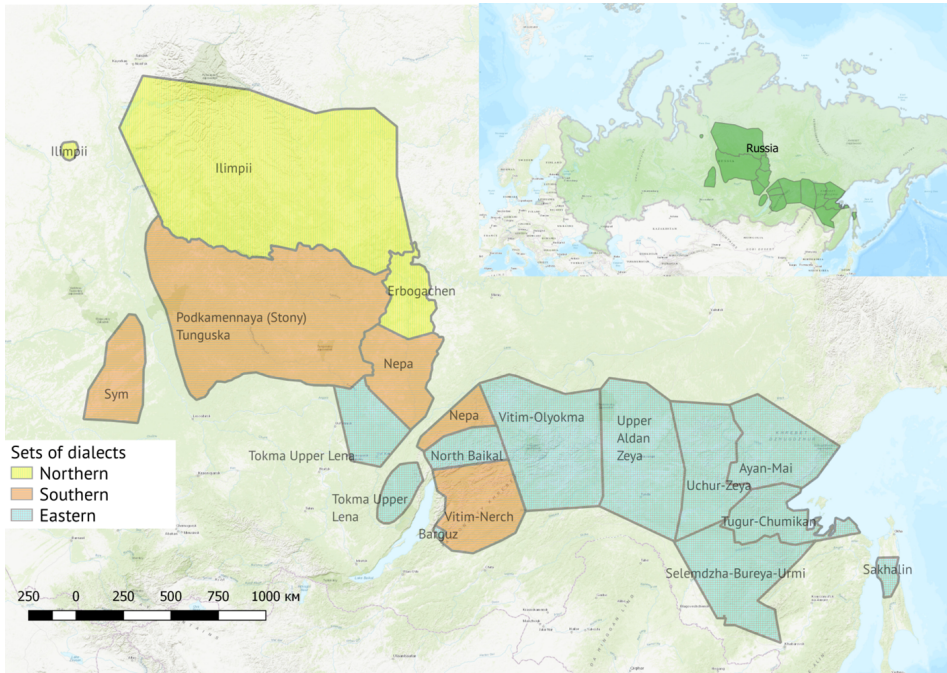


Figure 1: Evenki dialects of Russia (based on Vasilevich 1948, redrawn by Nadezhda Mamontova)

- postpositions (usually nominal stems with a locative meaning, having possessive suffixes)
- numerals/quantifiers
- particles (including what is called an adverb in traditional classifications)

A nominal wordform has the template shown in Table 1. For example:

- (1) $bəjə_1 \quad \emptyset_2 \quad -ŋi_3 \quad -l_4 \quad -ba_5 \quad -w_6$
 person -IND.POSS -PL -ACC -1SG.POSS
 ‘<told about> my people’
- (2) $əwi:1 \quad -wu_2 \quad \emptyset_3 \quad -r_4 \quad -ə_5 \quad -tin_6$
 play -NMLZ -PL -ACCIN -3PL.POSS
 ‘<made> toys for them’

Verbal wordforms can be finite or non-finite (participles and converbs), depending on whether it can be the only verbal form in an independent clause.¹ A finite verb form has the following template (Table 2).²

Table 1: Nominal template
(Bulatova & Grenoble 1999:
15)

1	stem
2	derivational affixes
3	alienable possession
4	number
5	case
6	inalienable possession

Table 2: Finite verbal
template (Bulatova &
Grenoble 1999: 27)

1	stem
2	derivational affixes
3	voice
4	aspect
5	mood+tense
6	mood+number

For example:

- (3) $\widehat{t}fa_1 -ti_2 -pka:3 -l_4 -d\partial_5 -n_6 =da:$
 tea -VBLZ -CAUS -INCH -NFUT -3SG -FOC
 ‘She started to give tea to drink.’

Non-finite verb forms can have personal or number endings, which depends on the actual participial or converbial form itself. A non-finite verb form has therefore the scheme shown in Table 3.

Table 3: Non-finite verbal template

1	stem
2	derivational affixes
3	voice
4	aspect
5	participial or converbial affix
6	person + number/ number
7	case (for participles)
8	inalienable possession (for participles)

¹Actually, there are rare cases of non-finite forms used independently in oral speech.

²Some aspect affixes can precede voice affixes. There can be several aspect affixes in a verb form (Tables 2–3).

(4) is an example of a converb with no personal endings, (5) is a converb with personal endings, and (6) is a participle.

(4) *bori*₁ \emptyset ₂ *-ma:t*₃ *-tʃə*₄ *-nə*₅ *-l*₆
 share -RECIP -IPFV -CVSIM -PL
 ‘<They live> sharing with each other.’

(5) *hokori*₁ *-w*₂ \emptyset ₃ \emptyset ₄ *-rak*₅ *-tin*₆
 lose -PASS -CVCOND -3PL
 ‘if they are lost.’

(6) *boyolo*₁ \emptyset ₂ \emptyset ₃ *-t*₄ *-tʃə*₄ *-ri*₅ \emptyset ₆ *-duk*₇ *-tin*₈
 deprive.of.share -DUR -IPFV -PSIM -ABL -3PL.POSS
 ‘because of their depriving of the share.’

All wordforms can be followed by a clitic (as in 3).

1.4 Aims

The aims of this work are as following:

- Describe various placeholder words in the Evenki language. To my knowledge, no comprehensive work of such kind has ever been done.
- Define the functions of the placeholder words and compare them to the functions observed typologically.

1.5 Methods

The study is mainly based on a corpus of texts, which have been recorded, transcribed and analyzed by a group of linguists, including the author of this work (Siberian-Lang corpus 2019). The material was recorded in 2007–2018 in Tomsk oblast, Krasnoyarsk krai and Irkutsk oblast. The corpus comprises mainly texts recorded from speakers of the Northern dialect group. A corpus of oral Evenki texts recorded in Krasnoyarsk krai by Nadezhda Mamontova in 2014 is also used (Corpora IEA 2019). Another source is descriptive grammars of the Evenki language: Konstantinova (1964), Nedjalkov (1997), and Bulatova & Grenoble (1999). Both corpora are more focused on the Northern and Southern Evenki dialects, while data on the Eastern dialects is scarce. Furthermore, the corpora of related Tungusic languages have been studied.

I conducted several elicitation experiments in Krasnoyarsk krai, Irkutsk oblast, and Khabarovsk krai. The design was straightforward: the speakers were given sentences containing placeholders and asked whether the sentences sounded acceptable and which words could be used instead. However, when analyzing the elicitation results, it should be taken into account that the status of the placeholder words is very low among the speakers. They are often referred to as “slips of the tongue” or “just insertions to connect the words together”. Sometimes a speaker says about a particular placeholder that there is no such word, although they still use it in their own speech. Still, some speakers recall “people using these words in the past when telling something”, and even try to distinguish the meanings of the placeholders.

2 Placeholder words in Evenki

In this section, placeholder words in Evenki will be described in detail according to the following plan, which roughly follows Podlesskaya (2010).

1. *Functions as a placeholder*: The placeholder can be used in various pragmatic circumstances, which will be further described in greater detail.
2. *Restrictions on the target word*: The target word is the word the speaker cannot recall. Therefore, it is replaced with a placeholder. Typologically, there can be placeholders which replace only words of specific classes (e. g., only nouns, or only proper nouns).
3. *Functions other than those of a placeholder*: Data from other languages show that some placeholder stems can also be used as interjections or hesitation markers without actually replacing any target word.
4. *Mirroring the grammatical shape of the target word*: The placeholder can copy inflectional and/or derivational affixes from the target word, thus reflecting its structure. Podlesskaya (2010) calls this mirroring. The target word is then sometimes recalled, so we can compare its surface realization to that of the placeholder. Typological studies show that there is sometimes no mirroring at all. In other cases, mirroring can be full (copying all affixes) or partial (copying only some of them).
5. *Frequency*: We can leverage corpus data to understand how frequent a particular placeholder is, and whether it depends on the dialect, on the speaker’s age, gender, or fluency in Evenki, as well as on the genre of the text.

6. *Dialectal distribution of a placeholder*: As shown above, the Evenki comprises various dialects that may display different behaviour in terms of placeholders.
7. *Possible source and comparative data from related languages*: In typological literature on placeholders, they are often traced to a pronoun or to a noun meaning ‘thing’. It is therefore crucial to trace the etymology of the placeholder, if possible. Evenki is a Northern Tungusic language, so I will use available data from other Northern as well as Southern Tungusic languages to indicate the etymology of common placeholders.

In all examples, the placeholder word will be put in **bold**, whereas the correspondent target phrase will be underlined. In translations, “whatsitsname” and “do that thing” will be used.

2.1 *aŋə / aŋi*

In the Evenki grammars, *aŋi* is described as a placeholder, though this exact term is not always used. In Konstantinova (1964: 265) it is called a demonstrative particle meaning ‘whatsitsname, something’. In Bulatova & Grenoble (1999: 24, 26) it is classified as an interrogative pronoun as well as a placeholder, and its use in both nominal and verbal roots is described. In the corpus texts, it is pronounced as either *aŋə* or *aŋi*. It seems to be more frequently pronounced as *aŋə* when it is used independently, without any affixes. Furthermore, the stem is sometimes shortened to *aŋ*, without the final vowel. However, (10) shows that the final vowel is not just a connecting or epenthetic vowel (otherwise the form would be **aŋ-tiki*: and not *aŋi-tki*:). Prosodically, *aŋi* is often followed with a pause. However, this can be justified by the speaker actively trying to recall the target word. Generally speaking, intonation in Evenki is understudied (see, for example Morozova & Androsova 2019). Therefore, I will not go into greater detail regarding intonation.

2.1.1 Functions as a placeholder

aŋi is used widely if the speaker cannot recall the exact word to ensure the fluency of the narrative. For example, in (7) there are two instances of *aŋi* for two nouns, which are both repaired on the spot. In (8), the first occurrence of the placeholder is repaired but the second is not.

- (7) *aŋi-l* *bəjə-l* *aŋi-l-tin* *igi-l-tin*
 whatsitsname-PL person-PL whatsitsname-PL-3PL.POSS voice-PL-3PL.POSS
do:ldi-w-ra-∅
 hear-PASS-NFUT-3PL
 ‘The whatsitsnames, the voices of whatsitsnames, of people were heard.’
 (G. K. Lapuko, Tura, 2008)
- (8) *umno: aŋi-wa* *bagdakə-wə* *aŋi-∅-m*
 once whatsitsname-ACC wild.reindeer-ACC whatsitsname-NFUT-1SG
 ‘I once did that thing to [killed] a whatsitsname, a wild reindeer.’ (S. M. Andreyeva, Strelka-Chunya, 2007)

2.1.2 Restrictions on the target word

As (9), (10), and (11) show, *aŋi* can substitute for both nouns (including proper nouns) and verbs.

- (9) *həwəki: hələ aŋi-l-duk* *sʲita-l-duk* *o-dʲa-fki:* *bi-sʲə*
 god INTJ whatsitsname-PL-ABL clay-PL-ABL make-IPFV-PHAB be-PANT
bəjə-l-bə
 person-PL-ACC
 ‘God was making people out of whatsitsname, out of clay.’ (V. Kh. Yoldogir, Chiringda, 2007)
- (10) *nuŋartin=tə aŋi-tki:* *huru-tʃo:-tin* *tar moskwa-tki:*
 3PL=FOC whatsitsname-ALL go.away-PST-3PL that Moscow-ALL
du:-la:-wər=tə
 house-LOC.ALL-RFL.PL=FOC
 ‘They went to whatsitsname, to Moscow, home.’ (G. K. Lapuko, Tura, 2008)
- (11) *tʃaŋit tar tʃaŋit-pa tarə aŋi-wa:t*
 bandit that bandit-ACC that.ACC whatsitsname-IMPER.1PL.INCL
tʃok-na:-ya:t
 kill-PRGRN-IMPER.1PL.INCL
 ‘Let us do that thing, let us go and kill that bandit (=bear).’ (S. M. Andreyeva, Strelka-Chunya, 2007)

In (12), it replaces an adjective: the speaker could not come up with the Evenki word and switched to Russian.

- (12) *aməgən=ta on aŋə skolskij*
 saddle=FOC how whatsitsname slipper.R
 ‘The saddle is how, whatsitsname, slippery.’ (I. K. Uvachan, Tutonchany, 2008)

There are no examples of *aŋi* replacing a numeral, a quantifier, or a postposition in our corpus or in the IEA RAS corpus.

In (13), *aŋi* may be considered to be replacing an adverb *duga* ‘in summer’. It is the only example of that kind in our corpus.

- (13) *hulaki:l koŋnomo-l aŋi duga o duga*
 fox-PL black-PL whatsitsname in.summer INTJ in.summer
aŋi-wki:-l
 whatsitsname-PHAB-PL
 ‘Black foxes whatsitsname, in summer, oh, in summer they usually do that thing.’ (L. V. Mikhaylova, Tura, 2008)

In (14), *aŋi* has the same affixes as the personal pronoun following it (note that the 3rd person pronoun form in Evenki has a possessive affix historically, which behaves just like a normal possessive suffix in nominal forms). However, it would be strange for a placeholder to replace a personal pronoun. Perhaps, the speaker wanted to say “When we were going past her grave...” and then said simply “When we were going past her...”. It is the only example in our corpus where the speaker uses a pronoun to “repair” the placeholder.

- (14) *tara aŋi-li:-n nuŋan-duli:-n*
 that.ACC whatsitsname-PROL-3SG.POSS 3SG-PROL-3SG.POSS
ŋənə-də-wun e:kun=məl təpkə-l-də-n
 go-PSIM-1PL.EXCL what=INDEF shout-INCH-NFUT-3SG
 ‘When we were going past whatsitsname, past her, something started to shout.’ (G. K. Lapuko, Tura, 2008)

2.1.3 Functions other than those of a placeholder

aŋi is sometimes used as an interjective hesitation marker as in Hayashi & Yoon (2010), when the speaker cites the direct speech of a character:

- (15) *tuy ŋikə-rə-n=da: gun-ə-n aŋi*
 thus do-NFUT-3SG=FOC say-NFUT-3SG whatsitsname
 ‘Having done this, (he) said: whatsitsname...’ (V. Kh. Yoldogir, Chiringda, 2007)

Sometimes, *aŋi* is used at the beginning of a new sentence (16) or at the end of a sentence (17) with seemingly no syntactic role or any actual placeholder function, being an interjection, marking hesitation and/or introducing a new topic.

- (16) *aŋe* *dʻur anŋapi:l uzə* *ŋənə-rə-∅*
 whatsitsname two year-PL already.R go-NFUT-3PL
 ‘Well, two years have already passed.’ (I. I. Tsurkan, Yerbogachyon, 2016)

- (17) *ŋinakin-tin əmə-hi-lɔi-sʲa* *aŋe*
 dog-3PL.POSS come-INCEP-COMIT-PANT whatsitsname
 ‘Their dog came with them, well.’ (V. Kh. Yoldogir, Chiringda, 2007)

In some of these examples, *aŋ* has a focus marker =*kə*:

- (18) *ə-kəldu* *ŋikagda ə-kəldu* *ŋi:-wə=də:* *aŋ=kə*
 NEG-IMPER.2PL never.R NEG-IMPER.2PL who-ACC=FOC whatsitsname=FOC
abizət-tə
 offend.R=PNEG
 ‘Never, well, never offend anybody.’ (T. A. Bogdanova, Potapovo, 2011)

According to Idiatov (2007: 300), who follows Bulatova & Grenoble (1999: 24), *aŋi* can be used as an interrogative word. However, such usages are lacking in our corpus.

2.1.4 Mirroring the grammatical shape of the target word

The questions of this section are: whether wordforms with *aŋi* can have all possible slots filled in; and which slots are copied from the target word. I must emphasize the fact that we cannot be 100% sure that the word recalled by the speaker is actually the target word. However, it will be our assumption. First, I will look into the slots of nominal and verbal wordforms. For nominal wordforms, there are no examples of *aŋi* taking the alienable possession suffix in our corpus. However, there are no examples where the target word is then recalled and actually has the alienable possession affix, either. Other slots can also be filled in. For example, in (19), a derivational intensifier affix is used together with the case and number suffixes.

- (19) *irəktə-l-ə* *aŋi-ka:ku:-r-ə* *o-dʻa-n*
 larch-PL-ACCIN whatsitsname-INTS-PL-ACCIN make-FUTCNT-3SG
 ‘He will make larches, whatsitsnames...’ (S. P. Mukto, Uchami, 2014)

In our corpus, intensifiers are the only non-inflectional affixes which are used in nominal *añi* wordforms.

As regards verbs, there are no examples of the derivational slot (including intensifiers) filled in for the *añi* verbal wordforms. Furthermore, the voice slot also remains empty in the corpus examples, although there are elicited sentences where the speaker uses a wordform with a non-empty voice slot (20).

- (20) *ə-do:-tin* *añi-dʰə-rə* *isə-w-dʰə-rə=do:*
 NEG-CVPUHP-3PL whatsitsname-IPFV-PNEG see-PASS-IPFV-PNEG=FOC
loku-sa-dʰa-ra=da:
 hang-STAT-IPFV-PNEG=FOC
 ‘So that they will not do that thing, be seen, hang.’ (S. P. Mukto, Uchami, 2014)

Aspect and mood/tense slots are, on the contrary, often filled. In (21), *añi* has non-empty aspect and tense slots, and in (22) the aspect and the mood (imperative) slots are filled.

- (21) *patom bu* *luhu* *añi-ñnə-rə-w*
 then.R 1PL.EXCL all.the.time whatsitsname-HAB-NFUT-1PL.EXCL
luhu
 all.the.time
 ‘Then we would all the time do that thing.’ (I. K. Uvachan, Tutonchany, 2008)
- (22) *dʰəm-mu:-l-mi* *añi-ñna-kal* *gu:-sə:*
 eat-DES-INCH-CVCOND whatsitsname-HAB-IMPER.2SG say-PANT
əri-ñ-mə-w *tuge: sʰiwu-ñna-kal*
 this-IND.POSS-ACC-1SG.POSS so lick-HAB-IMPER.2SG
 ‘If you get hungry, he said, do that thing, lick this your <paw> so.’ (V. K. Udygir, Ekongda, “The man and the bear-relative”) (IEA RAS³)

There are examples of participial (23) and converbial (24) forms with *añi*:

- (23) *bi tar do:ldi:-Ø-m* *añi-dʰə-ri-l-wə*
 1SG that hear-NFUT-1SG whatsitsname-IPFV-PSIM-PL-ACC
bu:ñi:-dʰə-ri-l-wə *straf* *ñə:ləwsʰi=ko:*
 howl-IPFV-PSIM-PL-ACC horrible.SLIP.R horrible=FOC
 ‘I heard doing that thing, howling, [it was] horrible.’ (S. M. Andreyeva, Strelka-Chunya, 2007)

³http://corpora.iea.ras.ru/corpora/describe_text.php?id=43

- (24) *e:-ja=wəl* *e:-dʒə-nə* *horol-dʒo-fki:* *tadu:=wər*
 what-ACCIN=INDEF what-IPFV-CV SIM whirl-IPFV-PHAB there=RFL.PL
ŋa:lə-l-dʒi-ji ***əŋ-dʒa-na***
 arm-PL-INSTR-RFL whatsitsname-IPFV-CV SIM
 ‘Doing something, he is whirling there, doing that thing with his arms.’
 (V. N. Udygir, Ekongda, 2007)

Clitic slots can be filled in *əŋi* nominal (25) and verbal (26) wordforms.

- (25) ***əŋi-l=da:*** *ŋi* *əmkə-n=də:* *asʲin bi-so:-n*
 whatsitsname-PL=FOC not.R cradle-3SG.POSS=FOC NEG be-PST-3SG
 ‘He had no whatsitsnames, no cradle.’ (S. P. Mukto, Uchami, 2014)
- (26) *tug=do: e:ku-r=wəl* *nunʌr-wa-tin* ***əŋi-tʃa=hintʃa***
 so=FOC what-PL=INDEF 3PL-ACC-3PL.POSS whatsitsname-PANT=PROB
huru-p-tʃo
 go.away-TR-PANT
 ‘So something must have done that thing to them, carried them away.’ (G. K. Lapuko, Tura, 2008)

As demonstrated by previous examples, *əŋi* can take nominal or verbal suffixes, mirroring the shape of the target. (11) shows that the mirroring can be partial: the inflectional affix (-*wa:t* ‘IMPER.1PL.INCL’) is copied whereas the derivational one (-*na*: ‘PRGRN’) is not. However, there are some examples where *əŋi* is used with no suffixes at all. In (27), both strategies are followed. It is worth noting that the same speaker also uses verbal affixes with *əŋi* in other examples.

- (27) *bi nunʌnman əŋi* *sabira-Ø-m* *i* *kuŋakan*
 1SG 3SG.ACC whatsitsname gather-NFUT-1SG and.R child
əŋ-du: *hapoki-ka:n-tiki: rezin-tiki: resinowij-du:*
 whatsitsnam-DAT.LOC boot-ATTEN-ALL rubber-ALL rubber-DAT.LOC
hisʲi-hi-ŋnə-Ø-m
 shove-INCEP-HAB-NFUT-1SG
 ‘I whatsitsname, gathered it (the antenna) and put it into whatsitsname, child’s rubber boot.’ (L. D. Utukogir, Khantayskoye Ozero, 2011)

In (28), it is hard to distinguish between the placeholder and the interjective use of *əŋi*.

- (28) *ekun ta-wər* *gun-dʻə-rə-n* *əŋi* *lutfa-l*
 what that-RFL.PL say-IPFV-NFUT-3SG whatsitsname Russian-PL
kokoldo-l-tin *zə*
 mitten-PL-3PL.POSS FOC.R
 ‘What’s that? – he says. – (It’s) whatsitsname, Russians’ mittens.’ (L. A. Yeryomina speaking to M. D. Turskaya, Khantayskoye Ozero, 2011)

If we denote the suffix set of *əŋi* with *AS* and the suffix set of the target word with *TS*, we can theoretically consider the following cases:

1. $AS = TS$ (the sets are equal)
2. $AS \subseteq TS$, $AS \neq TS$ (AS is a strict subset of TS)
3. $TS \subseteq AS$, $AS \neq TS$ (TS is a strict subset of AS)
4. $AS \cap TS = \emptyset$ (AS and TS have nothing in common)
5. $\text{not}(AS \subseteq TS)$, $\text{not}(TS \subseteq AS)$, $AS \cap TS \neq \emptyset$ (there is a non-empty intersection of suffixes in AS and TS)

Table 4 shows the distribution of these cases in our corpus for nominal and verbal forms separately.

Table 4: Suffix mirroring according to the corpus

	nominal forms		verbal forms	
1. equal sets	85	(45.2%)	51	(37.8%)
2. AS is a strict subset of TS	18	(9.6%)	19	(14.1%)
3. TS is a strict subset of AS	9	(4.8%)	6	(4.4%)
4. AS and TS have nothing in common	0		0	
5. there is a non-empty intersection of suffixes in AS and TS	0		0	
6. other cases (no target form, slips of the tongue etc.)	76	(40.4%)	59	(43.7%)
total	188	(100%)	135	(100%)

We compare the suffix sets only in case the Evenki target word was actually used. Therefore, the cases when the speaker did not actually pronounce the target word or shifted to Russian are included into “other cases”. However, even when the target word is lacking, the placeholder and interjective uses of *añi* can usually be distinguished with the help of *añi* forms and the context, such as the speaker’s explanations in Russian.

It can be seen that full mirroring occurs in most cases. The cases of partial mirroring can be explained with several reasons:

Case 1: The suffix which is absent from the placeholder form can be better described as derivational rather than inflectional. For example, in (29) the causative suffix in *ju:bdo:n* ‘to make go out’ lacks in the placeholder:

- (29) *to:li: dolbo: baldi:-ñahi:-w ami:-m gənnno:-sa:-n*
 then at.night be.born-CVSIM-1SG father-1SG.POSS fetch-PST-3SG
umuko:n atirka:nmə minə baldi:-da-raki:-w
 one old.woman-ACC 1SG.ACC be.born-IPFV-CVCOND-1SG
añi-da:-n *ju:-b-do:-n*
 whatsitsname-CVPURP-3SG go.out-CAUS-CVPURP-1SG
 ‘Then, at night, when I was born, my father went to fetch one old woman so that she would do that thing, make me go out. (A. I. Pankagir, Ekongda, 2007)’

Case 2: The placeholder can have intensifier and diminutive suffixes, serving the aims of emphasis (30, 31).

- (30) *itʃə-t-mi=ka tuyi añi-ka:kun tarə*
 see-DUR-CVCOND=FOC so whatsitsname-INTS TOT.ACC
itʃə-t-mi=do: gun-dəño:-n fsʲigda bəjə gun-dəño:-n
 see-DUR-CVCOND=FOC say-FUT-3SG always.R person say-FUT-3SG
tar wojennij nuñan
 that military 3SG
 ‘When someone sees – (he is) very whatsitsname, when someone sees, they will say... A person will always say that he is a military man.’ (G. K. Lapuko, Tura, 2008)

- (31) *a tar tuli-gido-n talu aŋi*
 and.R that outside-SIDE-3SG.POSS birch.bark whatsitsname
aŋi-s'i-ka:ku:n bi-fki: tar
 whatsitsname-ATR-INTS be-PHAB that
 ‘And on the outside there is usually birch bark whatsitsname, with
 whatsitsname.’ (S. P. Mukto, Uchami, 2014)

Case 3: the meaning of the suffix which is absent from the target form is incorporated in the word stem. For example, in verbal forms the most frequently omitted suffixes are the so-called inchoative, inceptive, durative, and stative aspects (called subaspects in Bulatova & Grenoble 1999: 30). In (32) the target word does not have the inchoative suffix, whereas the placeholder has it:

- (32) *taduk aŋ nuŋan aŋi-l-da-n*
 then whatsitsname 3SG whatsitsname-INCH-NFUT-3SG
himu:rga-ra-n tar faman famani-tka:n=tə
 become.silent-NFUT-3SG that shaman shaman-CHILD=FOC
 ‘Then she started doing that, became silent, that shaman, little
 shaman.’ (G. K. Lapuko, Tura, 2008)

We can suppose that the stem *himu:rga-* ‘become.silent’ already has an inchoative meaning, so it is not necessary to use the inchoative suffix. However, to prove this, a separate survey on the lexical restrictions for the stems in question should be carried out.

Finally, the target word used by the speaker may sometimes be not the target word originally intended. Thus, partial mirroring can show the speaker’s doubts, whereas the original intention cannot be retrieved.

2.1.5 Frequency

In our data, *aŋi* is quite frequent, occurring 350 times in a corpus of about 27,700 running words, i. e. about 12.6 times per one thousand words. This is much higher than the rates cited in Podlesskaya (2010) (5–6.7 per thousand), which may be explained by the lack of proficiency in some speakers. Actually, most speakers do not use the Evenki language in their daily life, and text generation presents difficulties for some of them, with lexical production being more challenging than following grammar rules. Many passive Evenki speakers have no trouble

declining a noun or conjugating a verb, including participial or converbial forms. However, recalling the exact lexemes demands much more effort from them. As a result, texts produced by such speakers might be grammatically correct but have nearly all meaning words replaced by placeholders.

2.1.6 Dialectal variation

According to the corpus, *ani* is used in the dialects of the Southern and Northern dialect groups: Sym, Podkamennaya Tunguska, and Ilimpeya dialects. However, the word seems to be absent from the Far Eastern Tugur-Chumikan and Sakhalin dialects: it does not occur in texts, and the speakers do not recognize it in context. We have little spoken data from other Eastern dialects.

2.1.7 Possible source and evidence from related languages

In Idiatov (2007: 299–302), the functions of *ani* as both a placeholder word and an interrogative pronoun are discussed. The author also states a hypothesis about its origin, tracing it to an old genitive form of a word originally meaning ‘thing’ or, alternatively, “a fossilized genitive of the ... ‘what’ root” (which can be found in other interrogative pronouns). *ani* can also be found in the Udeghe language, a relative of Evenki (Nikolaeva & Tolskaya 2001: 361, 362). In Udeghe, the target words for *ani* can be both verbs and nouns (including proper names), and *ani* tends to mirror the grammatical shape of the target. Furthermore, it can function as an indefinite pronoun.

In Uilta (Orok), a Southern Tungusic language, *anyu* is a placeholder word (Idiatov 2007: 301, citing Cincius 1975/77: I: 45). According to our Uilta field data, its target words can be both verbs and nouns, just like in Udeghe, and it also has the mirroring feature.

2.2 *unjun*

unjun is a named-entity placeholder. To my knowledge, this stem has not been reported in Evenki grammars yet.

2.2.1 Functions as a placeholder

In the texts, *unjun* substitutes proper nouns: names of people (33, 34) or animals in tales (35, 36), as well as geographical terms (37, 38).

- (33) *uŋun tare wot gus^ʃə-ja aŋi-tin am*
 whatsitsname2 that so.R Gusya-COLL whatsitsname-3PL.POSS SLIP
d̄edufka-tin ha:-Ø-ndə kosin-mo
 grandfather-3PL.POSS know-NFUT-2SG Kosin-ACC
 ‘Whatshisname, do you know, whatsitsname, the grandfather of the
 Gusya’s family, Kosin?’ (P. K. Pankagir speaking to V. P. Khukochar,
 Tutonchany, 2008)
- (34) *bəʃə uŋun-mə d̄əwit-pa ha:-Ø-ndə*
 person whatsitsname2-ACC David-ACC know-NFUT-2SG
 ‘Friend, do you know whatshisname, David?’ (L. F. Utukogir speaking to
 A. D. Chempogir, Khantayskoye Ozero, 2011)
- (35) *tar huru-rə-Ø d̄əgi huwuli-n uŋun-dula: tarə*
 that go.away-NFUT-3PL bird ALL-3SG.POSS whatsitsname2-LOC.ALL that
d̄antaki-la:
 wolverine-LOC.ALL
 ‘So all birds went to whatshisname, to the wolverine.’ (V. Kh. Yoldogir,
 Chiringda, 2007)
- (36) *tar gu:-sə uŋun he:ŋa:n gu:-sə:*
 that say-PANT whatsitsname2 burbot say-PANT
 ‘So said whatshisname, the burbot said. (it is not clear whether *uŋun*
 refers to the burbot or to the other character, the fox)’ (“Burbot and fox”,
 V. T. Yoldogir, Chiringda, 2011 (IEA RAS)⁴)
- (37) *e: nu ər-tiki: zə uŋun-tiki: bi-nə bi-rkə-Ø*
 INTJ INTJ this-ALL FOC whatsitsname2-ALL be-CV SIM be-PROB-3SG
nawerna ərə walok-tuk
 perhaps that Valyok-ABL
 ‘Yes, it <the settlement discussed previously> was perhaps in the
 direction of whatsitsname, in the direction from Valyok.’ (L. A.
 Yeryomina speaking to M. D. Turskaya, Khantayskoye Ozero, 2011)
- (38) *uŋun-duk ə aŋi-l d̄a-li-n*
 whatsitsname2-ABL INTJ whatsitsname-PL relative-PL-3SG.POSS
aŋi-duk gulə-l-duk əmə-rə-Ø=də:
 whatsitsname-ABL house-PL-ABL come-NFUT-3PL=FOC

⁴http://corpora.iea.ras.ru/corpora/describe_text.php?id=35

isʲə-no:[-rə-∅]

aŋi-la:

see-PRGRN[-NFUT-3PL] whatsitsname-LOC.ALL

‘His whatsitsname, relatives came from whatsitsname, village, came and went to see to whatsitsname.’ (G. K. Lapuko, Tura, 2008)

There is actually one example from Mutoray (Southern dialect group) where *uŋun* is probably a placeholder for a common noun meaning ‘hole in the ice’ and not a proper noun. However, it is hard to judge from the context as the speaker does not actually pronounce the word. Importantly, the narrator tells the tale in the presence of her husband, who makes comment to what she says, so this can also be regarded as a sort of a dialogue:

(39) *i tima nu bira uŋun=kə həgdi-məmə*
and.R in.morning.SLIP INTJ.R river whatsisname2=FOC big-INTS

i globokaj oj sunta mu:
and.R deep.R INTJ deep water

‘In the morning, well, the river, whatsitsname [a hole in the ice], large and deep water.’ (I. M. Yastrikova, Mutoray, 2007)

It can be seen from most of these examples that, when *uŋun* is used in dialogues, the speaker often asks the interlocutor to help recall the missing target. This brings *uŋun* into a sharp contrast with *aŋi*. With *aŋi*, the target word can usually be restored from the context, even if not pronounced. With *uŋun*, it is important to recall the exact name of a person or a place. This is perhaps the reason for its being used in dialogues with the inhabitants of the same settlement, who have the same background knowledge. We can say that *aŋi* is a placeholder for a word, whereas *uŋun* is a placeholder for the notion: *aŋi* helps make the narrative fluent acting like a joker, whereas *uŋun* brings the interlocutor’s attention to the word being missed.

Similar conclusions on the interactional use of a placeholder in Estonian have been made in Keevalik (2010).

2.2.2 Restrictions on the target constituent

As shown before, the target constituent is a proper noun.

2.2.3 Functions other than those of a placeholder

None have been found.

2.2.4 Mirroring the grammatical shape

uŋun mirrors the grammatical shape of a noun, copying case markers. Proper nouns in Evenki do not usually have alienable or inalienable possession markers. Plural number markers are theoretically possible but rare. Therefore, there is no surprise that *uŋun* has no number or possession markers in our data.

2.2.5 Frequency

In our data, *uŋun* is quite rare, occurring only 13 times in a corpus of about 27,700 running words. As shown above, it occurs mainly in dialogues between several Evenki speakers, and such dialogues are rare in our corpus. In the IEA RAS corpus, it only occurs 2 times in a corpus of 121,286 running words (the majority of the texts are, however, written texts), both times referring to the name of an animal in a tale.

2.2.6 Dialectal variation

In our corpus, *uŋun* is only found in the Ilimpeya dialect texts (Northern dialect group), and in Podkamennaya Tunguska texts (Southern dialect group). However, a speaker from Nakanno (Irkustk oblast, Yerbogachyon dialect, Northern dialect group) recalled this word being used in the past by elderly people, although she was not entirely sure. Speakers of the Tugur-Chumikan dialect (Khabarovsk krai, in the Far East of Russia, Eastern dialect group) did not accept the word.

2.2.7 Possible source and evidence from related languages

We cannot trace the origin of *uŋun*, and it is not mentioned in the comparative dictionary (Cincius 1975/77).

In Negidal, a close relative of Evenki, *uŋun* is used as a general purpose placeholder for both nominal and verbal stems, mirroring the target word grammar, e. g. in a text from the Negidal corpus (Pakendorf & Aralova 2017):⁵

(40) Negidal

net baka-ja-βun uŋun-ma

no.R find-NFUT-1PL.EXCL HESIT-ACC

‘No, we found a whatchamacallit.’ (A. V. Kazarova, Vladimirovka, 2017)

⁵<https://elar.soas.ac.uk/Record/MPI1084918>. The original glossing of *uŋun* as HESIT is preserved.

(41) Negidal

itce-mi hoŋte mesto-duki-n tak možno uŋun məjga-tca
see-SS.COND other place.R-ABL-3SG so.R be.able.to.R HESIT think-PST
možno tʃto rjukzak=to minə-βə mu-duk-in
be.able.to.R that.R rucksack.R=PTL.R 1SG.OBL-ACC.DEF water-ABL-3SG
uŋun-tca ʃaβu-tca-tca
HESIT-PST take-RES-PST

‘Looking from the side one could think that it was the rucksack that was holding me up.’ (A. V. Kazarova, Vladimirovka, 2017)

In the dictionary of Even, another close relative (Robbek & Robbek 2005: 271), *uŋ* is glossed as “interjection” which has the meaning “pause”. Matić (2008) shows that it is typical for the Eastern Even dialects. Arkady Taraboukine, a native speaker of Even born in Beryozovka and living in Anyuysk, gave the following examples of how it could be used.

(42) Even

tipiw bi bəri-ri-w uŋ-u halka-w
yesterday 1SG lose-NFUT-1SG whatsitsname2-ACC hammer-ACC

‘Yesterday I lost whatsitsname, a hammer.’ (A. Taraboukine, Beryozovka, 2020)

(43) Even

bi uŋ-di-m mərgət-tʃi-m
1SG whatsitsname2-PRS-1SG think-PRS-1SG

‘I am doing that thing, thinking.’ (A. Taraboukine, Beryozovka, 2020)

Therefore, in Even, just like in Negidal, *uŋ* is used as a general purpose placeholder for both nominal and verbal stems, mirroring the grammatical features of the target word.

We have no information of the stem *uŋ(un)* being used in Southern Tungusic languages.

To sum up, the *uŋ(un)* stem can be found in Northern Tungusic languages with its function ranging from a general placeholder in Even and Negidal to a proper noun placeholder in some Evenki dialects. Interestingly, it was not found in the Tugur-Chumikan dialect (at least in elicitation experiments), otherwise quite close to the Even language both geographically and linguistically. In all these languages, *uŋ(un)* mirrors the grammatical features of the target word.

2.3 *e:(kun)*

e:(kun) is an interrogative pronoun meaning ‘what/who’. *e:(kun)* can also have a shortened stem *e:-* (Konstantinova 1964: 137), mostly in oblique forms. In Poppe (1977) as well as in Cincius (1975/77: I: 286), *-kun* is considered to be a morpheme, with *e:* being the original stem. According to Idiatov (2007: 303–308), it can refer to objects, animals and to humans but only when questioning their “kind” (for example, their belonging to a clan). The meaning is different in various dialects, with Vanavara dialect (Southern dialect group) speakers more approving of its referring to humans. Indefinite and negative pronouns are formed from the interrogative pronominal base (Bulatova & Grenoble 1999: 25). *e:-/ə-* is also the stem of the question verb ‘what to do?’

2.3.1 Functions as a placeholder

e:(kun) serves as a placeholder for both nominal and verbal stems:

- (44) *dʉktʃa e:kun-ma-n* *ham-na-∅* *urkə-wə-n*
 tent what-ACC-3SG.POSS close-NFUT-3PL door-ACC-3SG.POSS
ham-na-∅
 close-NFUT-3PL

‘They close the tent’s whatsitsname, its door.’ (N. G. Yegorova, Yerbogachyon, 2016)

- (45) *ə-wki: minə ə-ra əyi-fko:n-ə*
 NEG-PHAB 1SG.ACC what-PNEG play-CAUS-PNEG
 ‘She did not let me do that thing, play.’ (V. I. Udygir, Kislokan, 2008)

2.3.2 Restrictions on the target constituent

In the examples considered, *e:(kun)* can substitute for both nominal and verbal roots.

2.3.3 Functions other than those of a placeholder

The functions of *e:(kun)* as an interrogative, indefinite or negative pronoun have already been discussed.

2.3.4 Mirroring the grammatical shape of the target word

The examples show that *e:(kun)* mirrors the grammatical shape, sometimes partially. Due to the scarcity of the data, I will not discuss the percentage of partial vs full mirroring. Like *əŋi*, *e:(kun)* is sometimes used as an interjective hesitation marker:

- (46) *e:kun=ka nawerna ŋinaki-r kiki-rka-l kuji:-ko:t-tʃə-nə-l*
 what=FOC perhaps.R dog-PL bite-PROB-3PL fight-INTS-IPFV-CVSIM-PL
 ‘Well, perhaps the dogs bit <it>, when they were fighting.’ (V. N. Saygotin, Bolshoye Sovetskoye Ozero, 2007)

An anonymous reviewer suggests that it could be a calque of Russian *что же* ‘what so’, used sometimes as an interjective hesitation marker. I think it might be difficult to prove or confute it with little data on how discourse markers are generally calqued in Evenki. However, I still think it is not so. *Chto zhe* sounds quite formal, and is not wide-spread in Russian colloquial speech. The speakers who use *e:(kun)=ka*, though bilingual in Evenki and Russian, are not exposed much to the formal Russian style. Actually, prosodically and functionally *e:(kun)=ka* more closely resembles Russian *это* *это* ‘this’ used very often as a hesitation marker, by native Evenki speakers, too, when they are speaking Russian.

Another important function of *e:(kun)* is its use when listing several objects of a kind, at the end of such enumerations, e. g.:

- (47) *muldi:-ka-r əragəri-t e:-wa=da doku-d’a-mi:=da*
 not.be.able-NMLZ-PL at.all-ADVZ what-ACC=FOC write-IPFV-CVCOND=FOC
e:-d’a-mi:=da
 what-IPFV-CVCOND=FOC
 ‘(They were) not able at all to write anything or do such things.’ (V. Kh. Yoldogir, Chiringda, 2007)

- (48) *umuko:-riktə əŋi tar ahi: mo:ni-n d’u-du:*
 one-LIM whatsitsname that woman RFL-3SG.POSS tent-DAT.LOC
bi-d’ə-ri tari-rikta bi-ŋki:-n sat-ti-fka:n-d’ə-nə
 be-IPFV-PSIM that-LIM be-PSTITER-3SG tea-VBLZ-CAUS-IPFV-CVSIM
ə-d’ə-nə
 what-IPFV-CVSIM
 ‘Only one whatsitsname, that woman, who was in her own tent, only she gave tea to drink and did such things.’ (E. K. Khukochar, Tura, 2014)

- (49) *tozə ʃimŋaka:n=li e:kun=li*
 also.R tale=Q.R what=Q.R
 ‘Also a tale or what.’ (I. I. Tsurkan, Yerbogachyon, 2016)

In (50), both stems of *e:(kun)* are used: *e:kun* as a placeholder and *e:-* in the enumeration.

- (50) *walok-tula: tozə e:ku-r-wa oldo-ŋi-l-wa e:-l-wa*
 Valyok-LOC.ALL also.R what-PL-ACC fish-IND.POSS-PL-ACC what-PL-ACC
əmə-wu-pki:-l bi-tʃo-l
 come-TR-PHAB-PL be-PANT-PL
 ‘They also carried whatsitsname, fish and such to Valyok.’ (L. A. Yeryomina speaking to M. D. Turskaya, Khantayskoye Ozero, 2011)

The enumeration function of *e:kun* is close to what is described for Udeghe in Tolskaya & Tolskaya (2008). In Udeghe a repetition of the verbal form with the interrogative ‘what’ is used in the formation of open alternative questions.

2.3.5 Frequency

In our data, *e:kun* and *e:-* are used 32 times (out of 27,700 running words) in the function of placeholders.

2.3.6 Dialectal variation

e:(kun) as a placeholder is used in texts from the Bolshoye Sovetskoye Lake, Sovrechka, Ekongda, and Kislokan (Ilimpeya dialect, Northern group), Yerbogachyon (Yerbogachyon dialect, Northern group), Sym and Bely Yar (Sym dialect, Southern group), and Poligus (Poligus dialect, Southern group). There is an intersection between *əŋi* and *e:kun* areas, although texts from Bolshoye Sovetskoye Lake and Sovrechka lack the otherwise very frequent *əŋi*, which suggests some dialectal variation.

2.3.7 Possible source and evidence from related languages

e:(kun) can be both used as a normal question word and as a placeholder by the same speakers. When used as a placeholder or an interjective hesitation marker, the *=ka* focus particle is sometimes attached, like in (46) or in (51):

- (51) *itʃə:-rə-w* *e:kun-ma e:kun-ma=ka* *kiran-tʃika:n-mə*
see-NFUT-IPL.EXCL what-ACC what-ACC=FOC eagle-CHILD-ACC
toʃo-t-tʃə-ri:-wə *dagda-du:*
sit-DUR-IPFV-PSIM-ACC pine-DAT.LOC

‘We saw *whatsitsname*, *whatsitsname*, a little eagle sitting on a pine.’ (G. P. Boyarin, Sym, 2009)

3 Conclusions

Evenki speakers employ various placeholders that mirror the grammatical form of the target word. These placeholders have different discourse functions: *əni* and *e:kun* are general purpose placeholders which provide speech fluency, whereas *uŋun* requires interaction from the interlocutor. There seems to be no difference between *əni* and *e:kun* when used as placeholders, although there is a hypothesis at a dialectal variation. *əni* and *e:kun* have usages other than those of a placeholder, which is typologically typical for placeholders, whereas *uŋun* is only registered as a placeholder in our materials.

uŋun, a placeholder with obscure etymology, has been found in western Evenki dialects, in Even and in Negidal. The westernmost and easternmost idioms have no contact nowadays, which suggests an ancient origin of *uŋun*. Interestingly, according to the considered materials, it is only in Evenki that *uŋun* has a special restriction on the target word, being a proper noun placeholder. *əni* is also quite wide-spread, as it is present in western Evenki dialects (i. e., in the Northern sub-branch of the Tungusic family) and in two languages of the Southern sub-branch: Uilta, and Udeghe. However, there is quite little data on placeholders in Tungusic languages in general. It is urgent to study discourse and, specifically, the use of placeholders in the Tungusic languages, especially given their endangered status and the decline of communication in these languages.

According to the corpus data, there are some regularities in placeholders copying intensifier affixes from the target word but not other derivational affixes, or, for example, voice slots. Nevertheless, it should be studied in elicitation experiments whether such copying is theoretically possible. The corpus data also suggests restrictions on the part of speech of the target word even for general purpose placeholders (*əni* and *e:kun*), which should also be tested with elicitation. However, direct elicitation experiments for the placeholders proved to be inefficient due to the low status of these words. A different experiment design, such as asking to fill in the gap, should be attempted. Another important lacuna is the prosodic features of the placeholders. In this paper, I do not look into prosodic

features of the placeholder verbs in great detail. It should also be studied using the available oral corpora with annotated multimedia content.

Non-standard abbreviations

Russian words are indicated with an R. Grammatical abbreviations include:

ACCIN	indefinite accusative	PANT	anterior participle
ATR	attributive	PHAB	habitual participle
FUTCNT	immediate future	PNEG	negative participle
CVCOND	conditional converb	PRGRN	peregrinative
CVPURP	converb of purpose	POSS	possession
CVSIM	converb of simultaneous action	PSIM	participle of simultaneous action
INCEP	inceptive	PSITER	iterative past
INCH	inchoative	TR	transitivizer
IND.POSS	indirect (alienable) possession		

Acknowledgements

This work was supported by the RSF (grant no 17-18-01649). I would like to thank the anonymous reviewers and the editors of the present volume for their valuable remarks.

References

- Bulatova, Nadezhda & Lenore Grenoble. 1999. *Evenki*. Munich: Lincom Europa.
- Cincius, Vera I. et al. 1975/77. *Sravnitel'nyj slovar' tunguso-man'chzhurskih jazykov: materialy k jetimologicheskomu slovarju [A comparative dictionary of the Tungus-Manchu languages: Materials toward an etymological dictionary]*. 2 vols. Leningrad: Nauka.
- Corpora at IEA RAS*. 2019. Retrieved from: <http://corpora.iea.ras.ru/corpora/structure.php>. (Accessed 08.10.2019.)
- Ethnologue: Evenki*. 2019. Retrieved from: <https://www.ethnologue.com/language/evn>. (Accessed 08.10.2019.)

- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 1–10. Amsterdam: John Benjamins.
- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction. In Nino Amiridze, Boyd Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 33–66. Amsterdam: John Benjamins.
- Idiatov, Dmitry. 2007. *A typology of non-selective interrogative pronominals*. Antwerp: Universiteit Antwerpen. (Doctoral dissertation).
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative “see”. In Nino Amiridze, Boyd Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Konstantinova, Olga A. 1964. *Evenkijskij jazyk: fonetika i morfologija [The Evenki language: Phonetics and morphology]*. Moscow, Leningrad: Nauka.
- Matić, Dejan. 2008. A note on hesitative forms in two Even dialects. In O.A. Osipova, A.V. Didenko & E.A. Krjukova (eds.), *Sravnitel’no-istoričeskoe i tipologičeskoe izučenie jazykov i kul’tur. Materialy meždunarodnoj naučnoj konferencii XXV Dul’zonovskie čtenija*, 234–239. Tomsk: Tomsk State University.
- Morozova, Olga N. & Svetlana V. Androsova. 2019. Intonacija voprosov v reči amurskih evenkov: osnovnoj ton [The prosody of questions in the Amur Evenks’ speech: Pitch]. *Theoretical and Applied Linguistics* 5(4). 84–96.
- Nedjalkov, Igor. 1997. *Evenki* (Descriptive Grammars). London: Routledge.
- Nikolaeva, Irina & Maria Tolskaya. 2001. *A grammar of Udihe* (Mouton Grammar Library 22). Berlin: Mouton de Gruyter.
- Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct northern Tungusic language of the Lower Amur*. Endangered Languages Archive. Available at <http://hdl.handle.net/2196/b644db81-725c-4031-935c-f33c763df152>. (Accessed 2019.10.08.)
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins.
- Poppe, Nicholas. 1977. On some Altaic designations of location. *Monumenta Serica* 33. 407–423.
- Robbek, Vasiliy A. & Maria E. Robbek. 2005. *Evenko-russkij slovar’ [Even-Russian dictionary]*. Leningrad: Nauka.
- Russian census. 2010. Retrieved from https://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm. (Accessed 2019.10.08.)

- Siberian-Lang corpus*. 2019. Retrieved from <http://gisly.net/corpus>. (Accessed 2019.10.08.)
- Tolskaya, Maria & Inna Tolskaya. 2008. Question particles or what? Open alternative questions in Manchu-Tungusic languages. In Edward J. Vajda (ed.), *Subordination and coordination strategies in North Asian languages* (Current Issues in Linguistic Theory Series 300), 97–108. Amsterdam: John Benjamins.
- Tsumagari, Toshirō. 1992. A basic vocabulary of Khamnigan and Oluguya Evenki in northern Inner Mongolia. *Bulletin of the Institute for the Study of North Eurasian Cultures* 21. 83–103.
- Vasilevich, Glafira M. 1948. *Očerki dialektov jevenkijskogo (tungusskogo) jazyka* [Essays on the grammar of the Evenki (Tungus) language]. Leningrad: Uchpedgiz.
- Whaley, Lindsay. 1998. *Endangered languages*. Retrieved from <http://www.dartmouth.edu/~trg/endangered.html>. (Accessed 2020.06.30.)
- Whaley, Lindsay & Fengxiang Li. 2000. Oroqen dialects. *Central Asiatic Journal* 44(1). 105–130.

Chapter 7

From consonant to tone: Laryngealized and pharyngealized vowels in Udihe

Elena Perekhvalskaya

LLACAN CNRS, France

This article gives a comprehensive analysis of laryngealized and pharyngealized vowels in the Udihe language. Their realization in different Udihe varieties is considered, and their etymology is traced. The classification of Udihe dialects is also discussed. The presence of pharyngealized vowels is one of the most important features that distinguishes the northern dialect cluster from the southern one. The loss of pharyngealized vowels has led to changes in the morphology and syntax of the dialects of the southern cluster. The analysis provides a basis for a complete picture of a dialectal continuum, which includes dialects of Udihe and the closely related Oroch language. The internal mechanisms of the dialectal continuum are presented, taking into account types of pronunciation in neighbouring varieties.

1 Introduction

Udihe (Udeghe, Udege) is a highly endangered Manchu-Tungusic language spoken in the southern part of the Russian Far East. The Udihe live in Khabarovskij Krai (districts: Imeni Lazo, Nanaisky) and Primorskij Krai (districts: Terneiskij, Požarskij, Krasnoarmejskij), and also in the Jewish Autonomous Region. The original name is *Udihe* or *Udie*.¹ The official Russian name is *Udegeiskij jazyk*.² In

¹Until the 1920s, the Udihe did not have a common self-designation but used clan names, usually derived from names of rivers.

²The name *Udege* is the transmission of the self-designation *Udihe*: the pharyngealized element was perceived as a consonant /ɣ/ and was written down with the Cyrillic letter <ɣ>, which reflected the local Russian pronunciation of /ɣ/ as a fricative consonant. In literary Russian <ɣ> denotes a plosive consonant. “The Russian form *Udege* is based, in a certain sense, on a phonetic misunderstanding” (Kormušin 1998: 5).



linguistic literature it is also known as *udeiskij* (Evgenij Šneider 1936) and *udyx-eiskij* (Igor Kormušin 1998). In the 2010 census Udihe named themselves as: *Udie*, *Ude*, *Udegeitsy*, *Udexe*, and *Udexeitsy*.

According to the latest censuses³ (1989, 2002 and 2010), the number of Udihe is constantly decreasing from 1,902 in 1989 to 1,496 in 2010. In 2010, 620 people were registered in the Khabarovskij Krai; 793 people lived in the Primorskij Krai. An additional 83 Udihe were registered outside of these territories, including 42 people in the Jewish Autonomous Region. The census data also reflect the steady decline of the language: according to the 1989 census, Udihe was spoken by 462 people, in 2002 it was 227 people, and in 2010 it was only 103 people. The 2010 census shows a sharp drop in the Udihe competence in the Khabarovskij Krai (from 96 to 16 people).

Traditionally, the Udihe were semi-nomads, moving within a limited territory, each along a particular river and its tributaries, thereby forming territorial groups which usually consisted of several families. The territorial groups are mostly named after the corresponding rivers: (1) Kur-Urmi, (2) Samarga, (3) An-juj, (4) Xungari, (5) Xor, (6) Bikin, (7) Iman, and (8) Sea shore (Namunka). In the 1930s, the Udihe were compelled to become sedentary: each territorial group was settled in a specially built permanent settlement: Kukan (Kur-Urmi), Bira (An-juj), Kun (Xungari), Agzu (Samarga), Gvasjugi (Xor), Sjain, Mitaxeza and Olon (Bikin), Sančixeza (Iman). The less numerous Sea shore Udihe were dispersed. At present, the largest Udihe settlements are: Agzu (Terneiskij district), constituting about 80% of the population of the village; Gvasjugi (Imeni Lazo district; 65% of the population), Krasnyj Jar and Olon (Požarskij district; 55% of the population). Each territorial group is characterized by a specific language variety. Dialectal differences primarily concern phonetics and vocabulary, and to a lesser extent morphology and syntax.

Morphologically, Udihe is an agglutinative language; the agglutination is combined with elements of fusion mainly in verb paradigms. The root, both nominal and verbal, always occupies the extreme left position in a word; it is followed by derivational and inflectional suffixes, which form a chain that can number up to six or seven (in the case of verb forms). In addition to synthetic forms the verb system contains analytic constructions with auxiliary verbs. The verbal negative construction consists of a negative verb and a main verb without any specific con-negative suffixes (for more details see Hölzl 2015).

The peculiarity of Udihe inside the Manchu-Tungusic group is largely due to its phonetics and phonology, primarily the existence of several series of vow-

³For an analysis of the census data, see Perekhval'skaya (2016).

els. The northern dialect cluster has four series of vowels: short, long (including diphthongoids), pharyngealized and laryngealized; the southern dialect cluster has three series: short, long and laryngealized. The phonological interpretation of these vowels is controversial (see Nikolaeva & Tolskaya 2001: 39–41).

The present article contains a comprehensive analysis of these vowels in Udihe dialects. It is shown that they developed out of tri-phonemic complexes of the V-C-V type, which are found in the closely related Oroch language.

When considering complex vowel phonemes, the phonological system of each territorial variety (dialect) is regarded as independent (Trudgill 1985). In each variety the full mode and the allegro modes of pronunciation are taken into account, which makes it possible to show that, roughly, the allegro mode of one variety corresponds to the full pronunciation mode of another variety which, in turn, creates a new allegro mode, etc.

The objectives of the article are 1) to give an overview of Udihe dialects and their clusters; 2) to display the anatomy of the “dialect continuum” by comparison of the modes of pronunciation in each territorial variety; 3) to show the relative character of the synchrony/diachrony dichotomy in a language description; 4) to demonstrate one of the mechanisms of tonogenesis in a previously atonal language.

2 The Udihe: Areal groups and dialects

2.1 Udihe and Oroch

The Udihe language area borders with Nanai, Ulcha and Ewenki, as well as, historically, Manchu dialects. Udihe had rather intensive contacts with these languages. Thus, Kur-Urmi Udihe situated in traditional Ewenki territory underwent significant influence of the latter. Bikin Udihe and Bikin Nanai (Kilen) acquired a number of similar features (Perekhval'skaya 2001). The linguistic border between Udihe, on one hand, and Nanai or Ewenki, on the other, are clear cut. Neither speakers nor linguists hesitate in attributing a variety to one or the other of these languages.

The situation of Udihe and Oroch is different.⁴ As there are no definite linguistic criteria for distinguishing “language” and “dialect”, it is worth considering the ethnic identity of Udihe and Oroch speakers.

⁴In addition, there is Kilen on the Chinese side, which has been heavily influenced by Udihe or Oroch (for details see Hölzl 2018). Negidal, most probably, also had an Oroch substrate (Pevnov 2012).

Traditionally, the Oroch lived along the sea coast and the Tumnin river. Their territory borders the Anjuj and Xungari Udihe area in the West and the Samarga Udihe in the South (see Figure 1). Culturally, the Udihe and Oroch are rather close. While the Nanai, who lived along large rivers Amur and Ussuri, were mainly fishermen, the Udihe and Oroch travelled along small taiga rivers, their main occupation being hunting; fishing and gathering were secondary occupations. Their neighbours in the North, the Ewenki, were reindeer breeders; neither Oroch nor Udihe were engaged in breeding.

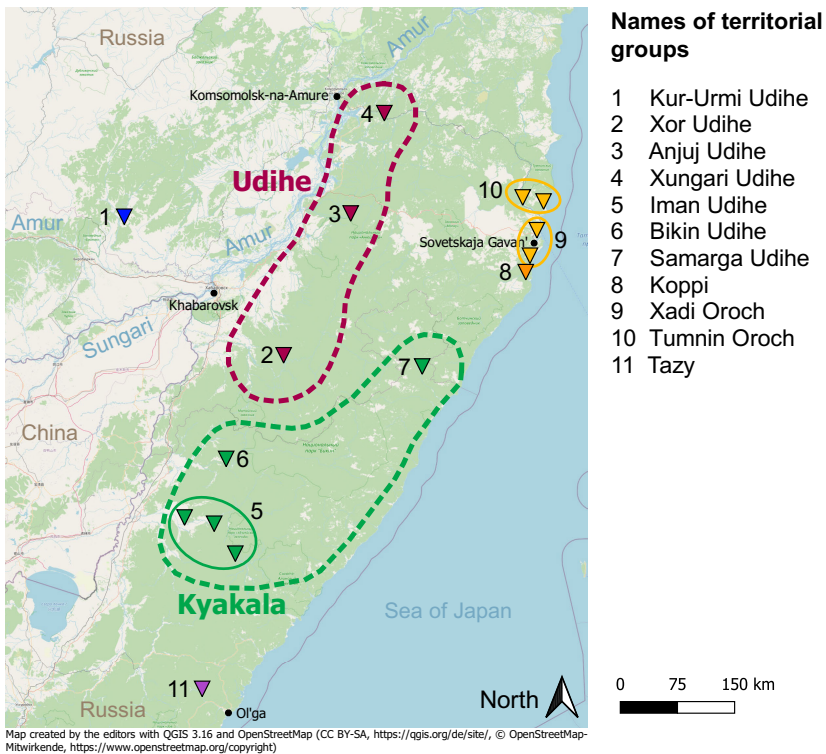


Figure 1: Udihe territorial groups

Previously, the Udihe and Oroch as well as other local ethnic groups had “clan identity”. “...the ethnonym Udihe (Udie) has been used since the 1930s. Before there was no general ethnic designation. Each areal group had its own self nomination: *hungakə* on Xungari, *bikiŋkə* on Bikin, *uniŋka* on Anjuj and so on” (Suliandziga et al. 2003: 142). The Oroch had no general ethnonym either. The Udihe call them *namuŋka* ‘sea shore dwellers’. This name was also used for the Udihe

living along small rivers that flow into the sea further to the south, so Oroch clans were not distinguished from the Udihe.

The first researchers did not separate the Udihe and the Oroch, which, apparently, reflected the real state of affairs. In the absence of a common self-designation, these people/peoples were called *Orochen* (*Orochon*) by the Russians. This name was given to the indigenous population living along the coast of the Tatar Strait and the Sea of Japan, by Jean-François de Lapérouse (Šrenk 1883: 142). This term is essentially erroneous, since it goes back to the Manchu-Tungusic word for reindeer, *oro(n)*. Neither the Udihe nor the Oroch were engaged in reindeer herding. Nevertheless, this ethnonym was used for some time.⁵ The *Orochen* were considered a separate ethnic group, along with the Gold (i.e., Nanai), Tungus (i.e., Ewenki and sometimes Even) or Gilyak (i.e., Nivkh).

In the modern scientific literature, the term *Udihe* appears for the first time in Sergej Brailovskij's work (Brailovskij 1901). He used the autonym of one of the groups of northern Udihe. Brailovskij also introduced the term *Tazy*⁶ as a synonym for Udihe. However, he did not separate the Udihe and the Oroch, and used the term *Oroch – Udihe*, and *Tazy* as synonymous. In the late 1920s, the campaign to change ethnonyms of Russian minorities was launched in the country. Old ethnonyms were assumed to be derogatory and were replaced usually by self-designations of respective peoples. Thus, Gold became Nanai, Gilyak became Nivkh, Tungus became Ewenki, Lamut became Even, etc. The *Orochen* were divided into three groups: Oroch, Udihe, and Tazy. This subdivision was apparently worked out by the famous geographer Vladimir Arseniev (Arsen'iev 1947-1949), who worked in the area.

This division is now universally recognized, and these ethnonyms are included in the list of Russian minorities. They were also recorded in Soviet passports as “nationality”. At present, when these languages are on the verge of extinction, and people themselves firmly know their “nationality”, this separation became reality. Still, the question arises how these idioms actually correlate.

⁵The term *Orochen*, referring to both the Udihe and the Oroch together, was used in all geographical, statistical, and other documents of the late 19th and early 20th centuries (see, for example, Šrenk 1883; Nadarov 1887; Margaritov 1888; Protodjakonov 1888; Przevalskij 1990 [1870]). It is worth mentioning that in Iman this designation is still used referring to the Udihe, being perceived as pejorative.

⁶The term *Tazy* goes back to Chinese 韃子 *dázi* ‘local resident of Primorye’; the word was already attested many hundreds of years ago in Chinese sources (Hölzl 2018: 116). *Tazy* is an ethnic group of Tungus-Manchu origin who have lost their native language and use a northern dialect of Chinese. *Tazy* were settled in the village of Mikhailovka, Olginskij district; about the *Tazy* language situation, see Belikov & Perekhval'skaya (1994).

The first dictionaries and other linguistic data on Udihe (Protodjakonov 1888; Leontovič 1898; Nadarov 1887; Margaritov 1888; Schmidt 1928), as well as a generalizing work of Brailovskij (Brailovskij 1901), did not separate Oroch and Udihe words.⁷ However, Brailovskij compared the data that he personally collected with words of other territorial groups, and came to the conclusion that the southern Udihe clans which had undergone Chinese influence were different from other groups. He combined northern Udihe (in modern terminology) and Oroch. At the same time, Brailovskij insisted on the cultural and linguistic unity of all “Oroch-Udihe”. The same was the point of view of Peter Schmidt (Schmidt 1928). The anthropologist Viktor Lar’kin also considered Oroch and Udihe two dialects of the same language, and divided Udihe into several sub-dialects (Lar’kin 1959: 5). Udihe and Oroch have been considered separate languages since the 1930s, beginning with works by Evgenij Šneider (Šneider 1936, 1937), Valentin Avrorin and Elena Lebedeva (Avrorin & Lebedeva 1978).

Regardless of whether Udihe and Oroch should be considered closely related languages or distant dialects of the same language, the fact remains that their territorial varieties form a dialect continuum. The Xadi (coastal) variety of Oroch is close to northern Udihe. The frontier dialect (Koppi variety) is described as either the most southern dialect of Oroch (Avrorin & Lebedeva 1978), or the most distant dialect of Udihe (Kormušin 1998). In fact, here the “official” border between Oroch and Udihe just coincides with the administrative border between the Khabarovskij and Primorskij Krai. Since the mouth of the Koppi river administratively is a part of the Khabarovskij Krai, local “Orochons” received the passport designation “Oroch” and are officially the Oroch. Until recently, the linguistic position of the Koppi variety remained unclear. In 2010, together with Natalia Kuznetsova, we conducted a study of the Koppi variety. Based on these data, I came to the conclusion that the Koppi variety combines features of Oroch and Udihe, being a transition from the northern dialects of Udihe to coastal varieties of Oroch. However, it shows more properties characteristic of Oroch. One of the main features being the preservation of etymological V-q-V and V-h-V complexes.

2.2 Udihe and Kekar (Kyakala)

Previously, the Udihe were also known as *Kekar* (*Kyakala*, *Kyakar* or *Kiyakara* in Manchu). “The Oroch call them *Ude* or *Kekar*, they call Oroch *Pæ*” (Emeljanov 1927). However, *Ude* and *Kekar* were not used as complete synonyms and referred not to one and the same but to two closely related ethnic groups. In 1998, Igor Kormušin wrote:

⁷There are newly found data on early Oroch (Alonso de la Fuente 2017).

Anthropologist Paul Schmidt in 1915 mentioned a remarkable fact, which did not attract due attention. Classifying the Manchu-Tungusic ethnic groups, he wrote that Oroch consist of three tribes: Oroch, Kyakar and Udihe. The term «Kyakar» [...] is preserved in Udihe in the form *kæ̃*⁸ (< *keka(r)*).⁹ As the legend says, there was also a legendary clan of the same name, which branched into several Udihe clans, localized mainly along the southern sea coast: *Amuliŋka*, *Geuŋka*, etc. If one takes into account that the *Udi*¹⁰ clan participated in the formation of Ulcha, and therefore should be localized much further to the north, then one should conclude that *Udi* and the *Kekar* correspond to the «Northern» and the «Southern» components of the Udihe ethnos respectively... (Kormušin 1998: 11–12, my translation – E.P.)

The anthropologist Anatolij Startsev suggested that initially there were three Udihe clans: *Udie*, *Kæ* and *Pianka* (Startsev 2004). According to Lar'kin (1959) the large *Kæ* clan divided into several clans: *Kančuga* (Kancuga), *Geonka*, *Kuinka* and *Suanka*. It is worth pointing out that in Xor (“Udihe proper”) there were only two clans: *Kjalundzjuga* (Kælundžuga) and *Kimonko* (Kimonko). The clan names *Kančuga*, *Geonka*, *Kuinka*, *Suanka* are usual among the Bikin Udihe and *K'æ* was registered in Iman.

It may be concluded that two distinct groups, Udihe and Kekar, were classified as one “nation” which is now called Udihe (or Udeghe, Udie). *Udihe* corresponds to the northern dialect cluster (Xor and Anjuj varieties); *Kekar* corresponds to the southern dialect cluster (Bikin, Iman and Samarga varieties). Very roughly, it can be said that northern dialects (“Udihe”) are closer to Oroch.

2.3 Udihe areal groups and dialects

2.3.1 Overview

Traditionally, the Udihe, being semi-nomads were spread across a fairly large territory: about a thousand kilometers from north to south. First researchers,

⁸Note that Kormušin used the apostrophe to mark laryngealized vowels after (not before) the character: *kæ̃*' (Shn. *k'eæ̃*, Sim. *Ki'a*).

⁹Janhunen has argued that it goes back to the word for ‘edge’, *kira > kija > kae (2012). However, *K'eæ̃* ‘clan *Kae*’ and *keæ̃* ‘edge’ are not homonyms. *K'eæ̃* contains the laryngealized /'eæ̃/ which points to the historical change VqV > V?V > V'V. It is mostly probable that the sequence *keka transformed into *k'eæ̃*.

¹⁰*udi* might be a word from the Manchu branch of Tungusic: Manchu *weji*, Alchuka *udi*, Bala *udi* ‘forest’. It seems there is no other Tungusic language that has a cognate of this word (Hölzl 2018: 121–122).

geographers, and anthropologists (see, for example, Arsen'iev 1947-1949: V, 81) indicated that dialectal differences in Udihe were so significant that the Udihe from different territorial groups hardly understood each other. However, modern studies showed that, with all the differences, the Udihe dialects are mutually intelligible (Simonov 1988; Perekhval'skaya 2010). Still, differences between Udihe dialects are not insignificant, and the mutual understanding between the dialects does not mean that they have identical systems (Trudgill 1985: 21–23).

Traditionally, Udihe dialects were named according to the river basins where they were spoken. Hunter-gatherer groups roamed within the basin of one river and acquired their specific language variety. The language of a larger areal group, however, was not uniform. Thus, Udihe clans living along the Bikin-river occupied specific smaller areas (along smaller rivers), and their language had specific features. There are still differences in the speech of those who came from the camps of Mitaxeza, Sjain, Olon, Sigou, Ulunga, Toholo, etc.

By the beginning of the 20th century, there were the following Udihe groups (Table 1): Kur-Urmi, Xor, Anjuj, Xungari (now Gur), Samarga, Bikin, Iman (now Bolshaja Ussurka).

Table 1: Udihe and Oroch dialects

Cover terms	Subgroups	Dialects
Udihe	Northern (Udihe)	Kur-Urmi
		Xor
	Southern (Kyakala)	Anjuj (Xungari)
		Iman
		Bikin Samarga
Oroch		Koppi Xadi Tumnin

In the 1930s, the Udihe were forcibly made sedentary: each areal group was settled in a specially built permanent settlement: Kukan (Kur-Urmi Udihe), Bira (Anjuj), Kun (Xungari), Agzu (Samarga), Gvasjugi (Xor), Sjain, Mitaxeza and Olon (Bikin), Sančixeza (Iman). The less numerous sea-shore Udihe were dispersed. In the 1960s and 70s in the course of the “consolidation of villages” campaign,

smaller Udihe villages were liquidated: Bira (Anjuj), Sančixeza (Iman), Sjain and Mitaxeza (Bikin). The Bikin Udihe resettled in the new Udihe village of Krasnyj Jar; and Anjuj and Iman Udihe were resettled into neighboring Russian villages. Therefore, a significant number of the Udihe were dispersed and came into daily contact with speakers of Russian. In the late 1930s the Kur-Urmi Udihe village of Kukan became a place of exile of political prisoners. After the building of the Khabarovsk-Sovgavan' railway, Kun, the village of the Xungari Udihe, became a railway station. Soon the Udihe were an insignificant part of the population in these villages.

At present, the largest Udihe settlements are: Agzu (Terneiskij district), where they constitute about 80% of the population of the village; Gvasjugi (Imeni Lazo district): 65% of the population, Krasnyj Jar and Olon (Požarskij district): 55% of the village population.

Each territorial group was characterized by a specific language variety. From a linguistic point of view, there are significant similarities between the Iman and Bikin dialects, on the one hand, and between the Xor and Anjuj dialects, on the other. They form the northern Udihe dialect cluster (Xor and Anjuj varieties), and the southern Udihe cluster (Bikin and Iman varieties). Samarga displays mixed features; however, it seems to be historically closer to the southern (Kekar) group. As for the Kur-Urmi dialect, it was heavily influenced by Ewenki. Orest Sunik expressed the idea of the proximity of Samarga and Xungari varieties (Sunik 1968: 231). According to Sunik, three dialect groups were distinguished in Udihe: Iman-Bikin, Xor-Anjuj and Samarga-Xungari. This statement cannot be verified because the Xungari variety has been completely lost and no data on it were published. From a purely a geographic point of view, the Xungari dialect should be placed in the northern cluster. Therefore, I will contrast the northern group (Anjuj, Xor) and the southern group (Bikin, Iman, Samarga) (2).

2.3.2 Dialect continuum

The linguistic reality is more complicated than the division of language into two dialect clusters. Territorial varieties of Udihe and Oroch form a “dialect continuum”. Neighboring varieties are linguistically rather close to each other, while the extreme points show significant differences. Moving from one variety to another, one can observe the gradual loss of certain linguistic features and the appearance of other features. This concerns all language levels: phonetics and phonology, morphology and syntax, vocabulary. In this article I will focus on the concrete realization of complex vowels in different varieties of Udihe.

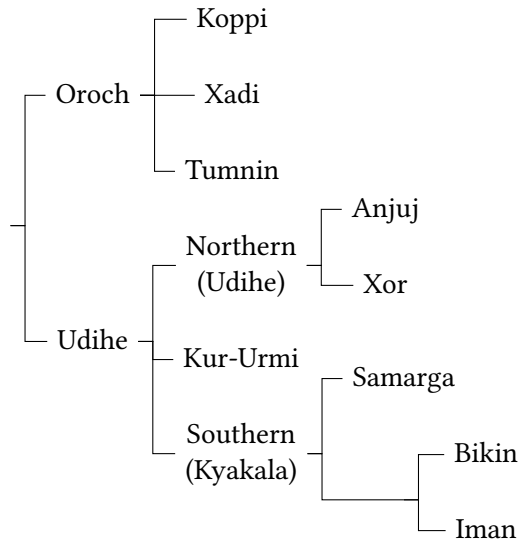


Figure 2: Classification of Udihe and Oroch varieties

3 Data

This work is based on the study of the following data:

1. Dictionaries

- Dictionaries of Oroch (Schmidt 1928; Avrorin & Lebedeva 1978);
- Dictionaries of Udihe (Nadarov 1887; Šneider 1936; Kormušin 1998; Simonov & Kialundziuga 1998; Girfanova 2001).
- Tungus-Manchu Comparative Dictionary. Materials for the etymological dictionary (Cincius 1975/77).

2. Descriptions

- Linguistic descriptions of Udihe varieties (Šneider 1936; Sunik 1968; Simonov 1988; Kormušin 1998; Nikolaeva 2000; Nikolaeva & Tolskaya 2001; Hölzl 2018).

3. Field data gathered in a number of field trips

- Tumnin and Xadi variety of Oroch, Khabarovskij Krai (2001, 2010); main speakers: Anatolij Namunka, Inna Akunka.

- Xor variety, Khabarovskij Krai (1989, 2011, 2013, 2014, 2018); main speakers: Valentina Kjalundzjuga, Nelli Andreeva (Kimonko), Polina Sun (Kjalundzjuga), Tatjana Dvojnova (Sulaindziuga).
- Koppi variety, Khabarovskij Krai (2006, 2010); main speaker Alexandr Ivaščenko (Pudza).
- Samarga variety, Primorskij Krai (1989, 2010, 2011); main speakers: Anatolij Kamandiga, Tatjana Kaza.
- Bikin variety, Primorskij Krai (more than 10 trips in the period 1984–2011); main speakers Aleksandr Kančuga, Nadežda Kukčenko, Aleksandr Pionka, Lydia Simanchuk, Susan Geonka, Anna Kančuga, Andrej Suanka, Evdikija Kjalundziga.
- Iman variety, Primorskij Krai (1989, 2010); main speakers: Ksenia Ajanka, Antonina Suanka, Aleksandr Kjalundziga.
- Field data were gathered with the help of Vladimir Belikov, Fatima Eloeva, Albina Girfanova, Irina Nikolaeva, Maria Tolskaya, Vlada Baranova, Kirill Maslinskij, Ksenia Viktorova, Valentin Vydrin, Natalia Kuznetsova.

4 Pharyngealized and laryngealized vowels in Udihe varieties

4.1 Udihe vowels

The peculiarity of Udihe inside the Manchu-Tungusic group is largely due to its phonetics and phonology, and especially the existence of several series of vowels. Laryngealized (glottalized) and pharyngealized (aspirated) vowels are features that clearly distinguish Udihe from Oroch and other Manchu-Tungusic languages (Zinder 1948: 58; Cincius 1949). However, the pharyngealized vowels in Udihe prove to be less stable than the laryngealized ones. Bikin and Iman varieties have lost pharyngealized vowels completely, while in Samarga they are kept only in some root morphemes. They are fully preserved only in the Xor dialect. On the contrary, laryngealized vowels are preserved in all varieties, although their specific realization may differ significantly (Šneider 1936; Simonov 1988; Nikolaeva 2000). In fact, the concrete realization of laryngealized vowels is one of the important features which distinguish Udihe dialects.

One of the most significant features that distinguish northern and southern varieties is the lack of pharyngealized long vowels in southern Udihe (1):

		‘fire’		‘button’
(1)	Xor	tō	~	to ^h o
	Bikin	tō	~	tō

Table 2: Vowel inventory of the Xor and Anjuj varieties (Šneider 1936: 83–86; Simonov 1988)

Description	Vowels								
Short	a	o	u	ə	i	e	æ	ø	(y)
Long	ā	ō	ū	ē	ī	iə	eæ	yø	yī
Long laryngealized	‘a	‘o		(‘ə) ^a					
Long pharyngealized	a ^h a	o ^h o	u ^h u	ə ^h ə	i ^h i	(i ^h e)	e ^h æ	y ^h ø	y ^h i

^aLaryngealized /‘ə/ is postulated by Nikolaeva and Tolskaya on the basis of one verb form: the perfect stem for verbs of the type *ətətə*- ‘to work’ – *ətət’ə* ‘he has worked’ (Nikolaeva & Tolskaya 2001: 40). However, this phoneme has a very narrow scope: it does not occur in any other position.

Table 3: Vowel inventory of the Bikin, Iman and Samarga varieties

Description	Vowels								
Short	a	o	u	ə	i	e	æ	ø	(y)
Long	ā	ō	ū	ē	ī	ē	æ	ō	
Long laryngealized	‘a	‘o		(‘ə)					

4.2 Interpretation of the Udihe complex vowels

The interpretation of the Udihe vocalic complexes, laryngealized and pharyngealized, has long been controversial. Trubetskoi’s phonology counts several dozen vowel phonemes in Udihe, as it was presented by Šneider (1936); in some other works the phonemic status of these complex vowels is not clarified (Nikolaeva 2000).

I regard suprasegmentals as an independent tier (Goldsmith 1976). Therefore, I postulate the appearance of the suprasegmental tier as a compensation for the simplification of the segmental tier. In Xor, Bikin and Iman varieties it was the

representation of the suprasegmental tier that underwent changes while the segment tier remained unchanged. The concrete realizations of pharyngealized and laryngealized long vowels differ according to the variety and also according to the mode of speech.

4.3 Pharyngealized long vowels

4.3.1 Etymology

Udihe pharyngealized vowels go back to a combination of three phonemes, V-s-V. In root morphemes:¹¹

- (2) Udihe *a^hanta* ‘woman’ ~ Oroch *asa*, Pl. *asa-nta*; Negidal *asīxān* ‘girl’; Ewenki *asī* ‘female’;
- (3) Udihe *a^hikta* ‘fir-tree’ ~ Oroch *asikta*, Ewenki *asikta*, *acēkān*, Nanai *xasikta* ‘fir-tree’;
- (4) Udihe *i^hi* ‘larch tree’ ~ Oroch *isi*, Negidal *isi*, Manchu *isi*, Ewenki *ismama* ‘dry larch tree’;
- (5) Udihe *o^hə*, *wo^hə* ‘deer-leg fur’ ~ Oroch *ōso*, Negidal *ōsa*, Ewenki *ōsa* ‘deer-leg fur’;
- (6) Udihe *u^həndə*, *wu^həndə* ‘to throw’ ~ Ewenki *usəndə* ‘to throw (weapon)’, Negidal *usə* ‘spear’.

In suffixes:

- (7) Udihe *bəli^hə* ‘girl (in tale)’ < **bəli-sə* ~ Negidal *bəlin*, Ewenki *bəlin* ‘hysteria’;
- (8) Udihe *tumi^hə* ‘path in mountains’ < **tumi-sə* ~ Negidal *tuasan* ‘footprint’, Ewenki *tū* ‘to step on’.
- (9) Udihe *ɲama^hi* ‘warm’ < *ɲama-si* ~ Oroch *ɲama*, *ɲama-si*, Nanai *ɲama*, Ewenki *ɲama* ‘warm’.

It is worth noting that the transition V-s-V → V^hV in root morphemes took place mostly when this complex was at the beginning of the word, in other words, when the root began with a vowel (see examples above). Later, a prosthetic glide

¹¹Hereinafter, the correspondences are given according to Cincius (1975/77).

could appear $o^h\theta > wo^h\theta$ 'deer-leg fur'; $i^hi > ji^hi$ 'larch tree', which already happened early in Udihe.

When a consonant was at the beginning of the word, this transition often did not take place, cf.:

- (10) Udihe *gaja* 'waterfowl, duck' ~ Oroch *gasa* 'waterfowl', Negidal *gasa* 'swan', Ewenki *gasa* 'crane';
- (11) Udihe *kəjə* 'word, language' ~ Oroch *xəsə* 'word', Negidal *xəsə* 'word, language', Nanai *xəsə* 'word, order', Manchu *xəsə* 'order, command'.¹²

Some words which have a consonant before the pharyngealized vowel do not have a convincing Tungus etymology:

- (12) $gə^hə$ 'bad', $dü^hi$ 'brain', $ʒa^hi$ 'wild boar', $tə^hu$ 'all'.

Or go back to different complexes:

- (13) Udihe to^ho 'button' ~ Negidal *toxon*, Nanai *toχō* 'button', Manchu *toχon* 'metallic button'.

Intervocalic *-s-* in Udihe goes back to *-č-*: *asa-* 'to fit' ~ Oroch *ača-*, Nanai *ača-* 'to come to'; Ewenki *arča-* 'to meet'.

4.3.2 Realization

Pharyngealization in different Udihe varieties can be realized as: a) a break of the sound by aspiration; b) breathy voice phonation, c) a "clean" long vowel. The concrete realizations of pharyngealized long vowels differ according to the variety and also according to the mode of speech.

Table 4 shows that each Udihe variety is characterized by two different modes of pronunciation: the full mode (FM) which is shown in the cell to the left and the allegro mode (AM) in the right cell.

Taking into account different tempo modes in each variety, Table 4 shows that the allegro mode of pronunciation of one variety corresponds to the full mode of pronunciation of the neighboring one, which produces a new allegro mode. It demonstrates the internal "anatomy" of the dialect continuum.

From a phonological point of view, the VhV sequence with a weakened consonant in the intervocal position is of particular interest. Acoustically it is a long vowel interrupted by aspiration. Its phonemic interpretation, however, can be twofold, depending on the variety analyzed.

¹²Šneider gives the forms $ga^hə$ 'duck' and $kə^hiə$ 'word', which are not confirmed by modern material. I did not find such forms in any of the varieties.

Table 4: Types of realizations of pharyngealized long vowels in different Udihe varieties and Oroch. Comments: V-s-V and V-h-V: sequences of three segments; V^hV: long vowels interrupted by aspiration; \bar{V} : long vowel with pharyngealized phonation (“breathy voice”); \bar{V} : long vowel.

	Oroch			Udihe		
Varieties	Xadi, Tumnin		Koppi (transitional)		Xor, Anjuj (Udihe)	
				Bikin, Iman (Kyakala)		
Stages	V-s-V	V-h-V	V ^h V	\bar{V}	\bar{V}	

In Koppi, this is an optional pronunciation variant characteristic of the allegro mode; the full mode of pronunciation is V-h-V (sequence of three phonemes). In the speech of Alexandr Ivashchenko, a Koppi speaker, sequences of this type were pronounced as three syllables in the full mode of pronunciation. In order to clarify a word Ivashchenko could chant it, clearly dividing these sequences into three syllables *abdæha*¹³ ‘leaf (of a tree)’ [ab.dæ.ha]. However, in the allegro mode, the V-h-V sequence contracted into a long vowel, interrupted by a brief aspiration [ab.dæ^ha]. See the following pronunciation of the word /abdæha/ ‘leaf (tree)’ in allegro (left) and full (right) modes of pronunciation.

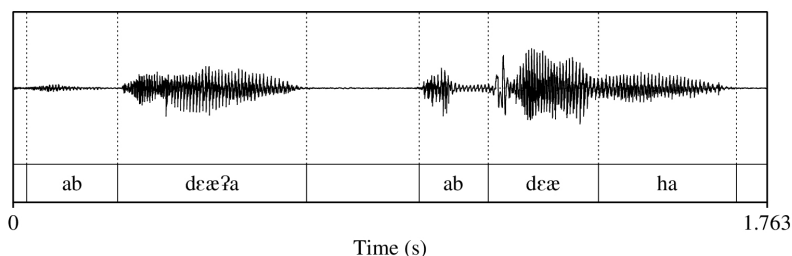


Figure 3: Koppi dialect, speaker Akexandr Ivashchenko: [abdæ^ha], [abdæha] ‘leaf’

Similar observations were made by Igor Kormušin:

In the fully marked type of pronunciation, if the vowels surrounding the pharyngeal consonant are similar, they are pronounced with equal length

¹³Hereinafter, aside from specific phonetic realizations, Udihe words are given in Šneider’s writing system.

and, in fact, form two syllables with **h** being voiced: *ahanta* (a-ḥan-ta) 'woman', *gehe* (ge-ḥe) 'bad', *oloho* (o-lo-ḥo) 'boiled fish', *ihi* (i-ḥi) 'larch'. In the fully normal type of pronunciation, **h** is articulated simultaneously with the second vowel, becoming a pharyngeal overtone in its initial part; at the same time, the pharyngeal consonant is fused with the previous vowel, so that a single complex sound is formed; as a result, the syllable border is aligned differently, combining two syllables into one: a^hanta, ge^he, olo^ho, i^hi. [...] in the normally abbreviated type of pronunciation, the surrounding vowels fuse into a long one, the pharyngeal consonant following it [...] a^hnta, ge^h, olo^h, i^h. This pronunciation creates conditions for the deletion of **h** [...] (Kormušin 1998: 64–65, my translation – E.P.).

Kormušin distinguished three pronunciation modes: fully marked, normally full and normally abbreviated. They correspond to chant, full style and allegro mode.

According to my data, none of the varieties exhibit coexistence of all the types of pronunciation that Kormušin singled out. Most likely, the researcher combined phenomena observed in different varieties.

Evgenij Šneider who worked in the 1930s with Anjuj Udihe interpreted the sequence VhV (full style in Anjuj) not as a sequence of two syllables, but as a long vowel interrupted by aspiration.

Of course, *h* in this sound complex is not an independent consonant [...] When comparing Udihe words with pharyngealized vowels with words of the same meaning in other Manchu-Tungusic languages, it turns out that [...] the two-syllable combination became monosyllabic, i.e., the transformation process *s* (ṣ) > *h* (ḥ) was accompanied by the contraction of the pair of identical vowels. This resulted in the emergence of a new category of vowels, for example *aha-* 'to catch up' (Ewenki *asa-*); *imaha* 'snow' (Oroch *imasa*); *ihi* 'larch' (Oroch, Manchu *iṣi*) [...] (Šneider 1937: 10–11, my translation – E.P.).

4.3.3 Realization of pharyngealized vowels

4.3.3.1 Koppi

The two types of pronunciation, V-h-V and VhV, seem to be characteristic of the northernmost dialects of Udihe: Koppi, and, apparently, Xungari. Most likely, at the end of the 19th century pronunciation of pharyngealized complexes as three segments V-h-V was also characteristic for Xor Udihe. In Nadarov's work we

find *яга* (*jaga*) ‘eye’ (Shn. *jehæ*, Sim. *jâ*), *нюге* (*niuge*) ‘nose’ (Shn. *nyhø*, Sim. *ɲiê*), *того* (*togo*) ‘button’ (Shn. *toho*, Sim. *tô*). It is not clear what kind of sound was represented by the Cyrillic letter «Г»; most likely it was a pharyngeal consonant, possibly voiced. In some cases, Nadarov did not note it, cf. another variant of the word ‘eye’ *я* (*ja*), *ниама* (*niama*) ‘leather jacket’ (Shn. *nehæma*, Sim. *ñâma* ‘leather’).¹⁴

4.3.3.2 Xor

In Xor Udihe a “new category of vowel” was formed. In the 1930s, the Udihe on the Xor River apparently pronounced VhV in the full pronunciation mode, and V^hV in the allegro mode. The full pronunciation mode of Xor Udihe was the basis of “literary” Udihe, in which several textbooks for primary school were published. Simonov, who worked with Xor Udihe since the late 1970s, noted at that period the V^hV variant was the full mode, and pharyngealized vowels were pronounced \bar{V} in allegro mode:

Pharyngealized vowels are pronounced with a sharp increase in intensity towards the end of the phonation. [...] When the aspiration is present, it is not in the middle of the vowel, but is superimposed on its entire second half. (Simonov 1988: 52, my translation – E.P.)

Simonov presented to the speakers words with a pharyngealized vowel, pronounced in two syllables: “words **je.hæ* (instead of *jâ* ‘eye’); **a.han.ta* (instead of *ânta* ‘woman’); **imo.ho* (instead of *imô* ‘fat’) were simply not understood by speakers” (Simonov 1988: 52, my translation – E.P.).¹⁵

In 2006, only one type of pronunciation of pharyngealized vowels was observed in the Xor variety. With the most complete pronunciation mode, a separate word could be pronounced as V^hV. However, even in this case, aspiration appears also after the vowel, cf. Figure 4.

Figure 4 shows that the final part loses vocalic characteristics turning into an aspiration. The final complex consists of a long vowel (250 milliseconds), the duration of which is almost twice of the initial short vowel (*u*). Compare pronouncing by the same speaker of the Accusative case form of the same word: *umahawa* (Figure 5).

¹⁴Simonov suggested that Nadarov recorded pharyngeal (*h*) only at the rhythmic boundaries of the word, but this does not explain the presence of doublets in Nadarov’s list of words: ‘eye’ *я* and *яга*; ‘nose’ *нюгу* and *нионё* (Nadarov 1887).

¹⁵This consideration was the reason for changing the type of writing for Xor Udihe made by Simonov; he introduced circumflex “ \hat{v} ” to mark breathy voice phonation (aspiration): *imô* ~ Shn. *imoho*.

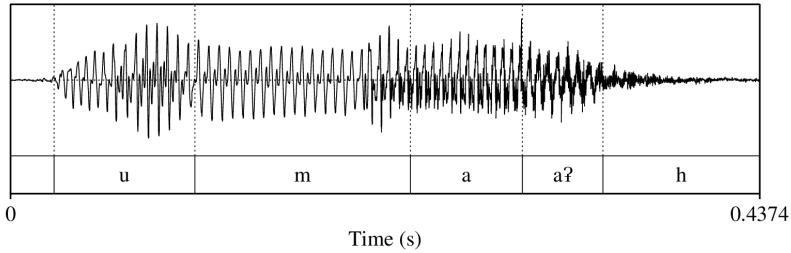


Figure 4: Xor variety. Speaker Valentina Kjalundzjuga: *umaha* [umaʔaḥ] ‘bone marrow’

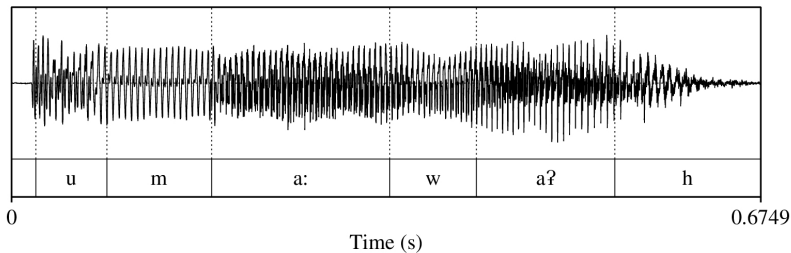


Figure 5: Xor variety. Speaker Valentina Kjalundzjuga: *umahawa* [umaawaḥ] ‘bone marrow ACC’

Figure 5 shows that the long vowel /ā/ in *umaha* has lost its pharyngealized quality, but the aspiration appears at the absolute end of the word. Such a transfer of aspiration to the end of the phonetic word may be an individual characteristic of the speaker, but most likely it reflects the pronunciation of pharyngealized vowels in the Xor variety. Kormušin also pointed out such a realization of pharyngealized vowels.¹⁶

Compare the realization of these two words in Figure 6. The principles of Autosegmental Phonology (Goldsmith 1976) explain this by the independent character of the suprasegmental level. Phonation characterizes the whole word and not any particular segment and is realized at the end of the word.

As pointed out by Kormušin, such a pronunciation creates the conditions for a loss of aspiration. This happened primarily with pharyngealized vowels in the final position, as in the examples above. Apparently, the loss of pharyngealization

¹⁶Nikolaeva questioned the possibility of such a realization: if pharyngealization is realized at the end of the word, it contradicts the basic phonotactic rules of Udihe. However, it is not a consonant, but a pharyngealized vowel (Nikolaeva 2000). This is an additional consideration in favor of the interpretation of these complexes as single phonemes.

occurred primarily in non-root morphemes. The pharyngealization, therefore, was lost in the personal markers 2SG, PL: $-ị > -i$; $-ụ > -u$; and in past tense suffixes: $-ạ- / -ọ- / -ə̣- > -a- / -o- / -ə-$.

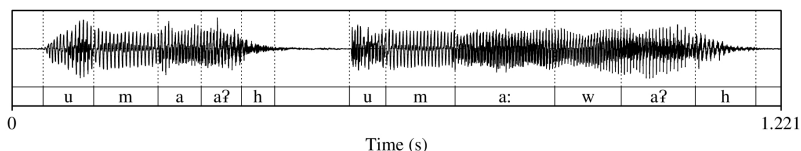


Figure 6: Xor variety. Speaker Valentina Kjalundzjuga: *umaha* [umaʰa] ‘bone marrow’; *umahawa* [umaawah] ‘bone marrow (ACC)’

Table 5: Personal possessive forms in Udihe varieties and in Oroch. The Oroch data are taken from Avrorin & Lebedeva (1968) and Avrorin & Boldyrev (2001). In both sources, alternative forms are given without comments. I suggest that forms which are closer to Udihe are characteristic of the Xadi dialect which is more innovative.

Form	Oroch	Xor Udihe FM	Xor Udihe AM, Bikin Udihe	Translation
1SG	təgə-wi, təgə-ji, təgə-i	təgə-i	təgə-i	‘my shirt’
2SG	təgə-si	təgə-ḷi	təgə-i	‘your shirt’
1PL.EXCL	təgə-mu, təgə-wu	təgə-u	təgə-u	‘our shirt’
2PL	təgə-su, təgə-hu	təgə-ḷu	təgə-u	‘your shirt’

In the Xor variety, this loss is characteristic of the allegro mode of pronunciation. Auxiliary and negative verbs, being the most frequent ones, were also “erased” as the full mode of their pronunciation was replaced by the allegro mode. Besides, in the speech of younger Xor variety speakers, pharyngealized vowels are totally lost and have been replaced with pure long vowels in all positions.

Published Xor Udihe texts demonstrate incoherence in marking the pharyngealized vowels, which reflects differences in modes of pronunciation. An example are folklore texts recorded mainly with the Xor Udihe (Simonov et al. 1998). In the vast majority of past tense forms, pharyngealization is not marked in suffixes, cf. in text No. 1: *oloktoni* (< **olokto-ho-ni*) ‘s/he cooked’; *andalati* (< **andala-ha-ti*)

‘they made friends’; *alasiēni* (< **alasi-hə-ni*) ‘s/he waited’ (Simonov et al. 1998: 74). Similarly, pharyngealized vowels in personal suffixes are also not marked in these texts. And in root morphemes, pharyngealized vowels are sequentially marked.

4.3.3.3 Samarga

In the Samarga variety, pharyngealization is kept only in some root morphemes.

4.3.3.4 Bikin and Iman

Pharyngealized vowels are completely lost in Bikin and Iman, where the corresponding complexes are pronounced as clear long vowels, cf. the word for ‘leaf’: Oroch [abdasa]; Koppī variety [abdəha] (FM) ~ [abdə^ha] (AM); Xor Udihe [abdæ^he] (FM) ~ [abdæ^hɛ] (AM); Bikin Udihe [abdæ^hæ]. While in Xor the loss of pharyngealized vowels is a recent phenomenon, and elder speakers still pronounce them at least under the full mode of pronunciation, in the Bikin and Iman varieties, pharyngealization was not characteristic for the speech of people born in the 1920–1930s. This means that pharyngealized vowels were lost at least a hundred years ago. In the Bikin variety, the etymological pharyngealized vowels were replaced by long ones, and there is a tendency for these vowels to become short (Nikolaeva 2000: 115–116; Tsumagari 2012).

In sum, the data presented show that Udihe varieties present different stages of one process: weakening of the consonant in the intervocalic position with the substitution of segment units by suprasegmental ones.

4.3.4 Loss of pharyngealization and its effects in morphology

Loss of pharyngealization had a significant impact on the morphological system of the southern dialects. The main consequence of the loss of pharyngealized vowels here was the formal coincidence of possessive suffixes of the first and second person singular and plural (Exclusive form) for vowel-final stems; cf. data in Table 6.

In southern Udihe, in order to clarify the “possessor”, personal pronouns are used. While in northern Udihe the use of personal pronouns indicates emphasis, in southern Udihe it is neutral. Therefore, southern Udihe displays a greater degree of analyzability.

Table 6: Fragment of the paradigm of the personal possessive conjugation of the noun *kusigə* ‘knife’ in Bikin and in Xor Udihe. Forms merged in Bikin Udihe are bold.

	Nominative		Accusative	
	Xor	Bikin	Xor	Bikin
1SG	kusigə-i	(bi) kusigə-i	kusigə-wə-i	kusigə-wə-i
2SG	kusigə-ḷ	(si) kusigə-i	kusigə-wə-ḷ	kusigə-wə-i
3SG	kusigə-ni	kusigə-ni	kusigə-wə-ni	kusigə-wə-ni
1PL.EXCL	kusigə-u	(bu) kusigə-u	kusigə-wə-u	kusigə-wə-u
1PL.INCL	kusigə-fi	kusigə-fi	kusigə-wə-fi	kusigə-wə-fi
2PL	kusigə-ḷ	(su) kusigə-u	kusigə-wə-ḷ	kusigə-wə-u
3PL	kusigə-ti	kusigə-ti	kusigə-wə-ti	kusigə-wə-ti

4.4 Laryngealized long vowels

4.4.1 Etymology

Udihe laryngealized vowels go back to the V-q-V complex, which was a three-phoneme combination and is present in many Tungusic languages, cf.:

- (14) Udihe *um'a* ‘fishing hook’ ~ Oroch *uməkə(n)*; Negidal *umaxa/umaka*; Ewenki *umika*;
- (15) Udihe *l'o-* ‘to hang’ ~ Oroch *lō/loko* [loqo]; Negidal *loko-*; Solon *loko-*; Kilen *loqo-*; Ewenki *loko-*;
- (16) Udihe *x'ai*¹⁷ ‘liver’ ~ Oroch *xākin* [xāqin]; Negidal *xāxin*; Kilen *xakī*; Ewenki *hakin*;
- (17) Bik. Udihe *g'ata* ‘cranberry’ ~ Oroch *gākta/gakka* [gaq:a]; Negidal *gaxakta*; Ewenki *gakakta*;
- (18) Udihe *g'ai* ‘crow’ ~ Oroch *gāki* [gāqi]; Negidal *gāxi*; Ewenki *gāki*;
- (19) Udihe *kakt'a* ‘half’ ~ Negidal *kaltaka/kaltaxa*; Ewenki *kaltaka*; Solon *xaltaxa*;
- (20) Udihe *s'ou* ‘scoop’ ~ Ewenki *sokowun/hokowun*; cf. Oroch *soko-* [soqo], Negidal *soxo-* ‘to scoop’.

¹⁷In sequences *a-q-i the first vowel holds the phonation: [a^ʔai] or [ʔai], in practical writing: 'ai.

It should be noted that only the uvular variant [q] of the phoneme /k/ transformed into the glottal stop and further created the creaky voice phonation. The velar [k] was preserved in Udihe as [x] and [k]:

- (21) Udihe *joxo* ‘kettle’ ~ Oroch. *iko*; Negidal *ix̄ā*; Ewenki *ik̄ā*.
(22) Udihe *jəxə* ‘to sing’ ~ Oroch *ikə*; Negidal *ix̄ā*; Ewenki *ik̄ā*.
(23) Udihe *cikə*- ‘to urinate’ ~ Negidal *cix̄ən*; Kilen *cik̄ə*; Ewenki *cik̄ən*-.

In Oroch, in accordance with an assimilation rule, the uvular allophone [q] occurs only after the vowels [a] and [o]. In other cases, the velar [k] appears. Apparently, a similar rule was also present in Udihe. The uvular [q] then transformed into the glottal stop. This explains why the series of laryngealized vowels in Udihe is limited to ‘o and ‘a.

Evgeny Šneider, on the basis of general system considerations, postulated the presence of the entire set of laryngealized long vowels, both simple and diphthongoids (Šneider 1936: 83). As Simonov showed, this does not correspond to the linguistic reality (Simonov 1988).

It is worth noting that “non-etymological” laryngealized vowels sporadically appear after the plosives *b*, *p*, *c*, if followed by the vowel *a*, cf.:

- (24) Udihe *b’ata* ‘boy’ ~ Oroch *bātu*, *baturi* ‘strong warrior’, Nanai *ātor* ‘strong warrior’; Manchu *baturu* ‘hero, warrior’;
(25) Udihe *c’aligi* ‘white’ ~ Oroch *čām* ‘white’; Nanai *čāyžā* ‘white’.¹⁸

It is also noticeable that many Udihe words with laryngealized vowels do not have a reliable Tungusic etymology. Often they are attested only in Udihe: ‘*ana* ‘boat’, ‘*d’a* ‘cotton wool’; ‘*gob’o* ‘fly’, ‘*asa* ‘bay’; ‘*t’añki* ‘middle’, ‘*s’ai* ‘salt’ and others. Still, these words are known in all Udihe dialects. The etymology of some other words is not very convincing, e.g. *od’o* ‘grandfather’ is compared with Oroch *adiyi*; Ulcha *ədəkə(n)* ‘father-in-law’ which is doubtful.

It may be supposed that Udihe had undergone influence of a substrate or adstrate non-Tungusic language which was also the source of non-Tungusic loanwords.

¹⁸However, consider Ewenki *bagadi* ‘strong, brave’, proto-Mongolian **bayatur* and proto-Mongolian **čayān* ‘white’. Based on a comparison of Udihe forms with historically attested Mongolian ones (as given by Cincius 1975/77), it may be imagined that Udihe retained a more ancient form. However, Udihe laryngealized vowels originate from -q- and not -g/y-.

4.4.2 Types of realization

Laryngealization in different Udihe varieties can be realized as: a) a break of the sound by the glottal stop; b) creaky voice phonation, c) the increased intensity in combination with the low/falling tone. The flattening effect of laryngealization is observed in all Udihe varieties. However, only in the most innovative varieties of Bikin and Iman, it became the main (and in Iman Udihe the only) distinctive feature [+tone]. Thus, these varieties should be classified as tonal, which is untypical for Manchu-Tungusic languages.

The concrete realizations of laryngealized long vowels differ according to the variety and also according to the mode of speech.

Table 7: Types of realizations of laryngealized long vowels in different Udihe varieties and Oroch. Comments: V-q-V and V-ʔ-V: sequences of three segments; V'V: long vowels interrupted by a glottal stop; \bar{V} : long vowel with laryngealized phonation (“creaky voice”); \dot{V} : long vowel with falling (low) tone.

	Oroch			Udihe	
Varieties	Xadi, Tumnin	Koppi (transitional)	Xor, Anjuj (Udihe)	Bikin, Iman (Kyakala)	
Stages	V-q-V	V-ʔ-V	V'V	\bar{V}	\dot{V}

Table 7 shows the two modes of pronunciation: the full mode (FM) which is shown in the cell to the left and the allegro mode (AM) in the right cell.

Udihe varieties and the closely related Oroch language represent changes of certain phonetic complexes “from consonant to tone”; each variety representing a certain stage of this process. The innovation was spreading, roughly, in the direction from north to south: Oroch → Koppi variety → Xor and Anjuj varieties → Bikin and Iman varieties.

Cf. the word for ‘dog’: Oroch [inaqi]; Koppi variety FM [unæʔi], AM [inə²i]; Xor variety FM [inæ²ai], AM [inǝɛi]; Bikin variety FM [inǝɛi], AM [inǝ²ei].

4.4.3 Realization of pharyngealized vowels

To study the realizations of laryngealized vowels in different Udihe varieties is a difficult task when based on written sources. In the case of pharyngealized vowels written sources provide more or less reliable information, but laryngealized

vowels are written with an apostrophe uniformly by all researchers; this spelling hides rather different types of realization.

4.4.3.1 Oroch and Koppi

According to Avrorin and Lebedeva, in Oroch the phoneme /k/ is realized as uvular [q] in the position after /a/, /ä/, /o/, between identical vowels, or before /i/ (Avrorin & Lebedeva 1978). This is also characteristic of the Koppi variety.

It should be noted that in slower speech, an aspiration [h] is clearly heard between the vowel and the uvular [q]: [na^hqi] 'dog', [beæ^hqa] 'river', [ara^hqi] 'strong spirit', [ga^hqi] 'crow'. A variant realization is a pause before [q], which is perceived as a "long stop": [maa[?]qi]. Here ([?]q) represents a preglottalized consonant. See Figures 7 and 8 on the pronunciation of the words [beæ^hqa] and [maa[?]qi] in Koppi.

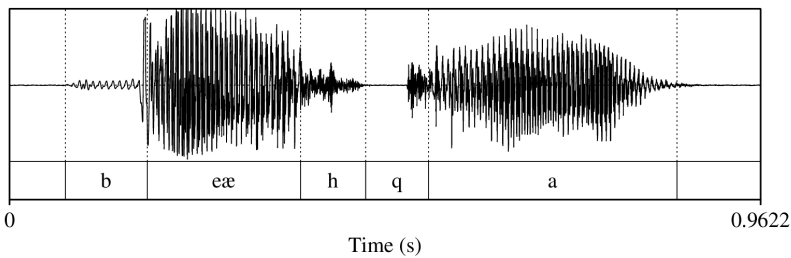


Figure 7: Koppi variety. Speaker Alexandr Ivaščenko: [beæ^hqa] 'river'

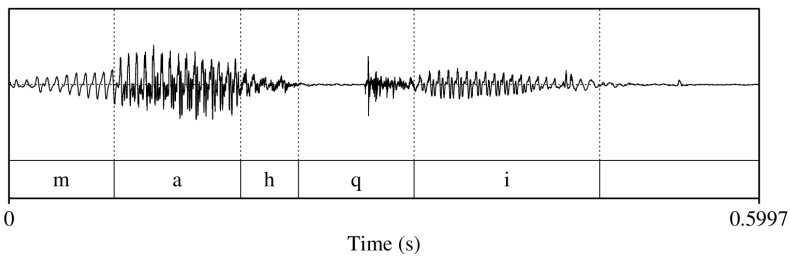


Figure 8: Koppi variety. Speaker Alexandr Ivaščenko: [maa[?]qi] 'there is no'

The spectrogram of [beæ^hqa] 'river' shows aspiration after a long diphthongoid [eæ], then there is a gap followed by the stop [q]. The spectrogram of [maa[?]qi] 'there is no', seems to present no aspiration, but the silence zone before the stop lasts for more than 70 milliseconds.

In other phonetic contexts the phoneme /k/ is realized as a velar [k]: [ukəhə] ‘door’: Bikin *ukā* ‘door, doorway’. The Koppi speaker never pronounced [q] as [ʔ]. Apparently, this pronunciation is not typical for the Oroch language, including the Koppi dialect. Obviously, we are dealing with a sequence of independent phonemes, and not with a complex sound in this variety.

4.4.3.2 Xor

Šneider described laryngealized vowels of Xor Udihe as having a stop interrupting the vowel; Lev R. Zinder and Margarita I. Matusevich¹⁹ showed that this stop occurs closer to the beginning of the vowel (Zinder 1948). At present, these vowels are pronounced in allegro mode with “creaky voice” phonation. See two variants of *in’ai* ‘dog’ pronounced by the same speaker within the same recording session. At first, the speaker clarified the word (full mode); later, she pronounced it more “carelessly” (allegro mode). It is worth noting that the speech tempo remained almost the same; it was the intensity of pronunciation and the tonal pattern that changed.

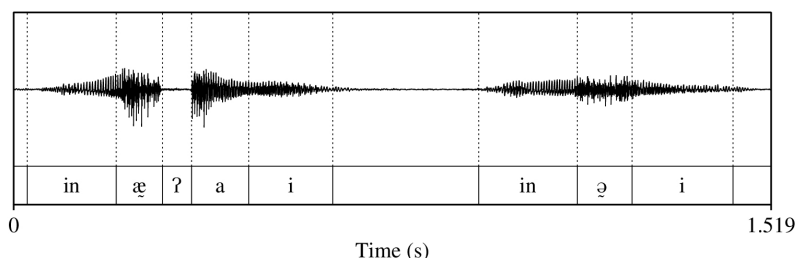


Figure 9: Xor variety. Speaker Valentina Kjalundzjuga: *in'ai* ‘dog’: FM [inæ^ʔai], AM [inəɕi]

In Figure 9, the creaky phonation zone can be seen in AM pronunciation. It should be noted that in FM the part of the vowel before the stop is also pronounced with creaky phonation.

These observations confirm the conclusions made by Zinder and Matusevich. Indeed, there is a pronunciation variant when the vowel is broken by a stop. Still, at present the most common way to pronounce a “laryngealized vowel” in

¹⁹The results of the study of the Udihe phonetic system, carried out in the 1930s by Lev Zinder and Margarita Matusevich in the laboratory of experimental phonetics of Leningrad State University, were not published. Partially they were included in Zinder (1948) and Kormušin (1998).

Xor Udihe (FM) is when the vowel is not broken with a stop, but with a glottal approximant, cf. the utterance by V.T. Kjalundzjuga of the word *bul'a* 'ash', see Figure 10.

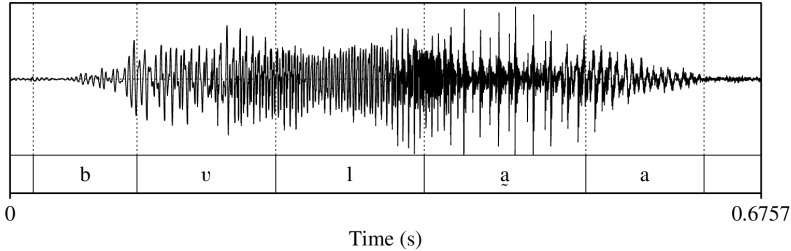


Figure 10: Xor variety. Speaker Valentina Kjalundzjuga: *bul'a* 'ash tree' [bulaʔa]

Thus, in Xor Udihe, three types of pronunciation of laryngealized vowels co-exist: 1) the vowel is interrupted by a stop; 2) by an approximant; 3) the vowel bears creaky voice phonation. However; the creaky phonation does not characterise the whole vowel, and happens in the place where the stop would have been pronounced under another mode of pronunciation.

4.4.3.3 Bikin, Iman

In varieties of the southern dialect cluster laryngealized vowels with glottal stop are not found. Specific realizations of laryngealized vowels are in fact a diagnostic feature that differentiate local varieties of Bikin Udihe. The Udihe came from different smaller camps before settling down in the village of Krasnyj Jar, and they still retain certain speech differences. Unfortunately, at present, it is difficult to make a detailed study of these varieties due to the poor preservation of the language and the small number of speakers. However, our language consultants distinguish people speaking *Sjain*, *Olon*, *Ulunga*, *Mitahiza*, *Sigou* and other varieties. Basing on the data that I have, it may be concluded that Udihe varieties located upstream the Bikin-river were more conservative in vocalism, and the most innovative one being *Olon*, the village lowest downstream.

The realization of laryngealized vowels in Bikin Udihe varies significantly. It may be a) a laryngeal spirant; b) creaky voice phonation; c) a sharp increase of intensity of the vowel in combination with a low tone. The latter [c] is typical primarily for people from *Olon*. Consider realizations of laryngealized vowels: laryngeal spirant and creaky voice phonation with a speaker of the *Sigou* variety.

Figure 11 shows that the laryngealized vowel is in the beginning of the word and is realized as a laryngeal spirant, clearly visible on the spectrogram. The

vowel [ā] is long and carries the creaky phonation. The next vowel [ɐ] is more close. The example presented concerns the full pronunciation mode in Bikin variety; in the allegro mode, laryngealized vowels are never pronounced as a laryngeal spirant.

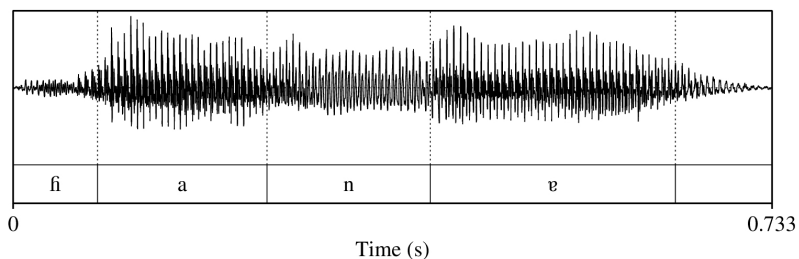


Figure 11: Bikin variety. Speaker Lydia Simanchuk: 'ana' 'boat' [fianɐ]

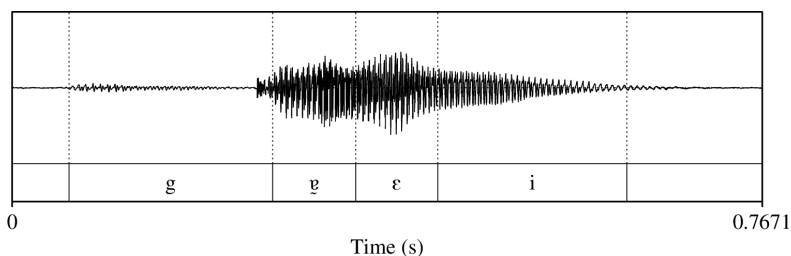


Figure 12: Bikin variety. Speaker Lydia Simanchuk: *g'ai* 'crow' [gɕɛi]

In Figure 12, the laryngealized vowel occurs in a CV syllable. It has an articulation that is more close and carries the creaky phonation.

It is interesting that in some Bikin variants the laryngealized element, as an independent suprasegmental unit, can change its position in the word. It can be located at the beginning of the vowel (see Figures 11, 12), and it can also move to the end of the vowel (see Figure 13). It may be an individual characteristic of the speaker, or, perhaps, characteristic of a local variety. Figure 12 presents the example of phonation realized at the second part of the vowel.

Figure 13 shows a decrease in pitch on the laryngealized vowel. This peculiarity of pronouncing laryngealized vowels in the Xor variety was noted by Galina Radchenko, who conducted an experimental study of the phonetics of this variety (Radčenko 1988: 37). Radchenko explained this phenomenon by the tone-lowering effect of laryngealization. This is even more obvious in the Bikin variety. Consider the pronunciation of *od'o* 'grandfather' by a speaker of the Sjain variety in Figure 14.

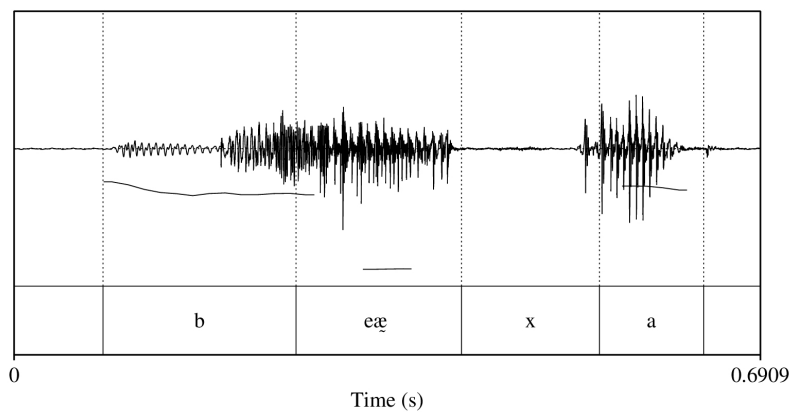


Figure 13: Bikin variety. Speaker Nadežda Kukčenko: *b'æsa* 'small river' [bæʂxa]

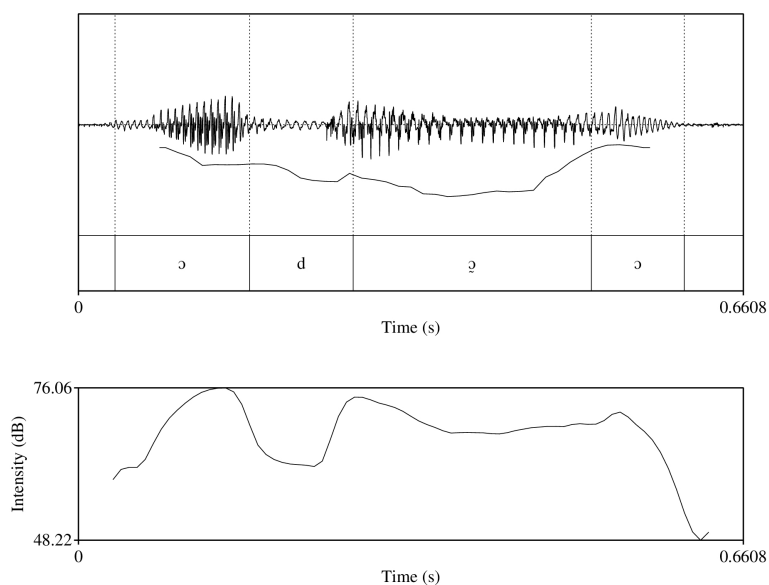


Figure 14: Bikin variety, speaker Alexandr Kančuga: *od'o* 'grandfather' [ɔdɔ]

7 From consonant to tone: Laryngealized and pharyngealized vowels in Udihe

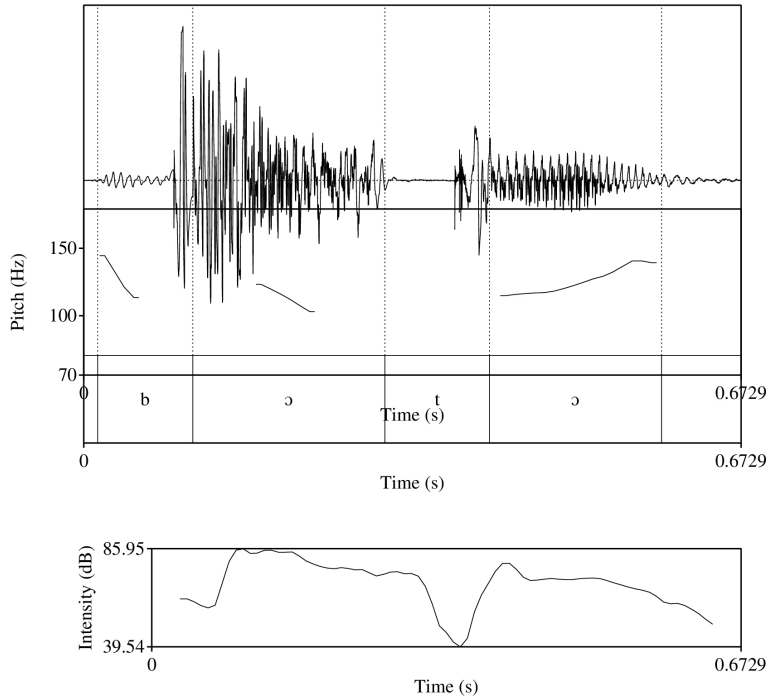


Figure 15: Bikin variety, speaker Alexandr Kančuga: *b'oto* 'ligneous mushroom' [bòtò]

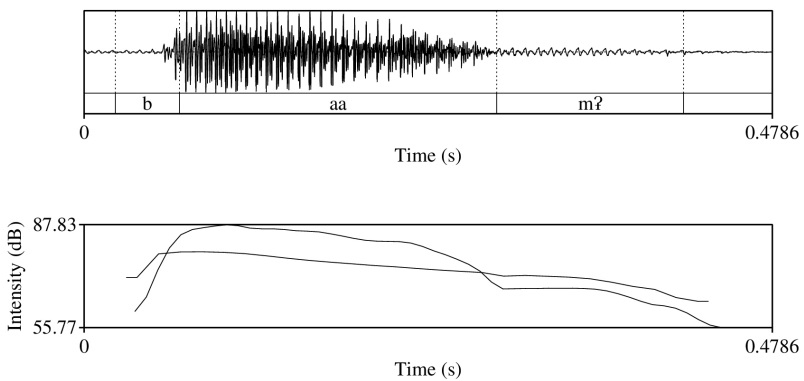


Figure 16: Bikin variety. Speaker Alexandr Pionka: [baamⁱ] 'I met'

Figure 14 shows a word pronounced in the full mode with creaky phonation. It shows that the laryngealized vowel is characterized by a high intensity and lowering of the pitch.

In the Xor variety low pitch was a side-effect of vowel laryngealization. In the Bikin variety, due to the gradual loss of creaky phonation in allegro mode, low pitch accompanied by a high intensity of pronunciation became the main distinctive feature of laryngealized vowels in some idiolects. Consider the following example: the word *b'oto* 'ligneous mushroom' pronounced by the same speaker.

The examples in Figures 14 and 15 present different tones (pitch movements): $\grave{\text{a}}-\grave{\text{a}}$ and $\grave{\text{a}}-\acute{\text{a}}$.

Tone raising on the second syllable as shown in Figure 14 was described by Šneider in 1936 who interpreted it as an exponent of musical accent in Udihe (Šneider 1936: 92). See details in Nikolaeva (2000: 134–137). This interpretation seems erroneous, since accent (stress) is connected with the hierarchy of syllables in a word. In Udihe, a word is characterized rather by a melodic pattern, which is closer to tone than to stress. Thus, we may conclude that in the vocalic systems of southern Udihe, tonal systems are under formation. This is most obvious in bisyllabic and polysyllabic words; however, it is also characteristic of monosyllabic words which have at least two moras (Simonov 1988). Consider the following example: the verb *b'aami* 'I met' pronounced by a speaker of the Olon variety which is the most innovative one and where phonation was lost.

The laryngealized vowel is realized by a sharp raise of intensity together with low tone, in other words, on the suprasegmental level. The change "from consonant to tone" is complete.

5 Discussion

Juha Janhunen suggested the appearance of tonal distinctions in Udihe are due to the Chinese influence (Janhunen 1999). His argument could be summarized as follows: a) Udihe is the southernmost of the Tungusic languages, and it was in contact with Chinese which is tonal; b) generally, tones in many Asian languages have arisen as "suprasegmental compensation" for the loss of segment sequences; c) four types of vowels of Udihe correspond to four tones of Chinese, as also noted by Radčenko (1988: 104). Janhunen pointed out that Chinese tones also have complex realizations and are characterized not only by changes in pitch, but also by the duration and the presence of different types of phonation. There are certain objections to this explanation.

First, the contraction of the V-s-V and V-q-V segment chains into a single complex vowel is already characteristic of the Koppi transitional dialect. And there was no Chinese influence in Koppi.

Second, the four types of vowels which correspond, according to Janhunen, to four tones of Chinese are found only in varieties of the northern dialect cluster. And, these are the varieties which were much less affected by Chinese influence than Bikin and Iman Udihe. Indeed, many features of the southern varieties can be explained by intensive contact with Chinese (for more details see Perekhval'skaya (2001). However, the influence of Chinese manifested itself, rather in the general trend to analytism, which was also noted by Tsumagari (2012: 83–84) and in a certain “simplification” of the system: alignment of paradigms by analogy, etc.

Still, the origin of Udihe vocalism is hard to explain. Using Edward Sapir's term “drift”, it can be said that in Udihe and in neighbouring Oroch varieties there was an influence of a certain “constant factor”. This had to be some peculiarities of articulation that were not characteristic of other Manchu-Tungusic languages. A large amount of Udihe common words are of non-Tungusic origin: *gə^hə* ‘bad’, *du^hi* ‘brain’, *ʒa^hi* ‘wild boar’, *tə^hu* ‘all’, ‘*ana* ‘boat’, ‘*d'a* ‘cotton wool’; *gob'o* ‘a fly’, ‘*asa* ‘bay’; *t'an^hki* ‘middle’, *s'ai* ‘salt’, *kæfakta* ‘firewood’ as well as the word *asasa* ‘thank you’ and some others. On a rather cautious assumption, Udihe was influenced by a non-Tungusic language, previously present in this area, but not Chinese.

6 Conclusions

In the dialects of the southern cluster, three types of vowels correspond to the four types of vowels characteristic of the northern dialect cluster of Udihe.

The decrease in pitch on a laryngealized vowel is characteristic of all Udihe varieties, but in the dialects of the northern cluster the low pitch was a side-effect of vowel laryngealization. In Bikin and Iman, due to the gradual loss of creaky phonation in allegro mode, low pitch became the main distinctive feature of laryngealized vowels in some idiolects.

All Udihe varieties are characterized by a specific prosodic structure of the word. The word has minimally two moras, and consist of an initial and final rhythmic part that differ by suprasegmental pattern. This was noted by researchers of Udihe before (Simonov 1988). Still it was often interpreted in terms of “stress” (accent): Šneider and Sunik wrote that an Udihe word has two stresses, one of which falls on the initial syllable of the word, and the other on the final

syllable (Šneider 1936; Sunik 1968). The term stress or accent is not appropriate here. It is the prosodic structure of the word that is contrastive: words with similar segmental chains can differ by their prosodic structure.

Contrastive prosodic patterns depend on the presence of a laryngealized vowel and on its place in the word. Contrastive prosodic patterns are, in fact, linguistic tones. Thus, Udihe and especially its southern varieties became a tonal language of the type of languages with low tone density, like Scandinavian dialects or Latvian.

Further research is hindered by the fact that suprasegmental patterns in modern versions of Udihe are lost due to the influence of Russian.

The study of Udihe varieties shows how conventional the line between synchronic and diachronic descriptions of language can be. A synchronic description and comparison of modern varieties can shed light on the history of these varieties.

The study of the Udihe dialect continuum reveals the internal mechanisms of language change. It becomes obvious that in all territorial varieties of the Udihe language, similar trends acted, but in different areas they appeared with different degrees of intensity. The internal “structure” of the dialect continuum has been demonstrated: the allegro-style of one dialect corresponds to the full style of the neighbouring dialect, which produces a new allegro-style, and so on.

It is easy to see that each dialect is an independent system that is not reducible to the system of another dialect. At the same time, mutual understanding between speakers of different dialects is preserved and can be quite easy.

As a result of these considerations, it becomes clear that the idea of the unity of “language” (the concept of “such and such a language”) in the absence of codification is often misleading and causes disputes among linguists. “One language” is an abstraction. In reality, there are specific systems – idiolects that can be combined into dialects, language varieties and separate languages. However, the higher the taxon, the more likely it is that the various systems are combined. The foregoing does not apply only to cases where the “language” means a codified norm.

Abbreviations

Sim.	data from Simonov & Kialundziuga (1998)	C	any consonant
Shn.	data from Šneider (1936)	ʋ	vowel with creaky voice phonation
AM	allegro mode of pronunciation	˘	vowel bearing low tone
FM	full mode of pronunciation	ʋ̃	vowel with breathy voice phonation
V	any vowel		

Acknowledgements

I use this opportunity and thank all the participants of my field trips to the Russian Far East. I would like to express my particular gratitude to Natalia Kuznetsova, who helped to collect high-quality data on the phonetics of Udihe varieties, as well as to Kirill Maslinskij, who greatly helped in the interpretation of the material.

The study was supported by the Russian Science Foundation, grant 20-18-00250 “Tonal languages of the world: on-line data base and atlas”.

References

- Alonso de la Fuente, José Andrés. 2017. An Oroch word-list lost and rediscovered: A critical edition of Tronson’s 1859 pseudo-Nivkh vocabulary. *Bulletin of SOAS* 80(1). 97–117.
- Arsen’iev, Vladimir K. 1947-1949. *Sočinenija v 6-ti tomah [Works in 6 volumes]*. 5 vols. Vladivostok.
- Avrorin, Valentin A. & Boris V. Boldyrev. 2001. *Grammatika oročskogo jazyka [Grammar of Oroch]*. Novosibirsk: RAN.
- Avrorin, Valentin A. & Elena P. Lebedeva. 1968. Oročskij jazyk [The Oroch language]. In Petr Ja. Skorik (ed.), *Jazyki narodov SSSR [Languages of the USSR]*, vol. 5, 191–209. Leningrad: Nauka.
- Avrorin, Valentin A. & Elena P. Lebedeva. 1978. *Oročskije teksty i slovar’ [Oroch dictionary and texts]*. Leningrad: Nauka.
- Belikov, Vladimir I. & Elena Perekhvalskaya. 1994. Tazov jazyk [The Tazy language]. In *Krasnaja kniga narodov Rossii. Entsiklopedičeskij slovar’-spravočnik*, 50–51. Moscow: Academia.
- Brailovskij, Sergej N. 1901. Tazy ili udihe. Opyt istoriko-ëtnografičeskogo issledovanija [Tazy or Udihe. A historical and ethnographic research]. *Zhivaja starina* 1-4. 129–216, 323–433.
- Cincius, Vera I. 1949. *Sravnitel’naja fonetika tunguso-man’čžurskix jazykov [Comparative phonetics of the Tungus-Manchu languages]*. Leningrad: Uchpergiz.
- Cincius, Vera I. 1975/77. *Sravnitel’nyj slovar’ tunguso-man’čžurskix jazykov: materialy k etimologičeskomu slovarju [Tungus-Manchu comparative dictionary. Materials for the etymological dictionary]*. 2 vols. Leningrad: Nauka.
- Emeljanov, Alexandr A. 1927. *Severnoje poberež’je primor’ja. iz pojezdki letom 1926 goda [Primorye’s Northern coast. Trip in the summer of 1926]*. *Ekonomičeskaja žizn’ Dal’nego Vostoka*.

- Girfanova, Albina Kh. 2001. *Slovar' udègejskogo jazyka [Udihe dictionary]*. St. Petersburg: Nauka.
- Goldsmith, John A. 1976. *Autosegmental phonology*. Cambridge: Massachusetts Institute of Technology (MIT). (Doctoral dissertation).
- Hözl, Andreas. 2015. A typology of negation in Tungusic. *Studies in Language* 39(1). 117–157.
- Hözl, Andreas. 2018. Udi, Udihe, and the language(s) of the Kyakala. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 15. 111–146.
- Janhunen, Juha. 1999. Tonogenesis in Northeast Asia: Udeghe as a tone language. *Journal de la Société Finno-Ougrienne* 88. 31–38.
- Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.
- Kormušin, Igor V. 1998. *Udyxejskij jazyk [The Udihe language]*. Moscow: Nauka.
- Lar'kin, Viktor G. 1959. *Udègejtsy. istoriko-ètnografičeskij očerk s serediny xix v. do nashih dnei [The Udihe. Historical and ethnographic essay from the middle of the 19th century to the present day]*. Vladivostok, Moscow: Institut ètnografii AN SSSR.
- Leontovič, Sergej. 1898. *Kratkij russko-orochenskij slovar' s grammatičeskoj zametkoj. narod bassejna reki Tumnin, vpadajuščej v Tatarskij proliv, severnee Imperatorskoj gavani [Concise Russian-Orochon dictionary with a grammatical sketch. People of the basin of the Tumnin River, which flows into the Tatar Strait, north of the Imperial Harbor]*. Vladivostok.
- Margaritov, Vasilij P. 1888. *Ob oročax Imperatorskoj gavani [On the Oroch of the Imperial Harbor]*. St. Petersburg.
- Nadarov, Ivan P. 1887. Severno-ussurijskij kraj (s kartoju) [North Ussuri Krai (with map)]. *Zapiski Imperatorskogo Russkogo Geografičeskogo občestva* 17(1).
- Nikolaeva, Irina. 2000. The vocalic system of Udihe. *Eurasian Studies Yearbook* 72. 114–142.
- Nikolaeva, Irina & Maria Tolskaya. 2001. *A grammar of Udihe* (Mouton Grammar Library 22). Berlin: Mouton de Gruyter.
- Perekhval'skaya, Elena. 2001. *The Bikin-river Sprachbund*. Paper presented at Sprache 23. Jahrestagung und Kognition der Deutschen Gesellschaft für Sprachwissenschaft, 2001.02.28-03.02, Leipzig.
- Perekhval'skaya, Elena. 2010. Aspirirovannye i glottalizovannye glasnye v dialektax udegejskogo jazyka [Pharyngealized and laryngealized vowels in Udihe dialects]. *Voprosy jazykoznanija* 6. 60–87.

- Perekhval'skaya, Elena. 2016. Udègejskij jazyk [the Udihe language]. In Vida J. Mikh'alchenko (ed.), *Jazyk i obščestvo. Entsiklopedija [Language and Society. Encyclopaedia]*, 514–521. Moscow: Azbukovnik.
- Pevnov, Aleksandr M. 2012. The problem of the localization of the Manchu-Tungusic homeland. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 17–40. Wiesbaden: Harrassowitz.
- Protodjakonov, Prokopij. 1888. *Kratkij russko-oročenskij slovar'* [Brief Russian-Oroch dictionary]. Kazan.
- Przevalskij, Nikolaj M. 1990 [1870]. *Putešestvie v ussurijskom kraje, 1867-1869 g.* [Travel in the Ussuri region in 1868-1869]. Vladivostok: Dal'nevostočnoje knižnoje izdatel'stvo.
- Radčenko, Galina L. 1988. Spornye voprosy fonetiki udègejskogo jazyka [Controversial questions of the Udihe phonetics]. In Mikhail D. Simonov (ed.), *Istoričeskije i tipologičeskije issledovanija po tunguso-man'čžurskim jazykam* [Historic and typological studies on Tungus-Manchu languages], 35–43. Novosibirsk: SO AN SSSR.
- Schmidt, Peter. 1928. The language of the Oroches. *Acta Universitatis Latviensis* 17. 17–62.
- Simonov, Mikhail D. 1988. Fonologičeskaja sistema udègejskogo jazyka [Phonological system of Udihe]. In Mikhail D. Simonov (ed.), *Historic and typological studies on Tungus-Manchu languages*, 44–88. Novosibirsk: SO AN SSSR.
- Simonov, Mikhail D. & Valentina Kialundziuga. 1998. *Slovar' udegejskogo jazyka v 3-h tomax* [Dictionary of Udihe in 3 volumes]. Stenshev.
- Simonov, Mikhail D., Valentina T. Kialundziuga & Marina M. Khasanova. 1998. *Fol'klor udegejtsev: nimanku, tenungu, exe* [Udihe folklore: nimanku, tenungu, exe]. Novosibirsk: Nauka.
- Šneider, Evgenij R. 1936. *Kratkij udejsko-russkij slovar' s priloženijem grammatičeskogo očerka* [Concise Udihe-Russian dictionary with a sketch on grammar]. Moscow, Leningrad: Učpedgiz.
- Šneider, Evgenij R. 1937. *Materialy po jazyku anjujskix ude* [Materials on the language of the Aniu Udihe] (Naučno-issledovatel'skaja asociacija Instituta Narodov Severa im. I.G. Smidoviča 11). Moscow, Leningrad.
- Šrenk, Leopold I. 1883. *Ob inorodtsah Amurskogo kraja* [On indigenous people of the Amur Krai]. St.Petersburg: Imperatorskaja akademija nauk.
- Startsev, Alexandr F. 2004. Nekotoryje aspekty problemy etnogeneza udegejtsev [Some aspects of the ethnogenesis of the Udihe]. *Rossija I ATR* 2. 75–86.

- Suliandziga, Rodion V., Daria Kudriashova & Pavel V. Suliandziga. 2003. *Korennye maločislennye narody Severa, Sibiri i Dal'nego Vostoka Rossijskoj Federatsii. Obzor sovremennogo položennja* [*Indigenous peoples of the North, Siberia and the Far East of the Russian Federation. Overview of the current situation*]. Moscow: North Indigenous Peoples Assistance Center.
- Sunik, Orest P. 1968. Udegejskij jazyk [The Udihe language]. In Petr Ja. Skorik (ed.), *Ĵazyki narodov SSSR*, vol. 5, 149–171. Leningrad: Nauka.
- Trudgill, Peter. 1985. *On dialect. Social and geographical perspective*. New York: New York University Press.
- Tsumagari, Toshiro. 2012. A note on Udihe phonology from an areal-typological perspective. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 81–88. Wiesbaden: Harrassowitz.
- Zinder, Lev R. 1948. Eksperimental'noe izučennje fonetiki severnyx jazykov [An experimental study of the phonetics of the northern languages]. *Izvestija Akademii Nauk SSSR. Otdelenie literatury i jazyka* 7(6). 579–581.

Chapter 8

Proto-Tungusic in time and space

Martine Robbeets

Max Planck Institute for the Science of Human History & Johannes Gutenberg
Universität Mainz

Sofia Oskolskaya

Institute for Linguistic Studies, Russian Academy of Sciences

Although there is a general consensus among historical comparative linguists that the Tungusic languages are genealogically related and descend from a common ancestral language, the internal structure of the family, its age, homeland and pre-historic cultural context remain subject to debate. In addition to four competing concepts of classification, the linguistic literature yields a wide range of time estimations for the family covering more than a millennium as well as four different proposals with regard to the location of the homeland covering Eastern Siberia and Manchuria. Here we will combine the power of traditional comparative historical linguistics and computational phylogenetics to shed light on the prehistory of the Tungusic languages. Our aim is to build on a recent Bayesian verification of the Tungusic family and examine its implications in determining a plausible time depth, location and cultural context of the ancestral proto-Tungusic speech community. We will compare spatial inferences based on two different statistically well-supported Tungusic classifications, namely one in which the break-up of Manchuric constitutes the first split in the family as well as a North-South classification with a northern branch including Even, Evenki, Negidal, Oroqen, Solon, Oroch and Udehe as opposed to a southern branch including Manchuric and Nanaic languages. Situating Proto-Tungusic in time and space, we will estimate the break-up of Proto-Tungusic in the beginning of the first millennium and place its homeland in the area around Lake Khanka. Our study pushes the field forward in answering some tantalizing questions about the prehistory of the Tungusic family, providing a quantitative basis for some conflicting hypotheses and in triangulating linguistics, archaeology and genetics into a holistic approach to the Tungusic past.



1 Introduction

The Tungusic language family is distributed over a vast area in China and Russia, ranging from the Sea of Okhotsk in the east to the Yenisei Basin in the west, and from the Bohai Sea in the south to the Arctic Ocean in the north. Figure 1 shows the distribution of 12 Tungusic languages, notably Oroch, Udehe, Hezhe,¹ Nanai, Orok, Ulch, Xibe, Even, Solon, Evenki, Negidal and Oroqen, as well as dialectal varieties.² Whereas the four Nanaic varieties, Hezhe (Heilongjiang), Najkhin Nanai (Middle-lower Amur), Kur-Urmi Nanai (Khabarovsk) and Bikin Nanai (Ussuri), are so diverse that it is not clear whether they should be considered dialects or separate languages, the Mомsky and Olsky Even doculects show less internal variation. The map in Figure 1 further shows two historical varieties, notably Jurchen and Manchu. Since written materials in Jurchen, the now extinct language of the Jin dynasty (1115–1234), are only partially deciphered, the earliest well documented stage is Manchu, the official language of the Qing dynasty (1636–1911).

There is a general consensus that the Tungusic languages are genealogically related and descend from a common ancestral language, conventionally called “Proto-Tungusic”. However, due to the wide geographical distribution and the considerable internal variation of these languages, the internal family structure along with its root age and homeland are subject to debate. Here we will combine the power of traditional comparative historical linguistics and computational phylogenetics to shed light on the prehistory of the Tungusic languages. Our aim is to build on recent Bayesian analyses of the Tungusic family and examine their implications for determining a plausible time depth and location of the ancestral proto-Tungusic speech community. In addition, we would like to shed light on the factors that drove early Tungusic language spread.

To this end, we will organize our paper as follows. In §2, we will summarize how the recent application of Bayesian inference methods was able to quantify the reliability of previously proposed classifications of the Tungusic family. In §3, we will compare the time range of Tungusic, previously inferred by various linguistic dating techniques, against the quantitative basis provided by Bayesian analysis. In §4, we will test various competing hypotheses concerning the possible homeland of Proto-Tungusic, applying the diversity hotspot principle on the best supported tree models. Finally, we will map our linguistic inferences about Tungusic prehistory on findings from archaeology and genetics in a holistic approach, for which we use the term “triangulation”.

¹We use the term *Hezhe* as a cover term for the Kilen and Hezhen dialects and used sources for both dialects.

²We used Natural Earth vector map data for the maps printed in this chapter, which are available in the public domain from <https://www.naturalearthdata.com/>.

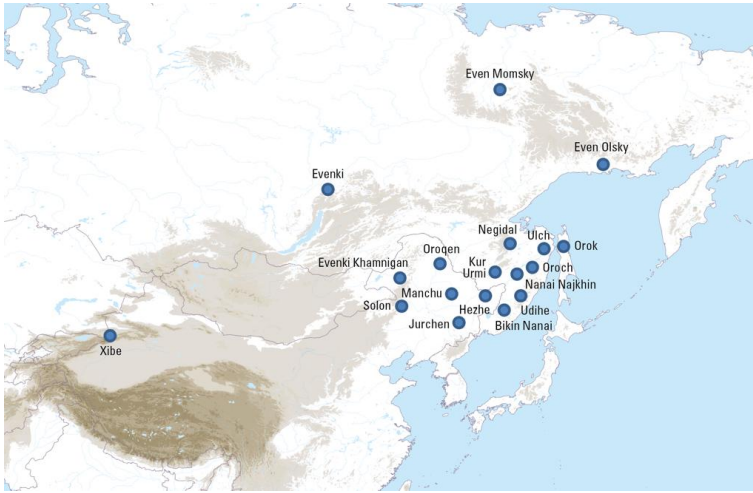


Figure 1: The distribution of the Tungusic languages

2 Family structure

2.1 Previous classifications

Previous studies of the internal taxonomy of the Tungusic family reach different results, particularly with respect to the early separation of the Manchuric (Jurchen, Manchu, Xibe) branch.³ Cincius (1949), Benzing (1956), Menges (1968: 27) and, more recently, Kormušin (1998), Georg (2004) and Janhunen (2012) proposed a binary north-south classification, in which the separation of Manchuric from the other Tungusic languages does not constitute the earliest split in the family (Figure 2a). The classical approaches by Cincius, Benzing and Menges separated a Northern branch consisting of Evenki, Even, Solon and Negidal from the rest of the Tungusic languages, while the more recent approaches added Oroch and Udehe to the Northern branch instead (Figure 2b). Sunik (1959: 333–335), Vasilevič (1960: 44), Doerfer (1978: 5), Vovin (1993: 102), Whaley et al. (1999: 291), Robbeets (2015), Robbeets & Bouckaert (2018), Dybo & Korovina (2019) and Whaley & Oskolskaya (2020) all argued for an early breakup between Manchuric and the rest of Tungusic, even if their precise configurations do not overlap in each detail (Figure 2c). Moreover, Ikegami (1974) proposed a polytopology, distinguishing as many as four branches, namely Manchuric, Evenic, Nanaic and Udeheic (Figure

³The term Manchuric is used elsewhere in literature to designate this branch, see among others Alonso de la Fuente (2011) and Robbeets & Savelyev (2020).

2d).⁴ The basic configuration of these four topologies is represented in Figure 2.

Except for Dybo & Korovina's (2019) classification, which is based on lexicostatistics, the majority of previous research is based on the traditional Maximum Parsimony method. This approach used by historical comparative linguists, seeks a tree that explains a dataset by minimizing the number of evolutionary changes required to produce the observed data. Lexicostatistics is an early and less reliable form of statistical tree-building, which uses the shared cognate proportion in a basic vocabulary list as a distance metric to estimate linguistic relationships.

The Bayesian method seeks to explain a set of observed data by quantifying how likely it is that they have been produced by a certain model. As this is a statistical approach in which all forms of uncertainty are expressed in terms of probability, it can contribute to the current state of the art by verifying which of the models in Figure 2 is best supported by the data by quantifying the statistical robustness of different proposals and by inferring absolute divergence dates. In this way, a Bayesian approach can provide a quantitative basis for previous classifications based on classical historical linguistic approaches. Here we aim at interpreting the results of a recent Bayesian analysis of the Tungusic family (Oskolskaya et al. 2022) and at inferring spatiotemporal and cultural patterns of Tungusic linguistic dispersal.

2.2 A recent Bayesian approach to the classification of the Tungusic languages

Oskolskaya et al. (2022) took a Bayesian approach to the classification of the Tungusic languages, based on the dataset of 254 basic vocabulary items collected for 21 Tungusic varieties. The maximum clade credibility tree in Figure 3, which is the best supported tree among all trees generated by applying different evolutionary models, summarizes the results of this study.⁵

The trees underlying Figure 3 were generated running the software BEAST 2.4.7 (Bouckaert et al. 2014), which only allows for a binary structure of splits. Jurchen, which is usually considered as a direct ancestor of Manchu is represented as a separate branch because a written standard is never considered directly ancestral to a spoken variety. The model thus assumes that Jurchen and Manchu-Xibe have separated from a common ancestor, spoken at some point in the past.

⁴Some authors, such as Ikegami and Janhunen used a different terminology for these groupings, but for reason of accessibility, we use a consistent terminology here.

⁵All datasets and coding details are accessible through the supplementary information in Oskolskaya et al. (2022).

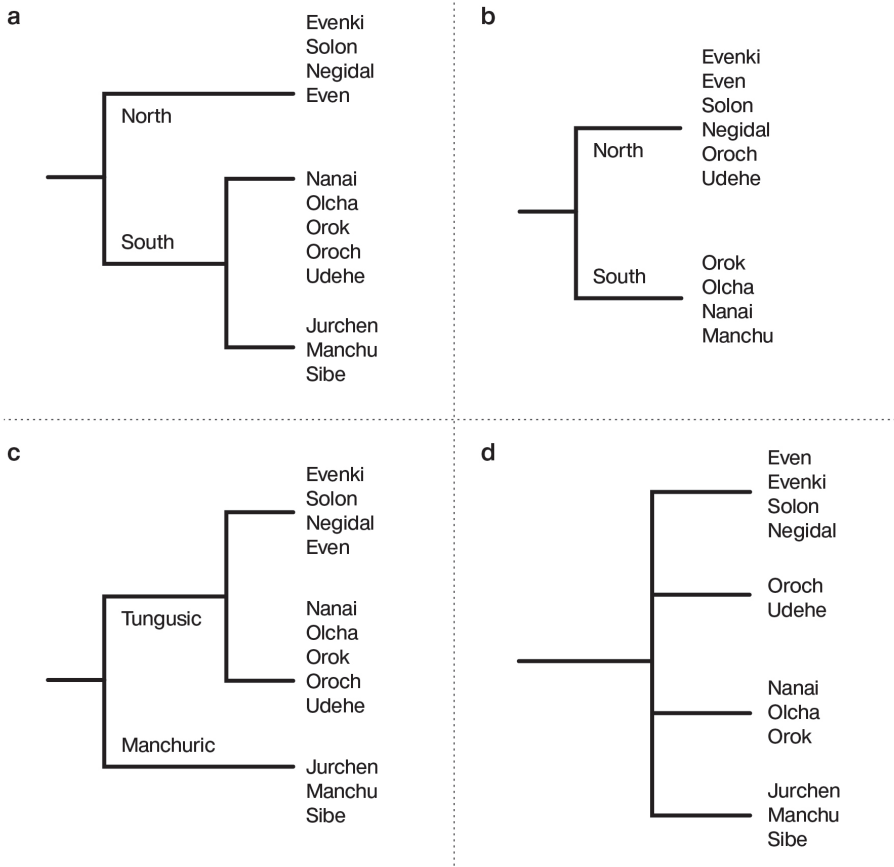


Figure 2: Basic configuration distinguishing four topologies for the Tungusic family: a. Classical North-South classification; b. Revised North-South classification; c. Manchu-Tungusic classification; d. Quadruple topology

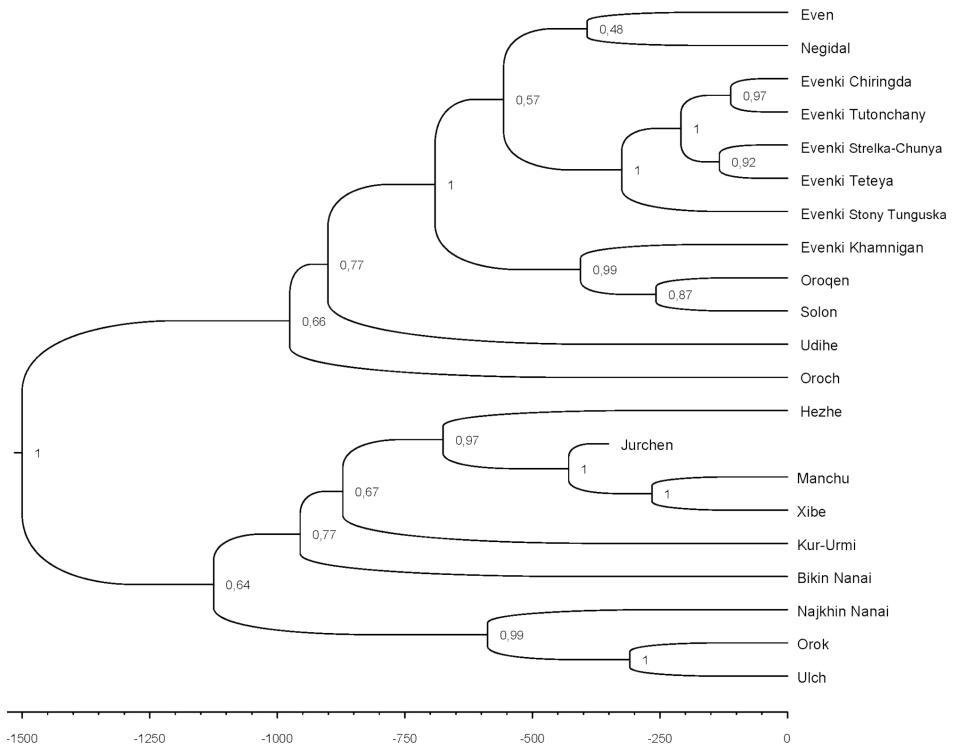


Figure 3: Maximum clade credibility tree (Binary covarion, no gamma variation, relaxed clock) for the Tungusic family (Oskolskaya et al. 2022)

The numbers on the nodes show the posterior probability, which qualifies the statistical robustness of each clade. The higher the number, the more probable the existence of the clade. We can thus safely establish a Northern Tungusic branch (posterior probability = 1), a Nanaic branch with Najkxin Nanai, Orok and Ulch (posterior probability = 0.99) and a Manchuric branch, probably including Hezhe besides Jurchen, Manchu and Xibe (posterior probability = 0.97). With posterior probabilities below 0,80, the exact position in the tree of Udihe, Oroch, Kur-Urmi, Bikin Nanai and the cluster Orok-Ulch-Nakjin Nanai is less secure, but it is interesting to note that the tree in Figure 3 does not support a monophyletic cluster composed by Oroch and Udihe.⁶

Taking into account the probabilities of the branches, the tree in Figure 3 supports two basic classifications previously proposed in the literature, the revised

⁶Perekhval'skaya (2022 [this volume]) proposes that Udihe, Oroch, and Kyakala are the ends of a continuum.

North-South classification (Figure 2b) and the Manchu-Tungusic classification (Figure 2c). According to Oskolskaya et al. (2022), the revised North-South classification with a binary split between Manchuric and Nanaic on the one hand and Northern Tungusic, Udehe and Oroch on the other is best supported (i.e. in 48,3% of generated trees), while the Manchu-Tungusic classification (Figure 2c) is also highly probable (31,1% of trees). The other two classifications, namely the classical North-South and the quadruple classification presented in Figures 2a and 2d, are excluded by this analysis.

3 Age

3.1 Previous dating

As shown in Table 1, different linguistic dating principles yield a time range for the primary breakup of Proto-Tungusic between 950 BC and AD 700. Applying lexicostatistic methods, Dybo & Korovina (2019) dated Proto-Tungusic to around 950 BC, while Korovina (2011) dated it to the sixth century BC, but other distance-based methods, such as the Automated Similarity Judgment Program (ASJP) yielded much younger dates, notably AD 681 (Holman et al. 2011: 854). Distance-based methods use a distant metric, such as the lost cognate proportion or the number of operations required to turn one string of phonemes into another, to infer the time depth of language separation. However, as they assume a constant rate of loss over time, their results are not generally accepted. Another loose dating principle that is not entirely foolproof is tracing the primary break-up of Proto-Tungusic back to certain ethnonym shifts. Referring to the name change in Chinese dynastic chronicles of the Tungusic ethnonym “Yilou” to “Wuji”, Robbeets (2015: 16–18) situated the break-up of Proto-Tungusic at the end of the Han period (206 BC–AD 220). On the basis of a rough measure of mutual intelligibility, Pevnov (2012: 32) estimated that Proto-Tungusic could not be younger than two thousand years.

Reconstructing the vocabulary of a proto-language, we can examine the cultural and ecological concepts revealed in it. The time when some of these concepts became available to the speakers of the proto-language can also serve as an indication for the language family’s time depth. In line with Janhunen’s (2012: 8) findings, an Iron Age dating of Proto-Tungusic is supported by the reconstruction of PTg **sele* ‘iron’, reflected in Evenki *sele*, Even *hel*, Neg. *sele*, Solon *sele*, Xibe *selə*, Manchu *sele*, Jur. **sele*, Ulch *sele*, Orok *sele*, Nanai *sele*, Oroch *sele* and Udihe *sele*. Although a single reconstruction, which is not backed up by other vocabulary items, cannot provide us with a reliable chronology, the Iron Age dating

of Proto-Tungusic is further corroborated by contact linguistics. As words have travelled between languages since prehistoric times, giving voice to new ideas and names to new products and practices, we expect to observe links between cultural diffusion and prehistoric borrowing. One possible candidate for such a link is PTg **murgi*, the reconstructed term for ‘barley and similar crops’. There is reason to believe that this word is borrowed from an Old Chinese donor word 來 **mərʔək* > **mərʔə* ‘a kind of wheat’ (Robbeets 2017b: 28–29). The linguistic reconstruction can be correlated to the archaeological evidence for barley being first imported through Chinese contact at the time of the Krounovskaya culture (600 BC–AD 200), situated in the Southern Primorye around Lake Khanka (Sergusheva & Vostretsov 2009: 214–215). This culture also marks the beginning of the Early Iron Age in the Russian Far East with the first uncontested finds of iron. Therefore, on the basis of contact studies, the break-up of Proto-Tungusic has been dated to the period between 600 BC and AD 200.

Table 1: The estimated time of separation of Proto-Tungusic according to different linguistic dating principles

Dating principle	Source	Time depth
Lexicostatistics	Dybo & Korovina 2019	950 BC
Lexicostatistics	Korovina 2011	600 BC
ASJP	Holman et al. 2011	AD 681
Ethnonym shift	Robbeets 2015	206 BC–AD 220
Intelligibility	Pevnov 2012	AD 1
Cultural reconstruction	Janhunen 2012	500 BC–AD 500
Contact linguistics	Robbeets et al. 2020	600 BC–AD 200

3.2 Dating through Bayesian inference

The Bayesian method calibrates the divergence time of the root and the nodes in a language family against known cases of language divergence over attested timespans and quantifies how likely it is that the inferred time depth falls within a certain density interval. For calculating the divergence time of the root and the nodes in Tungusic tree, Oskolskaya et al. (2022) calibrated the tree in three nodes.

First, they estimated a time depth for the transition from Jurchen into Manchu and inserted it as a calibration point. To this end, they used the time of the first known Manchu manuscript which is dated to 1599 (Gorelova 2002: 50). The logic

is that Jurchen is no longer attested by that period and had thus ceased to exist before 1599. Therefore, 351 years before present (conventionally before 1950) was used as an estimation for the separation time between Jurchen and Manchu–Xibe. However, as the model assumes that Jurchen and Manchu–Xibe already separated before the time that Jurchen was first attested, thus before 1185, a separation time that aligns with the model should be at least 765 years before present. This difference of at least 400 years is expected to yield a later date for the calculated time depth of Proto-Tungusic than the real one.

The second calibration point is the time of the split between the Xibe and Manchu languages, 186 years ago. This is based on the dating of the resettlement of Xibe populations to the northwest of China in 1764 (Gorelova 2002: 31).

The third calibration point is the time of the first break-up of Evenki, which had taken place already before 1723, i.e. more than 227 years ago. According to Vasilevič (1969), D. G. Messerschmidt in his fieldnotes in 1723 provided vocabulary collected from various Evenki people in different regions. His data show that there were at least two dialects that can be associated with the modern distinction between the Northern versus the Southern and Eastern dialects. Thus, the first break-up of Evenki had already taken place by that time.

Using these three calibration points to calibrate the tree, Bayesian analysis infers a time depth for the root and nodes in the Tungusic tree, as shown in Figure 4.⁷

All dates in Figure 4 indicate a time “before present”, conventionally before 1950. Each bar shows a time range, in which a specific split has taken place with a 95% higher posterior density interval (HPDI). The time depth of the primary split is particularly relevant for our present study. The credible interval for this split covers about 1800 years (737 BC–1154 AD), which implies that there are not enough data for more precise results. Nevertheless, the median of this bar is around 1500 BP, i.e. AD 450.

It should be noted that the three calibration points refer to relatively recent events and present upper limits after which the separation cannot have taken place. This could lead to a bias in the estimation of the time-depth of the Tungusic tree, by which the estimated age would be younger than the real age. This effect is increased by the fact that the split between Jurchen and Manchu–Xibe should be calibrated at least 400 years earlier. Therefore, it is probable that the actual break-up of Proto-Tungusic took place several centuries before the inferred date of AD 450, probably in the beginning of the first millennium AD, as implied in the hypotheses provided by Janhunen (2012), Pevnov (2012), Robbeets (2015) and Robbeets et al. (2020).

⁷The detailed information of this analysis is described in Oskolskaya et al. (2022).

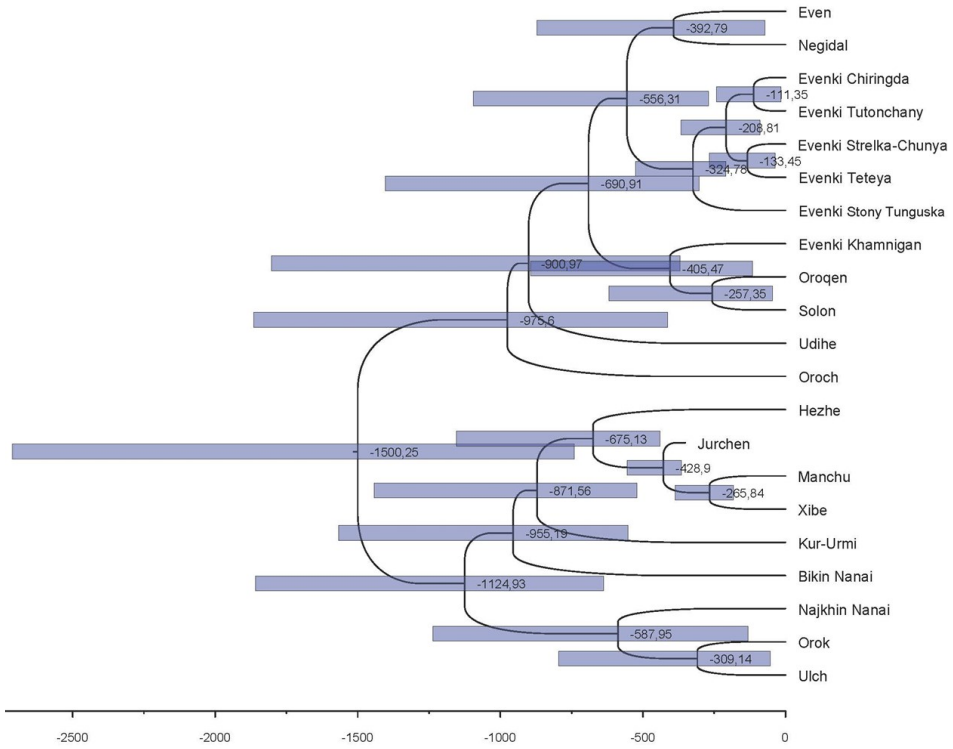


Figure 4: Bayesian age estimation for the nodes in the Tungusic tree (Oskolskaya et al. 2022)

4 Homeland

4.1 Previous proposals

As indicated on the map in Figure 5, there are four competing hypotheses with regard to the possible homeland of Proto-Tungusic, notably (1) the Baikal region (Vasilevič 1960; Menges 1968: 23; Derevyanko 1976; Helimski 1985: 279), (2) the Mid Amur and the lower part of the Upper Amur region (Tugolukov 1980; Janhunen 1996: 169; Korovina 2011; Pevnov 2012; Wichmann p.c., 2019.10.03; Pugach et al. 2016), (3) the region around Lake Khanka (Robbeets 2020; Wang & Robbeets 2020) and (4) the Yalu River region on the border between present-day Liaoning and Northern Korea (Janhunen 2012). The evidence in support of the Baikal region comes mainly from prehistoric contact linguistics assuming ethnolinguistic interaction with ancient speakers of Amuric, Samoyedic, Mongolic and Yeniseic. An original location in the Mid and Upper Amur region is supported by various

approaches: Janhunen (1996: 169) provides ethnolinguistic indications; Korovina (2011) reconstructs names for insects, reptiles, shellfish and fish; Pevnov (2012) combines toponyms with reconstructed river vocabulary and tree names; and Wichmann takes a computer-automated approach to the diversity hotspot principle. The region around Lake Khanka is supported by the diversity hotspot principle and cultural reconstruction, while the Yalu River region is proposed on the basis of a general northward trend of expansion.

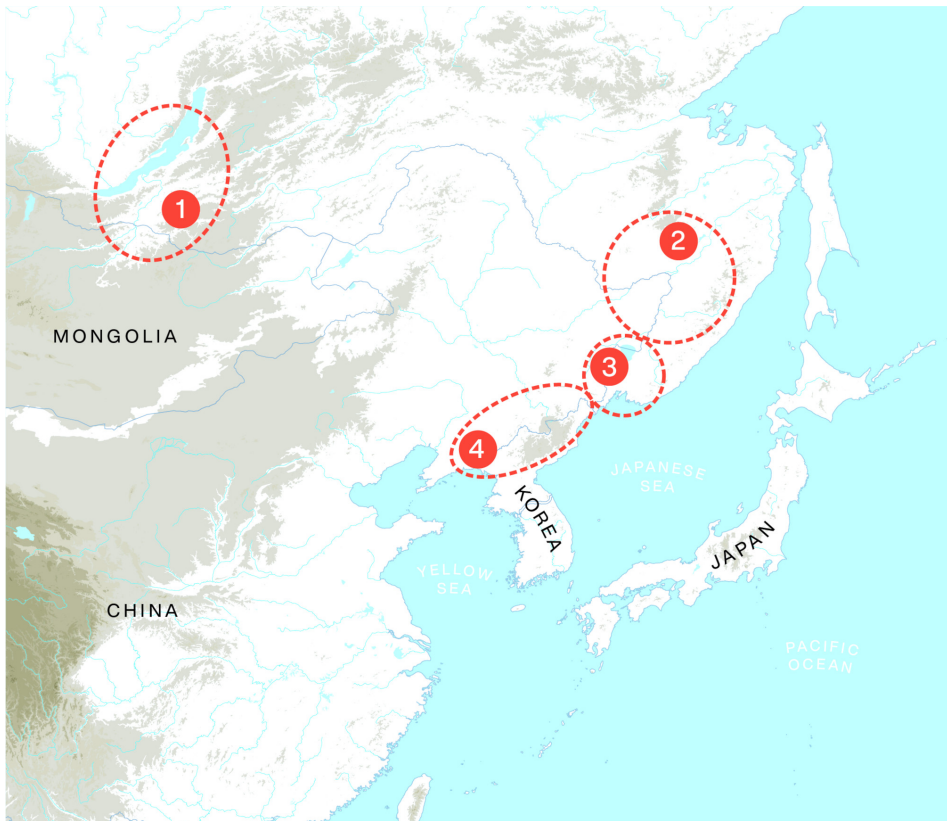


Figure 5: Proposed locations for the homeland of Proto-Tungusic: (1) the Baikal region; (2) the Mid Amur and the lower part of the Upper Amur region; (3) the region around Lake Khanka; (4) the Yalu River region

4.2 Diversity hotspot principle

The “diversity hotspot principle” is based on the assumption that the homeland is closest to where one finds the greatest diversity with regard to the deepest subgroups of the language family. It follows that the primary splits in the family are determinants for the location of the homeland. In the case of the Tungusic family, this implies that depending on which of the four classifications in Figure 2 we favor, the center of primary diversity – and thus also the inferred location of the homeland – will move on the map. The resulting locations are situated on the map in Figure 6. Whereas the classical north-south classification pushes the homeland to the north towards the Upper to Mid Amur region (Figure 6a), the revised north-south, the Manchu-Tungusic and the quadruple classifications pull it more southwards towards the Mid Amur region immediately north of Lake Khanka (Figure 6b), the area around Lake Khanka (Figure 6c) and Manchuria (Figure 6d), respectively.

Although the diversity hotspot principle can provide some clues about the homeland of a language family, it must also contend with several limitations (Wang & Robbeets 2020): the contemporary hotspot of linguistic diversity may diverge from the earlier one or the principle may be upset when a migration was suddenly directed over a long distance rather than representing slow, gradual and random movement into adjacent areas. The Xibe populations, for instance, were suddenly resettled to the west in the 18th century and northern Tungusic populations, such as the Evenki and Even, are extremely mobile. Nevertheless, we can gain more from applying the diversity hotspot principle to Tungusic, than we can lose from ignoring it, because it helps us to set up hypotheses about homelands and it makes us aware of the interconnection between internal family structure and the identification of a homeland.

Since our Bayesian analysis presented in Oskolskaya et al. (2022) and summarized in §2.2 supports the classifications represented in Figure 6b/c, it indicates that the original homeland of the Proto-Tungusic speech community was situated in the area around lake Khanka or immediately to the north of it. This location is corroborated by Wichmann et al.’s (2010) attempt to formalize the diversity hotspot principle in a computer-automated approach. Taking into account 11 Tungusic languages, notably Even, Evenki, Negidal, Oroqen, Naykxin Nanai, Oroch, Orok, Ulch, Udehe, Manchu and Xibe, they compute a diversity measure for each language and identify the homeland with the location of Naykxin Nanai, the language with the highest diversity measure.⁸ The location of Naykxin Nanai

⁸Wichmann et al.’s measure of diversity is derived as the proportion l/g of the linguistic distance l and the geographical distance g between two languages. By way of linguistic distance, they use the *Levenshtein distance*, which is the minimum number of substitutions necessary to transform one string of phonemes into another in a subset of 40 basic vocabulary items.

at latitude 49.28 and longitude 136.47 is equated with that of the Tungusic homeland. As more southern varieties such as Solon on the Nonni River in Inner Mongolia and Bikin Nanai in the southern part of the Primorye province as well as Hezhe in Heilongjiang are lacking from this analysis, we may expect the real diversity hotspot to be slightly more to the south than the inferred one.

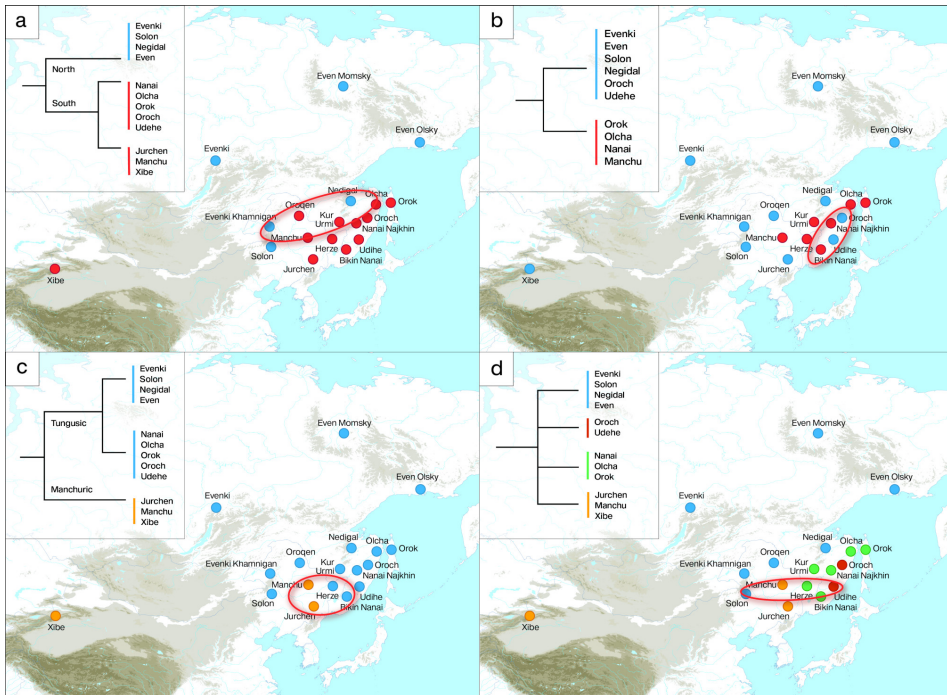


Figure 6: Diversity hotspot of the Tungusic languages under the four proposed classifications

5 Triangulation

5.1 Linguistics

Combining Bayesian inference with other linguistic approaches, we estimate that Proto-Tungusic was spoken in the area around or immediately north of lake Khanka in the beginning of the first millennium AD. As shown in Table 2, cultural reconstruction indicates that the speakers of Proto-Tungusic were familiar with agriculture. The use of words such as **pisike* ‘broomcorn millet (*Panicum*

miliaceum'), **jiya*- 'foxtail millet (*Setaria italica*)',⁹ **murgi* 'barley',¹⁰ **üse*- ~ *üsi*- 'to plant',¹¹ **üse* ~ *üsi* 'seed, seedling', **üsin* 'field for cultivation' and **tari*- 'to sow, plant, cultivate' implies that the speakers relied on plant cultivation for subsistence. The derivation of the Proto-Tungusic word **üse* ~ *üsi* 'seed, seedling' as a deverbal noun from the verb **üse*- ~ *üsi*- 'to plant' suggests that seeds were not just collected for consumption in the wild but that they were planted as part of a cultivation process. In addition, for some crop names such as 'barley' and 'broomcorn millet', we can argue that the word refers to the domesticated crop rather than to the wild variety of the plant because both crops are not native to the region and have been imported as domesticated crops. It is commonly assumed that broomcorn millet has been imported from the West Liao River region by the people who introduced the Zaisanovskaya culture (3200–1300 BCE) to the Russian far East (Sergusheva & Vostretsov 2009; Leipe et al. 2019; Li et al. 2020).

Whereas several agricultural terms are inherited, maritime vocabulary can often be explained as borrowed from Proto-Amuric or its descendant Nivkh. This

⁹The cognates Nanai *jie-kte* and Oroch *jie-kte* suggest front vocalization **jiye-kte* in Proto-Tungusic, while Negidal *ja:kta*, Solon *jakta* and Udehe *jakta* suggest original back vocalization **jiya-kta*. In addition to the vowel alternation in the in Oroch suffix *-kte* and *-kta*, this observation suggests that the alternation between **jiya-kta* and **jiye-kte* was already present at the Proto-Tungusic stage. It was probably due to assimilation with the initial front vowel in **jiya-kta*. Another example of such an alternation is found in Proto-Tungusic **niaya*- ~ *nিয়ে*-, which is reflected as *yaya*- or *ńaya*- 'to shamanize' in most Tungusic languages but as *leye*- 'to sing' in Manchu. Note that PTg **ye* regularly develops in Manchu *ye*, e.g. PTg **xeye*- 'to sink' and Ma. *eye*- 'to float, flow' (Cincius 1975/77: 440–442; Starostin et al. 2003) or PTg **jeye* 'sharp point; blade' and Manchu *jeyen* 'sharp point; blade' (Cincius 1975/77: 282–283; Starostin et al. 2003), etc. Manchu *je* 'foxtail millet (*Setaria italica*); grain' is thus considered a contraction from **jiye*.

¹⁰It is a universal tendency that consonant clusters are susceptible to variation through assimilation, metathesis or consonant loss. Consonant clusters across the Tungusic languages are no exception to this expectation. Although the reflex of the Proto-Tungusic **rg* cluster is relatively stable in most northern Tungusic languages, we find a variety of reflexes, such as *rg*, *gg*, *yg*, *yg* and *g* in Manchuric and other southern Tungusic languages, including an occasional *j* in Nanai, Oroch and Ulch, e.g., PTg **burga-kta* 'beard, moustache' reflected in Ulch *buja-qta*, Nanai *boja-qta*; PTg **serge-kte* 'nose bone' in Nanai *sejurē*; PTg **irge* 'brain, head' in Ulch *ije*, etc.

¹¹Given that the expected reflex of PTg **ü* is Udehe *i*, Udehe *uhi*- 'to sow, to plant a garden', *uhi* 'garden for cultivating plants' may represent cases of intra-Tungusic borrowing and the inherited reflex in Udihe may be *yehu*- 'to grow', given the palatal glide onset. Note however that, even if the majority reflex PTg **ü* is *i* in Oroch, there are a few instances where PTg **ü* is retained as *u* instead, e.g. PTg **erün* 'time' as Oroch *erū(n)* (Cincius 1975/77: 463–464; Starostin et al. 2003), PTg **xulbü*- 'to bind, arrange' as Oroch *ubbuna*- (Cincius 1977: 258; Starostin et al. 2003), PTg **xegün* 'nine' as Oroch *xuju(n)* (Cincius 1975/77: 352–353; Starostin et al. 2003), PTg **tüksa* 'house cover made of birch bark' as Oroch *tüksa* (Cincius 1975/77: 179; Starostin et al. 2003), etc. Therefore, this correspondence is included in the appended list of sound correspondences.

Table 2: The reconstruction of agricultural vocabulary in Proto-Tungusic (adapted from Robbeets et al. 2020: 765)

No.	Proto-Tungusic	Attested Tungusic words
(1)	* <i>pisike</i> 'broomcorn millet (Panicum miliaceum)'	Manchu <i>fisihe</i> ~ <i>fisike</i> 'glutinous millet, broomcorn millet (Panicum miliaceum)'; Ulch <i>pikse</i> ; Nanai <i>pikse</i> 'millet'; Kur-Urmi dialect <i>fisxe</i> 'broomcorn millet (Panicum miliaceum)'; Jurchen * <i>fise bele</i> 'yellow rice; coarse rice' (<i>bele</i> 'hulled rice, edible grain')
(2)	* <i>jiya-hta</i> ~ * <i>jiye-hte</i> 'foxtail millet (Setaria italica)'	Negidal <i>ja:kta</i> 'foxtail millet (Setaria italica), small millet; cleaned grain; oats flour'; Solon <i>jakta</i> 'porridge, food'; Xibe <i>je</i> 'foxtail millet (Setaria italica)'; Manchu <i>je</i> 'foxtail millet (Setaria italica); grain', Jurchen * <i>je bele</i> 'millet'; Nanai <i>jiekte</i> 'foxtail millet (Setaria italica)'; Oroch <i>jiekte</i> , <i>jekte</i> , <i>jiekta</i> 'millet', Udehe <i>jakta</i> 'foxtail millet (Setaria italica), porridge'
(3)	* <i>üse-</i> ~ * <i>üsi-</i> 'to plant' → * <i>üse</i> ~ * <i>üsi</i> 'seed, seedling' → * <i>üsi-n</i> 'field for cultivation'	Evenki <i>iħaw-</i> 'to grow (of people)'; Even <i>isu:-</i> ~ * <i>esu:-</i> 'to sprout, come out (of plants), blossom, grow', Negidal <i>isew-</i> 'to grow, become acclimatized (about plants); mature, grow up (about people)'; Xibe <i>use-</i> 'to sow seeds', <i>use</i> 'seed, grain', <i>usin</i> 'field, farmland'; Manchu <i>use-</i> 'to plant, seed (tr.)', <i>use</i> 'seed; insect egg', <i>usin</i> 'field for cultivation', Jurchen * <i>use</i> 'seedling', <i>usi-in</i> 'field', Ulch <i>use</i> 'seed', <i>usun</i> 'field, garden'; Nanai <i>use</i> 'seed', <i>usi</i> 'arable field for cultivation; private garden for cultivating vegetables', Oroch <i>usi</i> 'field (farm), garden', <i>usi-</i> 'to cultivate, till; hunt a bear'; Oroch <i>usi</i> 'seeds, grains', <i>usin</i> 'garden for cultivation', <i>usin-</i> 'to sow, to plant in a garden'; Udehe <i>yehu-</i> 'to grow', <i>uhi-</i> 'to sow, to plant a garden', <i>uhi</i> 'garden for cultivating plants'
(4)	* <i>tari-</i> 'to sow, plant, cultivate'	Evenki <i>tari-</i> ~ * <i>tare-</i> ~ * <i>tale-</i> 'to sow'; Solon <i>tari-</i> 'to sow seeds; to plant; to cultivate; to grow; to disseminate; to inject; to infect, to catch a disease', <i>tariŋko</i> 'injector'; Manchu <i>tari-</i> 'to cultivate, farm; to plow'; Jurchen * <i>tali-</i> ~ * <i>tari-</i> 'to sow, to plant, to cultivate'; Ulch <i>tari-</i> 'to sow, to plant'; Nanai <i>tari-</i> 'to sow seeds', <i>tariŋko</i> 'sowing machine', <i>tarici-</i> 'to sow (seeds) regularly'; Udehe <i>tali-</i> 'to plant a garden', <i>tali</i> 'garden for cultivating plants'
(5)	* <i>murgi</i> 'barley (Hordeum vulgare)'	Manchu <i>muji</i> 'barley (Hordeum vulgare)', Xibe <i>muji</i> 'barley (Hordeum vulgare)', Ulch <i>muji</i> 'oats', Nanai <i>muji</i> 'oats', Oroch <i>muji</i> 'barley', Jurchen * <i>miryei</i> 'product of agriculture', Solon <i>murgil</i> 'spring crops, spring-sown field'

is for instance the case for ancient loanwords, such as PTg **laamos* ‘wind (from the sea)’ from Proto-Amuric **lamos* > Nivkh *lams* ‘eastern wind’, PTg **kalimV* ‘whale’ from Proto-Amuric **kalimV* ‘whale’ > Nivkh *kalm* (*qalm*) ‘(small) whale’ (Janhunen 2016) and PTg **laska* ‘sea goby’ from Proto-Amuric **laskV* ‘goby’ > Nivkh *lask* ‘a goby (of middle size).

The direction of the borrowing is verifiably from Amuric into Tungusic given that Proto-Amuric **lamos* is derived from the simplex root **la* ‘wind’, while the Tungusic parallel is not segmentable and because the initial liquid phoneme **l-* is atypical for Proto-Tungusic.

Korovina (2011) further finds that fish species that inhabit the Pacific Ocean are not well distributed across the Tungusic languages and are often borrowed from Nivkh into one or more individual daughter languages (e.g. Oroqen *lokko* ‘flounder’ from Nivkh *lok* ‘flounder’, Oroqen *la:kka* ‘herring’ and Orok *la:qqa* ‘herring’ from Nivkh *laku* ‘herring’, etc.). This suggests that the speakers of Proto-Tungusic were farmers, who did not acquire maritime vocabulary until they came in contact with indigenous populations on the Pacific coast, some of which might have spoken an ancestral form of Nivkh.

Except for these maritime loanwords, Proto-Tungusic borrowed only few words from Proto-Amuric. By contrast, there are several indications of Proto-Amuric substratum interference in Proto-Tungusic. The evidence comes from atypical structural features in Tungusic that are likely to have developed through imperfect learning from Proto-Amuric. Among others, these features include the development of a word-initial liquid and velar nasal sound in Tungusic, the development of a distinction between ‘we (including the addressee)’ and ‘we (excluding the addressee)’ in first-person plural pronouns, the development of a distinction between alienable and inalienable possession and the development of marking possessive relations on the head noun instead of the dependent (Robbeets 2017b).

Whether one prefers to explain the numerous structural similarities between Tungusic and other Transeurasian languages by borrowing or inheritance, it is commonly agreed that Tungusic typology is of the Transeurasian (or “Altaic”) type (Robbeets 2017a). Transeurasian languages are typical dependent-marking languages, while Nivkh is – similar to Ainu, Asian North Pacific-Coast languages (e.g., Chukotko-Kamchatkan), wider Paleosiberian languages (e.g.,) and languages of the Northwest Pacific Coast (e.g., Salishan, Wakashan, Chimakuan, Athabaskan) – of the head-marking type. The features above are atypical for Tungusic and more proto-typical of Nivkh in the sense that they represent direct or indirect implicational tendencies of being of the head-marking type or that they more frequently occur in Ainu, other Asian North Pacific-Coast languages,

wider Paleosiberian languages and languages of the Northwest Pacific Coast than in languages of the Transeurasian type.

These linguistic observations thus suggest a situation of language shift whereby some ancestral speakers of Proto-Amuric abandoned their own language and adopted the Proto-Tungusic target language.

5.2 Archaeology

Is our association of Proto-Tungusic with incoming millet farmers who imposed their language on local fishers speaking Proto-Amuric supported by the archaeological record?

During the Middle to Late Hongshan periods (4000–3000 BC), the cultivation of broomcorn and foxtail millet dispersed from the West Liao River basin in North East China to the Primorye (Maritime) province of the Russian Far East (Sergusheva & Vostretsov 2009; Leipe et al. 2019; Li et al. 2020). The introduction of millet farming in the Primorye was combined with the adoption of Northeast Chinese material culture such as cord-marked pottery, spindle whorls and stone agricultural tools, especially mortars, pestles and constricted-waist hoes (Nelson et al. 2020; Li et al. 2020) and led to the establishment of the Zaisanovskaya culture (3200–1300 BCE). The linguistically inferred time depth in the beginning of the first millennium AD corresponds to the break-up time when Proto-Tungusic separated into its primary branches and thus ceased to exist, but it does not inform us about when the ancestral language arrived in the region or started to exist there. Considering Bayesian inference of the time depth of the split between Tungusic and Mongolo-Turkic at 3300 BC (Robbeets & Bouckaert 2018), it is inviting to associate the arrival of Proto-Tungusic in the Russian Far East with the beginning of the Zaisanovskaya culture (Mallory et al. 2019; Wang & Robbeets 2020; Cui et al. 2020; Li et al. 2020).

Observing the high productivity of rice vis-à-vis millets, archaeobotanists argue that rice tends to be spread more easily through cultural diffusion, while millets are more frequently spread by population migration (Fuller & Qin 2009; Stevens & Fuller 2017). This is explained by the fact that wet rice cultivation can absorb population increase through intensification of land use, while the increased production of millet tends to occur through the agricultural colonisation of new land. The assumption of actual population movements from the West Liao River Basin to the Primorye in the fourth millennium BC is further supported by increases in population density (Peterson et al. 2010; Miyamoto 2014; Drennan et al. 2017; Leipe et al. 2019).

Population migration and cultural diffusion are expected to yield different linguistic outcomes (Thomason & Kaufman 1988). In the first case, when human populations move into new areas along with their language and culture, language shift is frequently observed: local speakers abandon their own language in favour of the incoming target language. Due to imperfect learning, the abandoned language may leave some traces in the structure of the target language, a phenomenon called “substratum interference” (Van Coetsem 2000; Johanson 2002; Winford 2013). Nevertheless, the newly adopted language is genealogically related to the ancestral language of the migrants. By contrast, in the case of cultural diffusion, when certain elements of language and culture move into new areas without the intervention of a migrating population, local speakers frequently maintain their own language but borrow certain words from the model language. The assumption of language shift, whereby a part of the Proto-Amuric speakers abandoned their native language and shifted to the Proto-Tungusic target language is thus in line with a scenario of population migration. Therefore, the archaeological and linguistic observations converge in suggesting that the spread of the Proto-Tungusic farmers was driven by population migration.

At the end of the third century AD, there was a sharp cooling of the climate, which led to a worsening of the conditions for agriculture. This provided the impetus for a gradual migration of millet farmers to coastal regions across most of the Primorye. Based on archaeobotanical data (Yanušević et al. 1990), it appears that the coastal groups ceased to cultivate millets and wheat and returned to a subsistence strategy of hunting and fishing. If this event can be associated with the separation between Manchuric and Tungusic languages, as suggested by 31% of trees in the Bayesian phylogenetic analysis in Oskolskaya et al. (2022), it would explain why southern Tungusic populations on the Lower Amur such as the Nanai, Oroch and Udehe people were traditionally predominantly fishers and gatherers, rather than farmers.

Hudson (2020) proposed further details of later northern Tungusic expansions. The Evenki are widely distributed hunter-gatherers who also herd domesticated reindeer. According to Anderson (1999: 142) and Zgusta (2015: 166), they first herded wild reindeer around Lake Baikal then moved north ca. AD 1000, reaching the Arctic ocean by the 17th century. The Even probably separated from the Evenki in medieval times (Pakendorf 2007: 15–16), matching the separation estimated at 556 years ago in our Bayesian analysis (Figure 4). They further expanded with reindeer from 17th century onwards, mirroring the separation between Even and Negidal estimated at 393 years ago. Probably due to Russian colonial expansion, the Oroqen and Solon moved south from the Amur in the 17th century, mirroring the estimate of 405 years ago.

5.3 Genetics

The first applications of genetics to the study of human prehistory involved mitochondrial and Y-chromosomal DNA. Whereas mitochondrial DNA is passed down along the maternal line from mother to daughter to granddaughter (and from mother to son but not passed on from sons to their offspring), Y-chromosomal DNA goes along the paternal line from father to son to grandson. Sequencing the chemical building blocks of uniparental DNA from diverse people around the world and comparing the mutations across these sequences, geneticists can reconstruct family trees of maternal and paternal relationship. However, since mitochondrial DNA and Y-chromosomal DNA represent only a tiny proportion of the human genome and provide information on only one out of very numerous ancestors, they shed light on only a limited slice of human prehistory. In fact, our entire genome contains information about many diverse ancestors, not just the two whose lineages can be traced with mitochondrial and Y-chromosomal DNA. The recently acquired ability to sequence the whole genome – meaning, the entire genome analyzed at once instead of just small stretches of it such as mitochondrial and Y-chromosomal DNA – has given us access to richer information recorded into all 23 chromosomes of our genome and representing a multitude of ancestors. Whole genome analysis means a revolution in the study of the human past because it allows us to go beyond the tiny slice of the past sampled by our mtDNA and Y-chromosomal DNA. As recent genome-wide analyses of Tungusic speakers (Pugach et al. 2016; Siska et al. 2017; Wang et al. 2021; Wang & Robbeets 2020) are expected to tell a richer story than previous studies about their mtDNA (Starikovskaya et al. 2005; Sukernik et al. 2012; Duggan et al. 2013) and Y-chromosomal DNA (Malyarchuk et al. 2010; Duggan et al. 2013), we here focus our report on genome-wide analyses.

The Principal Component Analysis in Figure 7 visualizes the genetic distance between contemporary speakers of Tungusic languages and other present-day East Asian populations. In addition, it plots ancient genomes from the Devil's cave in the Southern Primorye dating back to the fifth and sixth millennium BC and from the Ust'-Ilda site near Lake Baikal dating back to the fourth and third millennium BC onto the contemporary Tungusic-speaking populations.

The contemporary Tungusic speakers in the Amur River Basin, such as the Hezhen, Nanai, Negidal, Oroqen and Ulch are genetically most similar to ancient genomes from the Southern Primorye dating back to the fifth and sixth millennium BC (Siska et al. 2017; Wang et al. 2021; Wang & Robbeets 2020). This is also true for the Nivkh people on Sakhalin island, even if their language is not of a Tungusic descent. The Xibe people in Xinjiang are shifted towards Han Chi-

nese populations due to Chinese influence but they are still very similar to the Amur Tungusic populations and close to the Devil’s Gate genome. Whereas some Eastern Evenki are similar to the Amur Tungusic populations, Baikal Evenki and Even populations are shifted towards West Eurasians, such as the Uyghur Turkic populations on the PCA. Wang & Robbeets (2020) estimate that they have about 14% to 35% West Eurasian related ancestry, but that their admixture is a very recent event, going back less than 200 years in time. The ancient genomes from the Ust’-Ilda site near Lake Baikal dating back to the fourth and third millennium BC show that they derive a large amount of Devil’s Gate related Amur-like ancestry and also have some admixture from West Eurasians. Their genetic profile is similar to Even people.

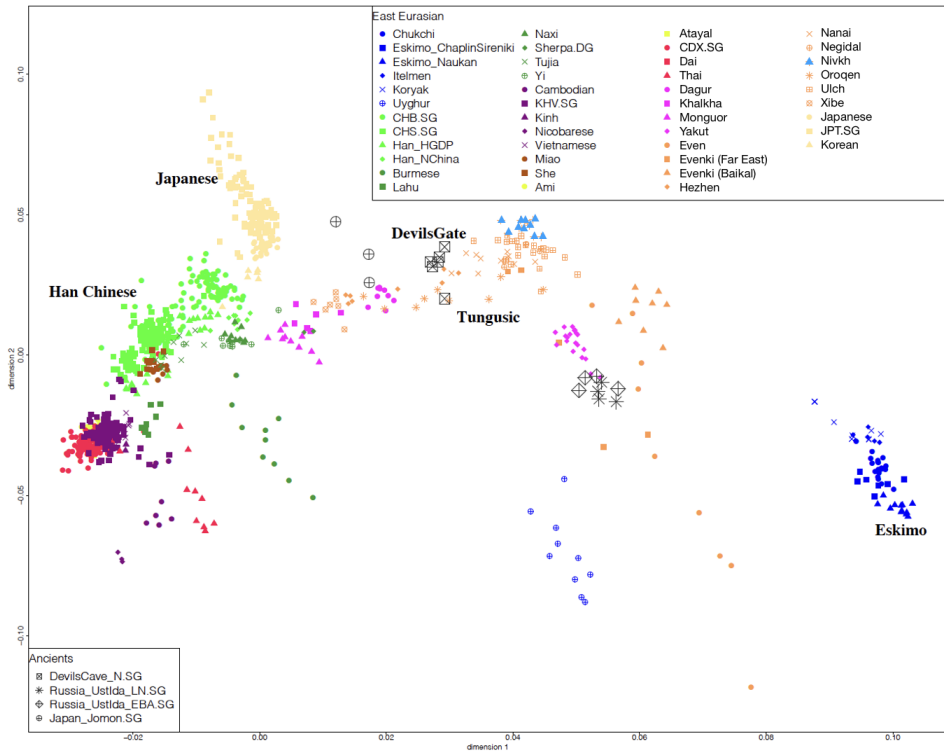


Figure 7: Principal Component Analysis of East Asian populations, projecting ancient Devil’s Gate and Ust’-Ilda genomes onto the present-day speakers of Tungusic languages (adapted from Wang & Robbeets 2020)

This genome-wide perspective is corroborated by analyses of mitochondrial and Y-chromosomal DNA. In the maternal line, there are only faint traces of a

genetic relationship between Tungusic-speaking populations in the Amur region, such as Negidal, Ulch and Udeghe and northern Tungusic populations, such as Even and Evenki, due to drift and admixture (Duggan et al. 2013). Nevertheless, the shared haplotypes found in these populations might be retentions from an earlier shared ancestral Tungusic population. Mitochondrial haplogroup frequencies show a cluster of Tungusic-speaking populations in the Amur region with Nivkh populations. The clustering of Even speakers with speakers of Yukaghir is seen as an implication of recent northward expansions of northern Tungusic speakers (Sukernik et al. 2012).

Tungusic speakers are further associated with the Y chromosomal haplogroup C3-M217, which is prevalent in Evenki and Even, as well as in other Tungusic speaking populations in the Amur River Basin including Oroqen, Ulch, Negidal, Udehe and Nanai (Malyarchuk et al. 2010; Duggan et al. 2013). This haplogroup is further well distributed among contemporary Mongolic and Nivkh-speaking populations and has been recovered in human remains of the Boisman culture (4825–2470 BC) in the Russian Far East (Yan et al. 2014; Wang et al. 2021).

It thus appears that both northern and southern Tungusic speaking populations share a proportion of their ancestry, which we refer to as the “Amur” genome. Mongolic and Turkic-speaking populations share a part of this “Amur” ancestry in spite of their increasing admixture with people of Western Eurasian ancestry from the first millennium BC onwards. (Jeong et al. 2018, 2019). Combined with recent analyses of ancient genomes from the West Liao River Basin (Ning et al. 2020), these results suggest that the Amur gene pool has long occupied the region from the Baikal to the West Liao River to the Russian Far East, at least for the last 10 000 years.

Since the Nivkh and the Tungusic-speaking populations share the same Amur ancestry, there are no traces of genetic admixture indicating population migration at the time of the agricultural dispersals. The long-term genetic continuity in the Amur basin is commonly used to argue against population migration and to support demic diffusion of agriculture into the Amur area (Siska et al. 2017). However, this conclusion does not take into account the increase in population density at the time of the agricultural expansions discussed in §5.2, which supports an alternative possibility that the incoming farmers may have shared an Amur-like genetic profile with the local populations (Cui et al. 2020; Wang & Robbeets 2020; Jeong et al. 2020; Li et al. 2020). Given the vast geographical reach of the Amur genetic profile, including the West Liao River region as well as the Russian Far East, a genetic admixture between Proto-Tungusic incoming farmers and Proto-Amuric local fishers would have led to an admixture of two

similar Amur genomes, like mixing two white paints together. Therefore, population migration and admixture are not expected to be visible in the genome. Bringing the archaeological, linguistic and genetic evidence together thus leaves room for agriculture-driven population migration and language shift spreading Proto-Tungusic to the Russian Far East in the Neolithic.

6 Conclusion

Quantitative methods, such as the Bayesian approach adopted in Oskolskaya et al. (2022), have much to offer: they can infer an internal family structure, calculate the statistical robustness of the proposed branches, estimate an absolute time depth within credible intervals without assuming a constant rate of change and help us to determine the location of the original homeland. Nevertheless, Bayesian results should be interpreted with caution, as they are dependent on the quality of the data input and the plausibility of the calibrations. Besides, they are limited in their abilities because even if they can infer information about the time and space of linguistic dispersals, they do not inform us about the natural and cultural environment of the ancient speakers: they can tell us where and when ancestral speech communities were located, but not why these people moved. In order to provide a better understanding of causalities in linguistic prehistory, we need to reinstate comparative historical linguistic tradition and, together with archaeology and genetics, integrate it into a holistic approach. In this paper, we attempted to take such an approach for the dispersal of Proto-Tungusic in time and space.

Combining the power of traditional comparative historical linguistics and computational phylogenetics, we used the recent Bayesian analysis provided by Oskolskaya et al. (2022) to quantify the likelihood of previously proposed classifications. We found that two classifications, namely the revised North-South classification (Figure 2b) and the Manchu-Tungusic classification (Figure 2b) were statistically robust, while other proposals could be excluded. Since we expressed the disagreement among different authors with regard to the exact configuration of the Tungusic tree in terms of probability, we were able to provide a quantitative basis to the ongoing discussions.

Chronologically, we estimated the break-up of Proto-Tungusic in the beginning of the first millennium AD and situated the homeland geographically in the area around or to the north of Lake Khanka.

Triangulating the linguistic evidence for Proto-Tungusic with evidence from archaeology and genetics, we argued for a language shift around 3300 BC, whereby some ancestral speakers of Proto-Amuric in the Russian Far East abandoned

their own language and adopted the Proto-Tungusic target language. The dispersal of Proto-Tungusic from the Liao River basin to the area around lake Khanka was probably caused by the expansion of millet agriculture and driven by population migration. The separation of the Manchuric branch, which may represent the first split in the family, can be associated with a return to hunting and fishing, in part because the conditions for agriculture worsened through climate change.

Abbreviations

PTg Proto-Tungusic

Acknowledgements

The research leading to these results has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 646612) granted to Martine Robbeets. We thank Patryk Czerwinski, Elena Klyachko and Matthew Miller for their help in collecting lexical data from the field and Ezequiel Koile for his advice on Bayesian methodology.

Appendix A Reconstruction of basic sound inventories for Proto-Tungusic

Table 3: Reconstruction of the basic consonant inventory of Proto-Tungusic (Robbeets 2020, supplementary files)

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*p-	f-	f-	f-	x-	x-	p-	f- x-	p-	p-	h-	ø-	h-	x-
*-p-	f	f b ø	v ø	p w ø	p f w	p	f	p	p	b w ø	p w g ø	p b w ø	p w
*b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-	b-
*-b-	b w ø	b f w ø	v ø	b w ø	b w ø	b w ø	w ø	b w ø	b w ø	b w ø	b p w ø	w ø	w ø
*t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-	t-

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*-t-	t	t	t s	t	t	t	t	t	t	t	t	t	t
*d-	d-	d-	d-	d-	d-	d-	d-	d-	d-	d-	d-	d-	d-
*ji-	ɟi-	ɟi-	ɟi-			ɟi-	ɟi-	ɟi-					
*-d-	d	d	d	d	d	d	d	d	d	d	d	d	d
*-ji-	ɟi	ɟi	ɟi			ɟi	ɟi	ɟi					
*k-	x-	x-	x-	k-	k-	k-	k-	k-	k-	k-	x-	k-	k- ø-
				ø-	ø-	ø-	ø-	ø-	ø-	ø-	ø-	ø-	
*-k-	k x	k x	k γ	k ø	k x	k γ	k γ	k ø	k γ	k	k x	k x	k x
	ø				γ ø	ø	ø		ø				
*g-	g-	g-	g-	g-	g-	g-	g-	g-	g-	g-	g-	g-	g- ŋ
				ŋ-	ŋ-	ø-	ø-	ŋ-	ŋ-	ŋ-	n-	ŋ-	
*-g-	γ w	γ w	ø	γ w	γ w	γ w	y ø	γ w	γ w	γ y	γ ø	γ	γ y
	ø	y ø		y ø	y ø	y ø		y ø	y ø				w
*č-	č-	č-	č-	č-	č-	č-	č-	č- >	č- >	č-	s-	č-	č-
								t-	t-				
*-č-	č	č	č	č	s	č	č s	č >	č	č	š	č	č
								t					
*x-	w-	w-	v-	x-	w-	x-	x-	x-	x-	ø-	ø-	ø-	ø-
	ø-	ø-	ø-	ø-	ø-	s-	s-	s-	s-				
*-x-	x	x	x k	k	ø	x ø	x k	x ø	x ø	k	x	k	k x
			γ										
*s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-	s-
*-s-	s	s	s	s	s h	s	s	s	s	s	s	x	s
					ø								
*m-	m-	m-	m-	m-	m-	m-	m-	m-	m-	m-	m-	m-	m-
						ŋ-			ŋ-				
*-m-	m	m	m	m	m	m	m	m	m	m	m	m	m
*n-	n-	n-	n-	n-	n-	n-	n-	n-	n-	n-	n-	n-	n-
						l-	l-	l-	l-			l-	
*-n-	n	n	n	n	n	n	n	n	n	n	n	n	n
*-r-	r	r	r	y ø	y ø	r	r	r	r	r	r	r	y ø
*-l-	l	l	l	l	l	l	l n	l	l	l	l	l	l
*-y-	y	y	y	y	y	y	y	y	y	y	y	y	y

Table 4: Reconstruction of the basic vowel inventory of Proto-Tungusic (Robbeets 2020, supplementary files)

PTg	J.	Ma.	X.	Or.	Ud.	N	B.N.	Ork.	Ul	Evn	Sol.	Evk.	Neg.
*a	a	a	a	a	a	a	a	a	a	a	a	a	a
*e	e	e	e	e	e	e	e	e	e	e	e	e	e
*o	o	o	o	o	o	o	o	o	o	o	o	o	o
*ö	u	u	u	ou	o	u	u	ou	ou	o	u	u	u
*u	u	u	u	u	u	u	u	u	u	uõ	uõ	uõ	uõ
*ü	uei	uei	ui	iu	i y ø	uo	io	iu	i u o	i	i	iu	io
*i	i	i	i	i	i	iᶥ	iᶥ	i	iᶥ	i	i	iᶥ	iᶥ
*i(y)a	ie i _a	iya ai i _a	ia a	iæ ei	iæ æ a	i _a ea	iæ	ᶥ:	ᶥ:	ia	ᶥ:	ᶥ:	ᶥ:

References

- Alonso de la Fuente, José A. 2011. *Tense, voice and Aktionsart in Tungusic. Another case of analysis to synthesis?* Wiesbaden: Harrassowitz.
- Anderson, Gregory D. S. 1999. The Evenki of the lower Enisei valley. In Richard B. Lee & Richard Daly (eds.), *The Cambridge encyclopedia of hunters and gatherers*, 142–146. Cambridge: Cambridge University Press.
- Benzing, Johannes. 1956. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik* (Akademie der Wissenschaften und der Literatur. Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 11). Mainz: Akademie der Wissenschaften und der Literatur.
- Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut & Alexei J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10(4). e1003537.
- Cincius, Vera I. 1949. *Sravnitel'naja fonetika tunguso-man'čžurskich jazykov [Comparative phonetics of Manchu-Tungus languages]*. Leningrad: Učpedgiz.
- Cincius, Vera I. 1975/77. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov [Comparative dictionary of the Tungus-Manchu languages]*. 2 vols. Leningrad: Nauka.

- Cui, Yinqiu, Fan Zhang, Pengcheng Ma, Linyuan Fan, Chao Ning, Quanchao Zhang, Wei Zhang, Lixin Wang & Martine Robbeets. 2020. Bioarchaeological perspective on the expansion of the Transeurasian languages in the Amur River Basin. *Evolutionary Human Sciences* 2. e15.
- Derevyanko, Anatoliy P. 1976. *Priamur'ye v drevnosti (do nachala nashej ery) [Priamur'e in antiquity before the beginning of our era]*. Novosibirsk: Nauka.
- Doerfer, Gerhard. 1978. Classification problems of Tungus. In Gerhard Doerfer & Michael Weiers (eds.), *Tungusica, vol.1: Beiträge zur nordasiatischen Kulturgeschichte*, 1–26. Wiesbaden: Harrassowitz.
- Drennan, Robert D., Xueming Lu & Christian E. Peterson. 2017. A place of pilgrimage? Niuheiliang and its role in Hongshan society. *Antiquity* 91. 43–56.
- Duggan, Ana T., Mark Whitten, Victor Wiebe, Michael Crawford, Anne Butthof, Victor Spitsyn, Sergey Makarov, Innokentiy Novgorodov, Vladimir Osakovskiy & Brigitte Pakendorf. 2013. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One* 8(12). e83570.
- Dybo, Anna V. & Evgeniya Korovina. 2019. *Altaic linguistics and mythology: geography or relationship?* Paper presented at the conference “Centenary of human population genetics”, Biology Department of Moscow State University May 29–June 1.
- Fuller, Dorian Q. & Ling Qin. 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeology* 41. 88–111.
- Georg, Stefan. 2004. Unreclassifying Tungusic. In Carsten Naehrer (ed.), *Proceedings of the First International Conference on Manchu-Tungus studies*, vol. 2 (Tungusica Sibirica 9), 45–57. Wiesbaden: Harrassowitz.
- Gorelova, Liliya M. 2002. *Manchu grammar*. Leiden: Brill.
- Helimski, Eugene. 1985. Samodiisko-tungusskie leksicheskie svyazi i ikh etnoistoricheskie implikatsii [Samoyedic-Tungusic lexical correspondences and their ethno-historical implications]. In Elizaveta I. Ubriatova (ed.), *Uraloaltaistika*, 206–212. Novosibirsk: Nauka.
- Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List & Dmitry Egorov. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52(6). 841–875.
- Hudson, Mark J. 2020. Language dispersals and the “Secondary Peoples’ Revolution”: A historical anthropology of the Transeurasian unity. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 806–813. Oxford: Oxford University Press.

- Ikegami, Jirō. 1974. Versuch einer Klassifikation der tungusischen Sprachen. In Gyorgy Hazai & Peter Zieme (eds.), *Sprache, Geschichte und Kultur der altaischen Völker* (Schriften zur Geschichte und Kultur des Alten Orients 5), 271–272. Berlin: Akademie Verlag.
- Janhunen, Juha. 1996. *Manchuria. An ethnic history* (Mémoires de la Société Finno-Ougrienne 222). Helsinki: Suomalais-Ugrilainen Seura.
- Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 5–16. Wiesbaden: Harrassowitz.
- Janhunen, Juha. 2016. Reconstructio externa linguae Ghiliacorum. *Studia Orientalia* 117. 3–27.
- Jeong, Choongwon, Oleg Balanovsky, Elena Lukianova, Nurzhibek Kahbatkyzy, Pavel Flegontov, Valery Zaporozhchenko, Alexander Immel, Chuan-Chao Wang, Olzhas Ixan, Elmira Khussainova, Bakhytzhhan Bekmanov, Victor Zhibert, Maria Lavryashina, Elvira Pocheshkhova, Yuldash Yusupov, Anastasiya Agdzhoyan, Sergey Koshel, Andrei Bukin, Pagbajabyn Nymadawa, Shahlo Turdikulova, Dilbar Dalimova, Mikhail Churnosov, Roza Skhalyakho, Denis Daragan, Yuri Bogunov, Anna Bogunova, Alexandr Shtrunov, Nadezhda Dubova, Maxat Zhabagin, Levon Yepiskoposyan, Vladimir Churakov, Nikolay Pislegin, Larissa Damba, Ludmila Saroyants, Khadizhat Dibirowa, Lubov Atramentova, Olga Utevska, Eldar Idrisov, Evgeniya Kamenshchikova, Irina Evseeva, Mait Metspalu, Alan K. Outram, Martine Robbeets, Leyla Djansugurova, Elena Balanovska, Stephan Schiffels, Wolfgang Haak, David Reich & Johannes Krause. 2019. The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution* 3. 966–976.
- Jeong, Choongwon, Chao Ning & Chuan-Chao Wang. 2020. The Oxford guide to the Transeurasian languages. In Martine Robbeets & Alexander Savelyev (eds.), *Transeurasian unity from a population genetic perspective*, 784–793. Oxford: Oxford University Press.
- Jeong, Choongwon, Shevan Wilkin, Tsend Amgalantugs, Abigail S. Bouwman, William Taylor, Richard W. Hagan, Sabri Bromage, Soninkhishig Tsolmon, Christian Trachsel, Jonas Grossmann, Judith Littleton, Cheryl A. Makarewicz, John Krigbaum, Marta Burri, Ashley Scott, Ganmaa Davaasambuu, Joshua Wright, Franziska Irmer, Erdene Myagmar, Nicole Boivin, Martine Robbeets, Frank J. Rühli, Johannes Krause, Bruno Frohlich, Jessica Hendy & Christina Warinner. 2018. Bronze age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proceedings of the National Academy of Sciences* 115(48). E11248–E11255.

- Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In Mari C. Jones & Edith Esch (eds.), *Language change. The interplay of internal, external and extra-linguistic factors*, 285–313. Berlin: Mouton de Gruyter.
- Kormušin, Igor V. 1998. *Udyhejskij jazyk [The Udihe language]*. Moscow: Nauka.
- Korovina, Evgeniya. 2011. *Leksika prirodnogo i kul'turnogo okruženiya v Tunguso-Man'čžurskich jazykach (v istoriko-tipologičeskom osveščanii) [Tungus-Manchu vocabulary related to natural environment and cultural activities (in historical and typological perspective)]*. Moscow: Russian State University of Humanities. BA dissertation.
- Leipe, Christian, T. Long, Elena A. Sergusheva, Maike Wagner & P.E. Tarasov. 2019. Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Science Advances* 5. eaax6225.
- Li, Tao, Martine Robbeets, Chao Ning, Irina Zhushchikhovskaya & Mark Hudson. 2020. Millet agriculture dispersed from Northeast China to the Russian Far East: Integrating archaeology, genetics, and linguistics. *Archaeological Research in Asia* 22. 100177.
- Mallory, James, Anna Dybo & Oleg Balanovsky. 2019. The impact of genetics research on archaeology and linguistics in Eurasia. *Russian Journal of Genetics* 55. 1472–1487.
- Malyarchuk, Boris, Miroslava Derenko, Galina Denisova, Marcin Wozniak, Tomasz Grzybowski, Irina Dambueva & Ilia Zakharov. 2010. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Annals of Human Genetics* 74. 539–546.
- Menges, Karl Heinrich. 1968. Tungusologie. In Walter Fuchs (ed.), *Die tungusischen Sprachen* (Handbuch der Orientalistik I.5,3), 21–256. Leiden: Brill.
- Miyamoto, Kazuo. 2014. The initial spread of early agriculture into Northeast Asia. *Asian Archaeology* 3. 1–12.
- Nelson, Sarah, Irina Zhushchikhovskaya, Tao Li, Chao Ning, Mark Hudson & Martine Robbeets. 2020. Tracing population movements in ancient East Asia through the linguistics and archaeology of textile production. *Evolutionary Human Sciences* 2. e5.
- Ning, Chao, Tianjiao Li, Ke Wang, Tao Li, Fan Zhang, Xiyan Wu, Shizhu Gao, Quanchao Zhang, Hai Zhang, Mark J. Hudson, Sihao Wu, Yanming Fang, Chen Liu, Chunyan Feng, Wei Li, Tao Han, Ruo Li, Jian Wei, Yonggang Zhu, Yawei Zhou, Chuan-Chao Wang, Shengying Fan, Zenglong Xiong, Guanghui Dong, Zhouyong Sun, Maolin Ye, Lei Sun, Xiaohong Wu, Fawei Liang, Yanpeng Cao, Xingtao Wei, Hong Zhu, Hui Zhou, Johannes Krause, Martine Robbeets, Choongwon Jeong & Yinqi Cui. 2020. Ancient genomes from northern China

- suggest links between subsistence changes and human migration. *Nature Communications* 11(1). 1–9.
- Oskolskaya, Sofia, Ezequiel Koile & Martine Robbeets. 2022. A Bayesian approach to the classification of Tungusic languages. *Diachronica* 39(1). 128–158.
- Pakendorf, Brigitte. 2007. *Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives*. Utrecht: LOT.
- Perekhvalskaya, Elena. 2022. From consonant to tone: Laryngealized and pharyngealized vowels in Udihe. In Andreas Hölzl & Thomas E. Payne (eds.), *Tungusic languages: Past and present*, 227–262. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053371.
- Peterson, Christian E., Xueming Lu, Robert D. Drennen & Da Zhu. 2010. Hongshan chiefly communities in Neolithic northeastern China. *Proceedings of the National Academy of Sciences USA* 107. 5756–5761.
- Pevnov, Aleksandr M. 2012. The problem of the localization of the Manchu-Tungusic homeland. In Andrej Malchukov & Lindsay Whaley (eds.), *Recent advances in Tungusic linguistics* (Turcologica 89), 17–40. Wiesbaden: Harrassowitz.
- Pugach, Irina, Rostislav Matveev, Viktor Spitsyn, Sergey Makarov, Innokentiy Novgorodov, Vladimir Osakovsky, Mark Stoneking & Brigitte Pakendorf. 2016. The complex admixture history and recent southern origins of Siberian populations. *Molecular Biology and Evolution* 33(7). 1777–1795.
- Robbeets, Martine. 2015. *Diachrony of verb morphology. Japanese and the Transeurasian languages* (Trends in linguistics 291). Berlin: De Gruyter Mouton.
- Robbeets, Martine. 2017a. Japanese, Korean and the Transeurasian languages. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics* (Cambridge Handbooks in Language and Linguistics), 586–626. Cambridge: Cambridge University Press.
- Robbeets, Martine. 2017b. Proto-Trans-Eurasian: Where and when? *Man in India* 97. 19–46.
- Robbeets, Martine. 2020. Basic vocabulary in the Transeurasian languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 645–659. Oxford: Oxford University Press.
- Robbeets, Martine & Remco Bouckaert. 2018. Bayesian phylolinguistics reveals the internal structure of the Transeurasian family. *Journal of Linguistic Evolution* 3. 145–162.
- Robbeets, Martine, Juha Janhunen, Alexander Savelyev & Evgeniya Korovina. 2020. The homelands of the individual proto-languages: Where, what and

- when? In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 753–771. Oxford: Oxford University Press.
- Robbeets, Martine & Alexander Savelyev (eds.). 2020. *The Oxford guide to the Transeurasian languages*. Oxford: Oxford University Press.
- Sergusheva, Elena A. & Yury E. Vostretsov. 2009. The advance of agriculture in the coastal zone of East Asia. In Andrew Fairbairn & Ehud Weiss (eds.), *From foragers to farmers: Papers in honour of Gordon C. Hillman*, 205–219. Oxford: Oxbow books.
- Siska, Veronika, Eppie R. Jones, Jeon Sungwon, Bhak Youngjune, Kim Hak-Min, Cho Yun Sung, Kim Hyunho, Lee Kyusang, Elizaveta Veselovskaya, Tatiana Balueva, Marcos Gallego-Llorente, Michael Hofreiter, Daniel G. Bradley, Anders Eriksson, Ron Pinhasi, Bhak Jong & Andrea Manica. 2017. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Science Advances* 3. 1–10.
- Starikovskaya, Elena B., Rem I. Sukernik, Olga A. Derbeneva, Natalia V. Volodko, Eduardo Ruiz-Pesini, Antonio Torroni, Michael D. Brown, Marie T. Lott, Seyed H. Hosseini, Kirsi Huoponen & Douglas C. Wallace. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of Human Genetics* 69(1). 67–89.
- Starostin, Sergei, Anna Dybo & Oleg Mudrak. 2003. *Etymological dictionary of the Altaic languages*. Leiden: Brill.
- Stevens, Chris J. & Dorian Q. Fuller. 2017. The spread of agriculture in eastern Asia: Archaeological bases for hypothetical farmer/language dispersals. *Language Dynamics and Change* 7. 152–186.
- Sukernik, Rem I., Natalia V. Volodko, Ilya O. Mazunin, Nikolai P. Eltsov, Stanislav V. Dryomov & Elena B. Starikovskaya. 2012. Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: Contribution to prehistory of native Siberians and their affinities to Native Americans. *American Journal of Physical Anthropology* 148. 123–138.
- Sunik, Orest P. 1959. *Tunguso-man'čžurskije jazyki [Manchu-Tungus languages]* (Mladopi'mennyje jazyki narodov SSSR). Moscow: Nauka.
- Thomason, Sarah G. & Terence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tugolukov, V. A. 1980. Etniceskie korni tungusov [Ethnic roots of the Tungus people]. In I. S. Gurvich (ed.), *Etnogenez narodov severa [Ethnogenesis of the peoples of the North]*, 152–176. Moscow: Nauka.
- Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag C. Winter.

- Vasilevič, Glafira M. 1960. *K voprosu o klassifikacii tunguso-man'čžurskih jazykov* [About the question of classification of Manchu-Tungus languages] (Voprosy jazykoznanija 2). Moscow: Nauka.
- Vasilevič, Glafira M. 1969. *Èvenki: Istoriko-etnografičeskie očerki (XVIII – načalo XX v.)* Leningrad: Nauka.
- Vovin, Alexander. 1993. Towards a new classification of Tungusic languages. *Eurasian Studies Yearbook* 65. 99–113.
- Wang, Chuan-Chao & Martine Robbeets. 2020. The homeland of Proto-Tungusic inferred from contemporary words and ancient genomes. *Evolutionary Human Sciences* 2. E8.
- Wang, Chuan-Chao, Hui-Yuan Yeh, Alexander N Popov, Hu-Qin Zhang, Hirofumi Matsumura, Kendra Sirak, Olivia Cheronet, Alexey Kovalev, Nadin Rohland, Alexander M. Kim, Rebecca Bernardos, Dashtseveg Tumen, Jing Zhao, Yi-Chang Liu, Jiun-Yu Liu, Matthew Mah, Swapan Mallick, Ke Wang, Zhao Zhang, Nicole Adamski, Nasreen Broomandkhoshbacht, Kimberly Callan, Brendan J. Culleton, Laurie Eccles, Ann Marie Lawson, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Shaoqing Wen, Shi Yan, Fatma Zalzal, Richard Chuang, Ching-Jung Huang, Chung-Ching Shiung, Yuri G. Nikitin, Andrei V. Tabarev, Alexey A. Tishkin, Song Lin, Zhou-Yong Sun, Xiao-Ming Wu, Tie-Lin Yang, Xi Hu, Liang Chen, Hua Du, Jamsranjav Bayarsaikhan, Enkhbayar Mijidorj, Diimaajav Erdenebaatar, Tumur-Ochir Iderkhangai, Erdene Myagmar, Hideaki Kanzawa-Kiriyama, Msato Nishino, Ken-ichi Shinoda, Olga A. Shubina, Jianxin Guo, Qiongying Deng, Longli Kang, Dawei Li, Dongna Li, Rong Lin, Wangwei Cai, Rukesh Shrestha, Ling-Xiang Wang, Lanhai Wei, Guangmao Xie, Hongbing Yao, Manfei Zhang, Guanglin He, Xiaomin Yang, Rong Hu, Martine Robbeets, Stephan Schiffels, Douglas J. Kennett, Li Jin, Hui Li, Johannes Krause, Ron Pinhasi & David Reich. 2021. The genomic formation of human populations in East Asia. *Nature* 591. 413–419.
- Whaley, Lindsay, Lenore A. Grenoble & Fengxiang Li. 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75. 286–321.
- Whaley, Lindsay & Sofia Oskolskaya. 2020. The classification of the Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 81–91. Oxford: Oxford University Press.
- Wichmann, Søren, André Müller & Veveka Velupillai. 2010. Homelands of the world's language families. A quantitative approach. *Diachronica* 27(2). 247–276.

- Winford, Donald. 2013. On the unity of contact phenomena: The case for imposition. In Carole de Feral (ed.), *In and out of Africa: Languages in question. In honour of Robert Nicolai*, 43–71. Leuven: Peeters.
- Yan, Shi, Chuan-Chao Wang, Hong-Xiang Zheng, Wei Wang, Zhen-Dong Qin, Lan-Hai Wei, Yi Wang, Xue-Dong Pan, Wen-Qing Fu, Yun-Gang He, Li-Jun Xiong, Wen-Fei Jin, Shi-Lin Li, and Yu An, Hui Li & Li Jin. 2014. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLOS One* 9(8). e105691.
- Yanušević, Zoya, Yuri Vostretsov & Sofiya V. Makarova. 1990. *The palaeo-ethnobotanical finds in Primorye*. Vladivostok: Nauka.
- 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.

Chapter 9

Historical language contact between Sibe and Khorchin

Veronika Zikmundová

Charles University

The Sibe of Xinjiang have been recognized as speakers of a Manchu variety by linguists. However, for the Sibe speakers themselves, the situation is more complicated. For certain reasons, the Sibe often present themselves as a group whose historical origins are different from the Manchus. Several mentions occur in historical sources about Sibe being vassals to the Khorchin Mongols before “becoming Manchus”. This has been used among the arguments for the non-Manchu identity of the Sibe.

In recent years, academic discussion has focused on the ethnic identity of the Manchus, and, to a lesser extent, also on the position of the Sibe in relation to the Manchus. In this paper I try to select out features of possible Khorchin, i.e. eastern Mongolian, origin, in Sibe which may have come from direct language contact. I discuss several morphological features of Mongolic origin which seem not to be shared by other Manchu varieties, and one remarkable Sibe feature of Khorchin origin (the emphatic prefix *me-*). In addition, I mention the existence of lexical evidence of direct contact which is found in more conservative layers of Sibe vocabulary. Another question concerns the significance of this evidence for imagining the Sibe history. The linguistic situation in central Manchuria during the period concerned (15th–16th centuries) suggests that if the shared features indeed come from this period, they may rather be remnants of an extinct linguistic environment characterized by intense Mongolic-Tungusic contacts than of bilateral contact between two distinct groups – Khorchins and Sibe.



1 Overview

Central Manchuria has been the home of many Mongolic- and Jurchenic-speaking¹ communities and the site of multiple and multi-layer contacts between these groups for several centuries.² During the period of the Yuan and Ming rule, namely between the 14th and 16th centuries, many demographic shifts happened which were probably followed by important changes in the linguistic situation, such as the growth of Mongolic influence in the area. These shifts supposedly resulted in new, both massive and small-scale, Jurchenic-Mongolic language contacts (cf. Janhunen 1996: 97). Most of these contact events are little, if at all, documented. However, in 20th century China, one of these little documented events received particular attention and different interpretations. This was the historical fact of the (probably) Jurchenic-speaking Sibe being vassals of the Khorchin Mongols. The present article is concerned with this contact event, its contexts and interpretations.

Modern SIBE is a Jurchenic diaspora language which has often been classified as an oral variety of Manchu. It is related to the other oral Manchu varieties which have been discovered in Manchuria during the 20th century. Sibe is spoken by 10,000–20,000 individuals in several localities close to the north-western border of China, detached by some 4000 kilometers from their original homeland in Manchuria.³ KHORCHIN, an eastern (Manchurian) variety of Mongolian, currently has about a million speakers who inhabit a large area of eastern Inner Mongolia, Jilin and Heilongjiang.

In the 16th and early 17th century (before the Qing administrative re-organization of Manchuria), most Jurchenic-speaking communities were grouped into several Jurchen tribal confederacies.⁴ Historical sources relate that in the same

¹The term *Jurchenic* was coined by Janhunen (1996: 154) as a term comprising both the documented Jurchen varieties and other, undocumented southern Tungusic languages whose existence Janhunen thus suggests. It seems convenient to use this term to refer to the branch of Tungusic languages which includes the extinct Jurchen varieties and their successor languages – written Manchu and several spoken Manchu varieties. These have been known under the names of Alchuka, Bala, Lalin, Aihui, Sanjiazi, Yibuqi and Sibe. Another little documented language, the Manchu Kyakala, has recently been suggested as belonging to this branch (Hözl & Hözl 2019).

²Janhunen (1996: 96–110) describes the setting of Manchuria during the Ming and Qing rule with several case studies of migrations and contact events, which show the ethnic and linguistic complexity in the area and enable us to estimate analogous, insufficiently documented migrations and language contact events.

³For descriptions of spoken Sibe see, for example, Norman (1974), Jang (2008), Zikmundová (2013), Kogura (2018).

⁴For an overview of the pre-Qing organization of the Jurchen tribes see, for example, Janhunen (1996: 98–100).

period, the Sibe were subject to the Mongolic Khorchin tribe and only in the 1690s were united with the rest of Jurchenic speakers (see §3.1). Linguistically, modern Sibe and modern Khorchin share certain features which may have originated at the time of their mutual contact during the Ming dynasty.

This article is an attempt to examine these similarities in their socio-linguistic and historical contexts and suggest an interpretation of their significance for Sibe studies. Further, I take the narrative of the historical Sibe-Khorchin contact and the search for possible linguistic evidence about it as a starting point for an attempt to outline some important traits of the linguistic situation in Central Manchuria before the 18th century.

First, in §3, the historical context of the supposed Sibe-Khorchin language contact is summarized and the political and socio-linguistic background of modern Sibe historiography is mentioned. I suggest that the period of historically documented pre-Qing contacts between the Sibe and the Khorchins has been assigned particular importance in the argumentation for ethnic origins distinct from those of the Manchus. In §4, the actual parallels in phonetics and morphology are listed. These are based, for the most part, on fieldwork data. Here I only mention features which Sibe shares with Khorchin and which are either not attested, or are marginal, in the other documented Manchu varieties. §5 gives examples of Mongolic loanwords in Sibe which are not documented in the other Manchu varieties. Some of them are Mongolic in general while others belong exclusively to the cultural sphere of the Manchurian Mongols. In the concluding part I discuss what these shared features can tell us about the linguistic situation in pre-Qing central Manchuria.

I suppose that the selected features may have resulted from a direct Mongolic influence on Sibe which was more intensive than the general Mongolic influence to which other Manchu varieties were exposed. However, concerning further interpretations of these shared features, they can be attributed both to pre-Manchu contact with Khorchin and to later contact with other Mongolic languages – Daur, Jungarian Chakhar and Öölöd. Independent internal developments cannot be ruled out either. Most importantly, in the light of historical data, it seems more plausible to interpret the shared features as remnants of a generally more Mongolic-influenced Jurchenic milieu which was otherwise lost due to language standardization, than as a proof of the historical Sibe-Khorchin contact.

2 Methodology

In search for the Sibe-Khorchin analogies, mainly corpora of Sibe and Khorchin fieldwork data were used. The Sibe part of these data, collected by myself in

the Xinjiang Sibe communities mainly with the purpose of grammar description, comes from the period between 1993 and 2009. The Khorchin part⁵ was collected between 2004–2015 both by the local consultant Bai Xiaomei and by myself. The Khorchin data were not elicited with the purpose of grammar description and therefore do not cover the whole Khorchin grammar which leaves some room for as yet undiscovered shared grammatical features. Additionally, if not stated otherwise, I use Khalkha Mongolian and Sanjiazhi Manchu⁶ data from my own fieldwork collections.

Distinctions between Sibe and written Manchu have been described, above all, by Jang Taeho (2008). During my work on Sibe grammar description I tried to systematically note features which not only distinguished Sibe from written Manchu, but which seemed likely to be of Mongolic origin. I subsequently searched for these features in the materials of Khorchin on one hand, and in other spoken Manchu materials on the other. I selected those features which are shared with Khorchin and, at the same time, either not attested or – compared to Sibe – marginal in the oral Manchurian varieties of Manchu.

In order to draw a plausible interpretation of the selected shared features I attempted to systematize the available information about the linguistic history of the area concerned and align the historical mentions of pre-Qing Sibe and Khorchin with more general patterns of developments in Ming Central Manchuria. Further, it seemed to be important to assess the value of the official Sibe historiography and its accent on the non-Jurchen origins of the Sibe for the interpretation of the Sibe-Khorchin contact history. Fortunately, recently published works such as Zhuangsheng (2019) and Sárközi (2019) offer a much-needed insight into the motivation of the indigenous Sibe historiography.

3 The historical and socio-linguistic background of the Sibe-Khorchin language contact

Below I give basic data about the two languages involved in the supposed language contact episode, including some historical facts that pertain to the general linguistic situation in the area and time concerned. I also note the socio-historical contexts of the official self-presentation of the modern Sibe people as a group of non-Jurchen origin.

⁵The Khorchin data comprise approximately 10 hours of lengthy interviews on historical and cultural topics.

⁶The village of Sanjiazhi (Fuyu county) is one of the last locations in Heilongjiang where a form of Manchu is still spoken by several elderly individuals.

3.1 Sibe

At present, two groups of people in China at two different locations are officially recognized as members of the Sibe ethnic group. The larger of these groups inhabits certain areas in Northeastern China (Manchuria) and are speakers of Mandarin. The smaller group of Sibe⁷, some 30,000 individuals, live in the most faraway corner of China – the Ili valley on the border with Kazakhstan. These Sibe are not only more-or-less fluent speakers of a Manchu variety, but also preservers of a specific Manchurian culture. This paper is concerned with the latter – Xinjiang or Jungarian – Sibe⁸ group.

Comparative data from other living or recently extinct Manchu varieties (e.g. Wang 2005; Zhao 1989; Mu 1985, 1986a,b, 1987, 1988; Hölzl & Hölzl 2019) allow Sibe to be classified as one of the Bannermen Manchu⁹ varieties together with Sanjiazi Manchu, Aihui Manchu, Yibuqi Manchu and Lalin/Jing Manchu. Historically, these varieties, in contrast to other modern Jurchenic languages, seem to have been forms of a standard spoken language used in Manchu military garrisons. Knowledge of written Manchu, which was widespread in the Manchurian garrisons as well as in the Xinjiang Sibe enclaves, is probably responsible for the relatively little diversity among all Bannermen Manchu varieties. Most of the differences between Sibe and written Manchu (cf. Jang 2008) are in fact shared by Sanjiazi, Aihui and Yibuqi and may therefore be interpreted in terms of differences between the spoken language on one hand and the written form on the other, rather similar to the difference between written (Classical) Mongolian and the modern spoken forms of Mongolian. Furthermore, similar to the situation in Mongolian, it may be assumed that, besides reflecting an earlier shape of the spoken language, some of the features in written Manchu may be orthographic conventions rather than of records of the actual pronunciation.¹⁰

⁷The ancestors of this group were moved from Manchuria to Xinjiang in 1764 as soldiers of the Manchu army with the task of manning the frontier garrisons on the border with Russia. For detailed accounts of the history of the Xinjiang Sibe see, for example, Sárközi (2019) or Zhuangsheng (2019).

⁸The term *Jungarian Sibe* is employed by Janhunen (1996: 49).

⁹Cf. e.g. Zhao (1989). Chinese authors use the term *Qiren Manyu* ‘Bannermen Manchu’ to distinguish the standard Manchu language from the varieties used in communities of Manchu/Jurchen civilians whose language was not subject to so intensive standardization, such as Alchuka or Bala.

¹⁰An example of this – the difference between the notation and the actual pronunciation of the Manchu past tense forms – was analysed by Kubo Tomoyuki in his lecture (Charles University Oct 4 2019). It should also be noted that the Manchu writing system, similar to the Mongolian script, ignores most allophones of the spoken forms.

This homogeneity of the Bannermen Manchu varieties notwithstanding, several distinctions exist between Sibe on the one hand and the other Manchu varieties on the other. These distinctions comprise phonetic, morphosyntactic and lexical features. Some of these features are likely to have originated in contact with Mongolic languages.

3.1.1 The historical background of the Sibe

The Sibe are first found in Central Manchuria, in the areas of Qiqihar and historical Bedune (the modern Fuyu city). The first substantial evidence about them is a note about the inclusion of the Sibe into the Manchu military system in 1692, found in the Records of Girin (Zhuangsheng 2019: 51; Sárközi 2019: 8). In noting this event, the source gives the retrospective detail that Sibe and Gūwalca¹¹ had been Khorchin vassals. The transfer of Sibe and Gūwalca from the Khorchin under direct Manchu administration was mediated by the Second Neichi Toyin¹², in whose biography the description of the event is given (Ujeed 2013: 232–233). This is the historical base of the narrative about the Sibe vassalage to the Khorchin. Except for these accounts, other brief mentions confirm the relationship of the Sibe and Gūwalca to the Khorchin (Gorelova 2002: 35) – namely the account of the battle of Gure (1593) when Sibe and Gūwalca fought together with the Khorchins and the Hūlun Jurchens against Nurhaci, and a mention of the Sibe and Gūwalca as Khorchin vassals in the biography of the all-important Buddhist missionary to the Khorchin, the First Neichi Toyin (between 1636 and 1653, cf. Heissig 1980: 36).

Especially the account of the Battle of Gure places the Sibe into the context of the Hūlun Jurchens, about who Crossley (2006: 65) writes: “The majority of Hūluns were Jurchen in origin but by the late 1500s spoke a distinct dialect, with a much larger portion of Mongolian loan-words, and among them were found a very high incidence of Mongolian names, marriage into Mongolian-speaking lineages (either Khorchin or Kharachin), and extensive acculturation with the Khorchin or Kharachin populations generally.” The Khorchin and Kharachin were, in their majority, descendants of the Ujiyed and Uriangkhan Mongols respectively (see below).

¹¹The Gūwalca (known as *Khuulchin* in Mongolian sources, cf. Ujeed 2013: 232–233) are mentioned together with the Sibe in the early Qing period. By the 19th century they have disappeared, possibly due to merger with the Sibe. Their language is not documented at all but they are generally considered to be linguistically related to the Sibe (Zhuangsheng, p.c. August 2019).

¹²For a detailed description of the activities of the Second Neichi Toyin (1671–1703), a successor and re-incarnation of the famous Buddhist missionary to the eastern Mongols, the First Neichi Toyin, see Ujeed (2013).

Consequently, the Sibe, together with the Gūwalca, were probably involved in the intensive contact processes on the borders between the Mongolic- and Tungusic-dominated parts of Manchuria (Janhunen 1996: 98–99). The historical accounts of the event of incorporation of the Sibe and the Gūwalca into the Manchu banners state that these two groups were related to the Jurchens. These people, whatever their political status was, can thus probably be taken as representatives of Jurchenic groups of the Mongolic-influenced area. They were acculturated by Mongols who, in their turn, were linguistically and culturally Tungusic-influenced, and themselves were, in part, Mongolized Tungusic speakers (see below). Interestingly, Crossley (2006: 65)¹³ notes that “the Jurchens of Nurgaci’s time used the word *Mongol* (*monggo*) for the Hūluns”, which could have likewise influenced the traditional self-perception of the Sibe¹⁴. In 1636–1638, the Sibe, together with the Gūwalca, the Daur and possibly other originally Hūlun groups (cf. Crossley 2006: 69–70), were incorporated into the newly created Mongol Eight Banners, to be transferred to the Manchu Eight Banners in 1692.

While the abovementioned historical sources confirm the fact that the Sibe were Khorchin subjects, they do not give details about this relationship and its duration. It is, however, clear that Sibe lived in a Mongolic-influenced environment for two or three centuries before becoming Manchu bannermen. After becoming Manchu army soldiers, they were divided into several groups and relocated into several military garrisons in Manchuria and Inner Mongolia (Gorelova 2002: 36). There they were organized into the Sibe banners. Initially, the Gūwalca had their own banners but later were probably merged into the Sibe banners (Zhuangsheng, p.c. August 2019), in this way disappearing from history. In different garrisons the Sibe came into contact with different – Tungusic and Mongolic – speakers. As Manchu bannermen they probably participated in the processes described by Atwood (2005: 9–12), and others. These processes involved, on one hand, intensive merging which resulted in the common milieu of Manchu bannermen, also known as Qizu, literally ‘Banner ethnic group’, in the beginning of the 20th century, cf. Chengzhi (2021). High prestige of Standard Manchu was one of several important traits of this milieu. On the other hand, identification with particular banners created the notions of Sibe, Solon, Daur and other groups based on administrative affiliation rather than origin and language. Thus “Sibe”

¹³Crossley (2006: 65) quotes the source *Huangqing kaiguo fanglüe* 3.3a. written by Agui et al. For a brief description of the ethnic setting of Central Manchuria in late Ming based on contemporary sources see Crossley (2006: 64–66).

¹⁴This tradition of viewing the Jurchenic groups of central Manchuria as Mongols may also stand behind the appellation “Sibege Mongols” for a sinicized group of Manchurian Sibe mentioned in Lattimore (1935: 225–227).

in the Qing period largely referred to people affiliated with the Sibe banners which could include people of different linguistic background. The thus constituted Sibe identity was distinct from that of the Manchus and rather close to that of the Daur, Solon and Butha (cf. Elliott 2001: 85). In 1764, 1000 individual soldiers were picked up from different Sibe banners (Sárközi 2019: 9) and with their families were transferred to their present location in Xinjiang. Closer study of these developments leads historians to question the continuity between the pre-Qing Sibe and the modern Sibe in Xinjiang (e.g. Chengzhi 2012: 257–268).

During the Qing period Sibe came into close contact with other Mongolic groups, such as the Daur, the Chakhar or the Öölöd. Nevertheless, Standard Manchu became their first language. Throughout the Qing rule and until modern times, Sibe have been known for their solid Manchu skills (Zhuangsheng 2019: 51).

3.1.2 The socio-linguistic background of the narrative about the non-Jurchenic origins of the Sibe and of the Sibe-Khorchin contacts

In the beginning of the 20th century, the fact that the Sibe people in the vicinity of *Ghulja* (Mongolian *Ili hot*, Chinese *Yining shi*) spoke Manchu had been widely recognized by the speakers themselves (e.g. Donjina 1989; Porter 2018: 10–12), as well as by foreign travelers and researchers (e.g. Kałużyński 1987). Historical sources confirm that Sibe spoke Manchu as at least one of their languages during the whole Qing era (Zhuangsheng 2019: 51). However, in 1990, when I visited the Xinjiang Sibe community for the first time, any relationship to Manchus was generally denied in the official discourse among Sibe intellectuals. The language of the Sibe was called *Sibe*. Moreover, several of my Sibe consultants were suggesting that Sibe originally spoke a Mongolic or Mongolic-related language. The remarkable difference between the written Manchu language (known by many in the older generation of Sibe) and spoken Sibe¹⁵ was mentioned in support of this idea. Sibe was presented as a language on its own, distinct from Manchu. Publications influential in Sibe society described Sibe culture without the Manchu context and studies of Sibe history argued for an ethnic origin distinct from that of the Manchus.¹⁶

¹⁵This difference involves not only features which seem to reflect diachronic processes such as vowel reduction or consonant weakening, but also features which call for other interpretations such as dialectal variation (namely in lexicon and morphology). In the 1990s the Sibe were generally not aware that many of these distinctions were shared by the oral varieties of Manchuria.

¹⁶The basic comprehensive description of Sibe folk culture is *Xibozu minsu – Sibe uksurai an tacin* (He & Tong 1989), the main description of Sibe ethnic history was *Xibozu jianshi/Sibe uksurai šolokon suduri* (Wu et al. 1985).

As Zhuangsheng (2019: 58–70) has shown, this narrative came into being at the beginning of the 20th century and became essential in the context of the creation of the 55 ethnic minorities during the 1970s. Evidence for a distinct origin and a history as an ethnic group of its own was required in order to be officially recognized as an ethnic minority and enjoy the advantages associated with this status. Another reason why the Sibe strongly denied common origins with the Manchus was the persecution of ethnic Manchus which started in Republican China and continued into the PRC period. Zhuangsheng (2019: 58–71) describes how the Sibe intellectuals worked on collecting historical evidence for writing a *Sibe history*. He concludes (2019: 71–72) that Sibe as a political or ethnic entity indeed occur in historical sources since early 17th century. However, the whole narrative about their relationship to the presumably Mongolic-related Xianbei and their early history since the 3rd century¹⁷ was made up without any historical basis, and with very little background in oral tradition. This narrative has become part of the modern Sibe identity.

3.1.3 A story of a “different original language”: The case of the *jivš* language

The story of the extinct *jivš* language is an example of a detail from Sibe oral tradition that became an important part of the Sibe “ethnic narrative” and (linguistic) self-consciousness as a non-Manchu group.¹⁸

¹⁷The official Sibe history uses several unclear mentions found in oral tradition to argue that the ethnonym *Sibe* is related to the name of the Xianbei, a presumably nomadic group from western Manchuria which ruled over the Mongolian grasslands in the 2nd century. The Xianbei language has been most often interpreted as Mongolic (e.g. Janhunen 2010: 281). This hypothetical Xianbei connection of the Sibe has been used in support of the argumentation for a non-Jurchen origin of the Sibe.

¹⁸As for the possible identity of this enigmatic language, the Inner Mongolian linguist Otgonchecheg suggested a connection to the Chipchin (Bargu: *šivšin*), an exonym used for the Old Bargu (a Buryat-related Mongolic group) during the Qing. Otgonchecheg, who did fieldwork in Chabchal in order to collect data of the *jivš* language, did not publish her research due to the lack of evidence. From a historical point of view it is plausible that a group of Chipchin Bargu bannermen was incorporated into the Sibe banners. However, the Sibe scholar Su Deshan (1984), based on his fieldwork in the Fifth banner, maintains that the term *jivš gisun* referred merely to a layer of Khorchin loanwords which was thicker in some groups of Sibe than in others. Su Deshan, following a “folk” explanation, interprets the word *jivš* as ‘double, additional’ and the term *jivš gisun* as ‘additional words, synonyms’. Small pieces of evidence from more recent fieldwork (Guo Junxiao, Chengzhi, p.c. September 2020) suggest that the notion of *jivš gisun* is still remembered in the Fifth banner, currently pointing to a mixture of Mongolian loanwords and Literary Manchu expressions which are marginal, though not entirely unknown, among the rest of the Chabchal speakers. Guo Junxiao, a Sibe speaker (p.c. 2020) describes *jivš gisun* as a group of “unfamiliar, Mongolian-sounding words” while the unpublished data collected by Chengzhi (2020) include lexical items such as *saxaxuri* ‘whitish’ (< written Manchu *sahahūri*) and *xurdun* ‘quick’ (< written Mongol *qurdun*, Khorchin *xurden*, vs. Sibe *xudun*, written Manchu *hūdun*).

The inhabitants of the Fifth banner, one of the eight administrative units of Chabchal, speak Sibe with a (for a native speaker) remarkably different pronunciation. The difference supposedly consists of lesser reduction and generally greater closeness to written Manchu. Sibe speakers from other banners often quote the example of the written Manchu word *aliyaha* ‘waited’ which is pronounced as *aliaxa* in the Fifth banner but *alixe* in the rest of Chabchal. Oral tradition explains this by saying that Sibe of the Fifth banner were originally speakers of a different language and therefore were taught Standard Manchu as a new language. This caused their pronunciation in the spoken language being closer to the literary language. Oral tradition calls their original language *jivš gisun* ‘the *jivš* language’ (written form *jibsi gisun*), and holds that it had disappeared by the end of the 19th century. Different ‘folk’ hypotheses exist about this language, such as that *jivš gisun* was a “Mongol language, perhaps something like Khorchin or Daur” or that it was a “secret language which consisted of repeating every word twice.” (fieldwork data February 1995). Moreover, now and then a statement is heard or read that *jivš gisun* was the original language of the Sibe.

Whatever the historical roots of the *jivš* case, it has become part of the popular narrative of Sibe identity. Even today the statement about *jivš* as the original language of the Sibe, attributed to a source called “minjian” (folk), is repeated on Sibe social media,¹⁹ which testifies to its lasting popularity.

3.2 Khorchin

Khorchin Mongol, spoken by close to a million of speakers and thus being the largest and most influential Mongolian dialect after Khalkha, is less researched than Sibe. The Khorchin speech community differs from most other Mongolian speech communities in that it has a long tradition of sedentary or semi-sedentary life-style. Two important descriptions of Khorchin are Bayančogtu (2002) and Caidengduerji (2014), the latter being an unpublished dissertation.²⁰

At present, Khorchin is spoken over a large territory in Inner Mongolia and the neighboring provinces of Jilin and Heilongjiang. The locations with the greatest concentration of speakers are the administrative unit of Tongliao City and the Hinggan League in Inner Mongolia. The varieties spoken in these two areas

¹⁹ *musei te gisuremaha gisun oci manju gisun inu, musei da gisun oci jibsi gisun, manju gisun waka* ‘the language we speak now is Manchu, but our original language is the Jibsi language, not Manchu’. (E.g. http://blog.sina.com.cn/s/blog_4aa943a1010008yv.html. Last access 28.10.2020.)

²⁰ Other studies and materials of Khorchin include, for example, Brosig (2014a,b) and Yamakoshi (2015).

slightly differ from each other. Khorchin is close to two other large eastern Mongolian varieties – Kharachin and Baarin – and the three, including a number of their sub-varieties, share some important differences from the rest of Mongolian. The Tongliao variety, in particular, is hardly intelligible to speakers of most other modern Mongolian languages.

However, the available descriptions of Khorchin present a picture of a rather regular variety of modern Mongolian and do not give sufficient explanation for the mutual unintelligibility with standard varieties such as Khalkha.

In my observation, two main factors may be responsible for the surface difference of Tongliao Khorchin from other modern Mongolian varieties. First, Khorchin retains, with certain exceptions such as the loss of the vowel *ö*, the general phonological structure that goes back to Proto-Mongolic (e.g. Janhunen 2003b: 4). However, extensive processes on the phonetic level such as consonant weakening, vowel shifts and vowel reduction fundamentally change its shape in speech. Second, Khorchin in most rural areas is profoundly influenced by Chinese with which it has been in close contact for several centuries. Chinese influence is mostly manifested in syntax (e.g. paratactic constructions instead of chains of clauses connected by non-finite verbal forms, which are typical for most other modern Mongolian languages) and vocabulary. Depending on the topic and circumstances, the speech of a Khorchin speaker may consist of about fifty percent of words of Chinese origin. These features are not readily seen in the descriptions but are important for shaping the performance of Khorchin speakers which then radically differs from the speech of, for example, a Khalkha speaker.

3.2.1 Historical background of the Khorchin

The Khorchin population seems to have initially been composed of two main elements. The first, the most important according to Khorchin historians, and the one which gave the group its name and proclaimed identity, is the Mongol noble lineage descended from Khasar and their subjects. In the 13th century Khasar, the younger brother of Genghis Khan, was granted the lands around Lake Hulun and the Ergune river as an appanage, hence approximately the area of the modern administrative unit of Hulunbuir.²¹ During the Ming dynasty, probably in connection to the period of internal conflicts in Mongolia (Caidengduerji 2014: 29), the main part of the Khorchins crossed the Khingan mountains to the east and

²¹It is often difficult to establish the precise location of the lands of particular nomadic peoples in this period. In the case of the Khasar lineage, however, archaeologists have interpreted at least two important sites in the Ergune valley as towns built by Khasar's descendants (e.g. Kradin 2018: 227–227).

settled in the Nonni valley where they became the overlords of the local Mongol population. The local Mongols, the second important – and probably more numerous – element in the composition of Khorchins, were the Ujjiyed of the Fuyu Guard²² (Atwood 2004: 306). The Fuyu guard was one of the Three Guards – administrative units in Manchuria loosely controlled by the Chinese (Ming) court. The population of the Three Guards was referred to as either Mongol or Uriangkhan, but comprised, besides Mongols, groups of Tungusic origin.²³ Therefore, in imagining the linguistic situation during the Ming, it seems important that the population of the Three guards, which later²⁴ “became the ancestors of many eastern Inner Mongolian peoples” (Atwood 2004: 35), was probably largely homogenous in terms of language and culture²⁵ which contained elements of Tungusic origin (Crossley 2006: 82). In addition to this picture, the Three Guards were geographically close to the former Khitan territories, and their settlement in the area probably goes back to times when Khitans still existed as a distinct entity. Therefore a certain Khitan influence on Khorchin cannot be excluded.

Consequently, the remarkable features shared by the eastern Mongolian dialects – Khorchin, Kharachin and Baarin – may in fact have originated in the language of the Three Guard Mongols who have been continually exposed to local Manchurian influences since as early as the Yuan period.

Since the 15th century the Khorchins often intermarried with Jurchens (Crossley 2006: 65). Since their arrival they started migrating from the Nonni valley southwards, into their present territory in the Liao valley. According to a contemporary account of a Korean observer, they were “dressed in furs, with their felt yurts on wagons, moving their herds toward appropriate pastures. Many, he noted, were also agricultural and would sow fields in the spring to which they

²²The Fuyu guard, situated close to the present Qiqihar in the Nonni valley, was one of the three “loose rein” guards (the Fuyu guard of the Ujjiyed people, the Taining guard of the Ongniuts and the Döyün guard of the Uriangkhan) established in Manchuria by the Ming. The “Guards” were groups of former subjects of the Yuan empire who were identified as Mongols and after the fall of the Yuan rule became tributaries of the new Ming dynasty (Atwood 2004: 536).

²³Crossley (2006: 64) refers to the Ming authors Xiao Daheng and Ye Xianggao for a definition of “Mongols” in the Ming era, concluding that: “[...] some Mongolian-speaking communities were not nomadic but agricultural; many groups who migrated with “Mongols” were speakers of Turkic or Tungusic languages; many living among the Mongols were Han or the descendants of Han, who had been taken by the hundreds of thousands by eastern Mongol raiders in northern China.”

²⁴For the detailed descriptions of the migrations of the Three Guards and their mixing with other Mongols see Atwood (2004: 304, 410).

²⁵The Three Guard Mongols were mostly sedentary and practiced agriculture (Atwood 2004: 535).

expected to return in the fall to reap a meager crop of wheat or millet.” (Crossley 2006: 66). During the Qing period the Khorchins took over the Liao valley and thanks to their alliance with the Manchus politically dominated the area. At the same time groups of outsiders settled on this territory and were integrated and assimilated by the Khorchins (Caidengduoerji 2014: 37). These immigrants were both large groups of Manchus and Chinese and smaller groups or individuals of other ethnic origin such as Sibe, Ewenki, or Koreans. In the beginning of the 20th century the Khorchin area became one of the main targets of the Qing New Policies, which involved an unrestricted immigration of Han Chinese and further sedentarization of the local Mongols. Even during the 20th century, however, many immigrants kept adopting the Khorchin language and culture.

4 Evidence of Sibe-Khorchin contacts

In this section I list some shared features of modern Sibe and modern Khorchin, which may have resulted from mutual contacts between the ancestors of the two modern groups. These features, in my opinion, indeed point in the direction of direct contact of some kind. Historically and linguistically, these features remain open to different interpretations. When taking into consideration the available evidence about “ethnic” and “linguistic” mobility in Manchuria, especially within the Eight Banners,²⁶ it is rather clear that it is impossible to entirely separate the

²⁶In Qing-time Manchuria large-scale migrations and resettlements are documented, such as the abovementioned resettlement of Sibe, Khorchin migrations, or the massive Daur and Solon migration into the Qiqihar area in the 17th century. In addition, evidence of countless shifts of small groups and individuals among the Qing garrisons is scattered across historical sources. Another factor important for linguistic developments are frequent intermarriages among members of different banners which were supported by the strict rules of exogamy in Tungusic-speaking groups. Among these, intermarriages between Sibe and Manchu bannermen seem to have been common (He Rongwei, p.c. June 2020). Intermarriages between Khorchin and Manchu speakers are generally known to have been frequent (Shuangshan, p.c. August 2015). If we take the longest-surviving “banner society” – that of Hulun Buir – as a model for the linguistic situation in the Manchurian Banner communities, we may assume that not only many bilingual couples lived in the Banners but most of the bannermen were, to a certain degree, familiar with other languages. The supposed constant language contact between the Sibe and Manchu bannermen and the Khorchins rules out the possibility of independent developments of these languages and any clear-cut evidence for the earlier direct contacts between the Sibe and the Khorchin.

It also needs to be taken into account that the available data of spoken Manchurian Manchu represent tiny pieces of a once broad continuum of local varieties, and that much of the data available were collected from semi-speakers and rememberers, and thus cannot supply a complete picture of Manchurian Manchu.

linguistic developments in Sibe from the other Manchu varieties. However, the features listed below are central and massive in Sibe while, if attested, marginal in Manchurian Manchu.

4.1 Manchu influence on Khorchin?

For obvious reasons – namely the absence of any Sibe data before the 20th century – any specifically Sibe influence on Khorchin cannot be determined. In the context of the historical developments described above, strong influence of Manchurian Tungusic varieties might be expected. Quite surprisingly, however, little influence is seen on the lexical level. While Chinese loanwords form a significant part of the Khorchin vocabulary, Manchu loanwords do not seem to exceed several tens. Words used in everyday life such as *lah* for the brick bed (Chin. *kang*) from Manchu *nahan* or kinship terms such as *eme* for mother (Manchu *eme*) have been noted by native linguists (Bayančoytu 2002: 25). Some Manchu loanwords are connected to shamanic practices, such as *samaan* ‘shaman’ from Manchu *saman*, *sarg* ‘home altar’ from Manchu *sarha* or the verb *magsi-* ‘to perform shamanic dance’ from Manchu *maksi-* ‘to dance.’ On the level of morphology and morphosyntax, the general typological similarity of Manchu and Mongolian makes it difficult to single out instances of mutual influence.

The small number of Jurchenic loanwords in general may, at least partly, be attributed to the standardization forces during the Qing dynasty which affected Mongolian (proper)²⁷ speakers not less than Manchu speakers. In spite of the fact that the Mongolian script was invented before the Yuan times, it became widely used only since the 16th century with the spread of Buddhism, accompanied by translations of literary works into Classical Mongolian. At the same time, original compositions of didactic and other character were written and read in Mongolian-speaking societies. The influence of Classical Mongolian could have brought the vocabulary of the (politically) Mongol groups of Manchuria closer to other Mongolian varieties (Crossley 2006: 83).

In terms of contact features, research into phonetic peculiarities of Khorchin and their relationship to the language environment of Manchuria may prove more rewarding. It seems worthwhile to analyze Khorchin phonetic and phonological differences from other Mongolian varieties in the context of other eastern Mongolic idioms (Baarin, Buryat, Daur), in the context of Manchu varieties, Manchurian Mandarin and possibly even the of language of the Korean minority of China.

²⁷In contrast to Mongolian proper, the Mongolic Daur language was not affected by standardization, instead borrowing many Manchu words.

Below I just note two features which are similarly typical for Sibe among Manchu varieties as for Khorchin among Mongol dialects and may therefore be added among the candidates for results of direct Sibe-Khorchin language contact.

4.2 Shared phonetic developments in Khorchin and Sibe

Generally speaking, Sibe and Khorchin are phonetically strikingly similar, which seems to be caused for the most part by the Manchurian influence on Khorchin. For example, Khorchin is perhaps the only Mongolian variety where the intervocalic cluster *ŋg* is pronounced as syllable-initial [ŋ], as in [moŋol] ‘Mongol’. Still, however, two of the shared features may be interpreted as results of phonetic processes that Manchurian Manchu has avoided.

4.2.1 Change of closing diphthongs into opening diphthongs

In Sibe, the equivalent of the written Manchu diphthong *ai* is often pronounced as *iä*, e.g. written Manchu *bayimbi*²⁸ [pajmbi] vs. Sibe *biäm* [pjɛm] ‘to look for’, etc. This is valid for approximately half of the reflections of the written Manchu *ai*. The rest either remains as *äi/ai* or is monophthongized. Some instances of retention of the closing diphthong are in the word-initial position (e.g. written Manchu *ai*, Sibe *ai* ‘what’), others come after uvulars (e.g. written Manchu *kaicambi*, Sibe *qaicem/qacim* ‘to shout’), or apparently belong to a more literary style (e.g. written Manchu *saikan*, Sibe *saiken* ‘beautiful’). In other cases such as the written Manchu *baita*, Sibe *bäit* there is no immediately apparent reason. The “reversal” also took place in a few cases of the closing diphthong *oi* (e.g. written Manchu *boihon*, Sibe *bioxun* ‘dust’). These changes fit into the context of the overall phonetic tendencies in Sibe (vowel raising and fronting, e.g. written Manchu *omimbi*, Sibe *eimim/iemim* ‘to drink’).

In contrast to Sibe, in the spoken Manchurian varieties of Sanjiazi, Aihui and Yibuqi monophthongization of the written Manchu diphthongs occurs (e.g. written Manchu *sain*, Sanjiazi *sän* ‘good’), but there are no cases of “reversal” of the diphthongs.

Unlike Manchurian Manchu but quite similarly to Sibe, Khorchin has a strong tendency towards vowel fronting and raising (Janhunen 2012: 60–61). Closing diphthongs of written Mongol (which are either retained or monophthongized in the central Mongolian varieties such as Khalkha) are, at least in some Khorchin varieties, almost regularly reversed, e.g. written Khalkha *naim*, Khorchin

²⁸Unlike the pronunciation in spoken varieties, academic pronunciation of written Manchu unpacks the diphthong.

nie:m ‘eight’ or written Khalkha *meiren*, Khorchin *mie:rin* (title of an official). The reversal may involve change of vowel quality such as written Khalkha *xoit*, Khorchin *xie:t* ‘north’.

Janhunen (2012: 45) notes that the tendency towards vowel fronting is seen in Mongolian in general but this process has been most complete in the eastern dialects including Khorchin. Similarly, reversal of diphthongs occasionally happens in other Mongolian varieties but has become regular in Khorchin. The described feature of Sibe may therefore be interpreted as a diachronic change that happened during the period of influence of the eastern Mongolian phonetic environment but was halted when the Sibe left this particular environment.

4.2.2 Dissimilation of the cluster *čx*

There is another phonetic development that occurs in Sibe and Khorchin but is found neither in other Manchu varieties, nor in any other Mongolian variety. In spoken Sibe the consonant clusters *čk* and *čx*, which result from vowel elision, often change into the sequence *šk*, e.g. written Manchu *tacikū*, Sibe *tačqu/tašqu* ‘school’ or written Manchu *tacihabi*, Sibe *tačxei/tašqe* ‘studied’. The dissimilated forms are used in quick and less careful speech, while the careful pronunciation retains the original consonants. In Khorchin, the cluster *čx* in the Mongolian deverbal suffix *-čix/-čx-* (quick or intensive action) in quick speech is sometimes dissimilated in a similar way. e.g. *yavšgen/yavčxen* cf. written Khalkha *yavčixna* ‘will leave’. While this may be just a parallel development, it certainly contributes to the similarity of the two languages.

4.3 Potential Khorchin influence on Sibe grammar

In the next part I list those features of Sibe grammar which have analogies in Khorchin and are not shared by, or are marginal in, the other oral Manchu varieties.

4.3.1 The emphatic prefix *mV-* (used with deictics)

4.3.1.1 The prefix *mV-* in Sibe

Sibe has the element *me-/mu-* which is added to the beginning of some deictic expressions. Generally it adds emphasis to the deictics and is possibly best translated as ‘just, exactly’, sometimes ‘the very’. Its use is often analogous to the Chinese particle *jiù* ‘just, exactly’, sometimes also ‘the same’.

The prefix is at least partly productive. Below I list forms encountered in my fieldwork material with examples:

- *mere* ‘exactly this’, from *ere* ‘this’ (1).

(1) *mere jilgan mim-be eme diower amxe-we-xa-qv.*
 just.this sound 1SG-ACC one night sleep-CAUS-PTCP.PFV-NEG
 ‘It was exactly this thing which did not let me sleep the whole night.’

- *metere* ‘exactly that’ from *tere* ‘that’ (2).

(2) *metere baite-we giser-maie.*
 just.that matter-ACC speak-PROG
 ‘This is exactly what I am speaking about; I am speaking about the same thing.’

- *merange* ‘exactly like this’ from *erange* ‘like this’ (3).

(3) *min-i uwe=da merange.*
 1SG-GEN fate=FOC just.like.this
 ‘This is exactly what my fate is (I cannot change it).’

- *meterange* ‘exactly like that’ from *terange* ‘like that’ (4).

(4) *meterange=da are!*
 Just.like.that=FOC write.IMP
 ‘Write it exactly in that way!/ Just write it in that way!’

- *meske* ‘just this much’ from *eske* ‘this much’ (5).

(5) *bilxa=ni meske ma.*
 neck=3SG.POSS just.this.much thick
 ‘His neck is just this thick. (This form is usually used when demonstrating the degree of something with a gesture.)’

- A lexicalized expression formed in the same way is *mujagen* ‘just now’ from *jaqen* ‘a while ago’.

The form *mere* ‘exactly this’ is further used as means of emphasis with different types of expressions, both with deictics (6) and with other words (7), (8). In this case it rather adds emphasis to the whole sentence than to its determinandum.

- (6) *mere ewade=da yinde!*
EMPH here=FOC stay.overnight.IMP
'Just stay here overnight! (emphasis for the sake of expressing hospitality).'
- (7) *mere xancide min-i bo-de ji-xei.*
EMPH recently 1SG-GEN house-DAT.LOC come-PST
'He came to my place during the very last couple of days.'
- (8) *mere feksi-m.*
EMPH run-NPST
'He took a flight/ immediately started running.' (emphasis in storytelling)

This feature is very likely borrowed from Khorchin, where the element *m(V)*- has an analogous function.

4.3.1.2 The prefix *mV*- in Khorchin

According to Bayančoytu (2002: 148–151), in Khorchin this prefix is fully productive with demonstratives. In his description the author gives a list of more than 120 possible forms. Below I give examples from my fieldwork material:

- *men/mun* 'exactly this' from *en* 'this' (9).

(9) *Tongliao-nii laajii-gii men dotor avšir-č.baina.*
Tongliao-GEN waste-ACC just.this inside bring-PRS.PROG
'It is (exactly) inside this (fence) they are bringing the waste from Tongliao.'
- *meter* 'exactly that' from *ter* 'that' (10).

(10) *meter modon.eel šii.*
just.that PN EMPH
'It was that very Modon eel.'
- *miim* 'just like this' from *iim* 'like this' (11).

(11) *huu miim miim budun.*
all just.like.this just.like.this thick
'They were all just this thick.' (showing)'

- *mitiim* ‘just like that’ from *tiim* ‘like that’ (12).

(12) *mitiim sanaa-tai ir-jee.*
 just.like.that idea-COM come-PST
 ‘I came exactly with this idea in mind (I came exactly for this purpose).’

- *mengeed* ‘(doing) in this very way’ = ‘just like this’ from *engeed* ‘(doing) in this way’ (13).

(13) *mengeed neg tangs mod ux-jee.*
 just.in.this.way one row tree die-PST
 ‘And in this very way the whole row of trees died.’

- *metgej* ‘(doing) in that very way’ from *tegej* ‘(doing) in that very way’ (14).

(14) *metgej or-j ir-sen šdee.*
 just.in.this.way enter-CVB.IPFV come-PTCP.PFV EMPH
 ‘This is the very road we took on the way here.’

- *mudii* ‘exactly this much’ from *udii* ‘this much’ (15).

(15) *mudii gonjgoil-son.*
 just.this.much be.oblong-PTCP.PFV
 ‘(Its shape was) oblong, this long (showing).’

- *mendegu/mundugu* ‘exactly from here’ from *ende-gu* [here-NMLZR] ‘coming from here, local’ (16).

(16) *nienie-nii ug suugaal ger bol mende-gu ii?*
 grandmother-GEN original seat home TOP just.here-NMLZR Q
 ‘Grandmother, are you originally from this very place?’

The forms listed above are mostly found in eastern Mongolian dialects, even though in recent years they started being occasionally used by speakers of other Inner Mongolian varieties. The word *meter*, which is also used as a filler, is so prominent that Mongols in some other parts of Inner Mongolia used to mock Khorchin soldiers by calling them *Meteruud* ‘the Meters’.

This element *mV-* has most probably evolved from the Mongolian emphatic pronoun *mön* (written Mongol ‘the same, just this’, Poppe (2006: 51), Proto-Mongolic ‘the very, the same’, Janhunen (2003b: 20). In modern Mongolian proper it has been mostly used as an (often emphatic) copula, e.g.

- (17) *bi Dorj mön.*
1SG Dorj COP
‘I am (indeed) Dorj.’

and as an emphatic particle, e.g.

- (18) *Ulaanbaatar utaa-güi bol mön goyo.*
Ulaanbaatar smog-PRIV TOP PTC nice
‘It would be really nice if Ulaanbaatar was without smog.’

While combining the particle *mön* with deictics is occasionally found in many of the modern Mongol varieties (e.g. Khalkha *mön ter xün* ‘that very person’), its grammaticalization into a kind of prefix has only taken place in Khorchin and the adjacent eastern Mongolian varieties. In other spoken Manchu varieties mainly the form *meter* is attested (Wang 2005: 155) but seems to be marginal compared to its massive use in Sibe. Another interesting question is that of the Sibe word *menjang* ‘indeed, truly’ which is used in positions corresponding to the use of the word *mön* in Mongolian. This expression is attested in written Manchu in the form *mujangga*. No plausible Jurchen etymology for this word seems to be at hand, therefore a connection to the Mongolian form *mön* may be considered. In the whole, the above-mentioned Sibe set of emphatic deictic expressions is one of the candidates for a proof of direct and intensive contact between the ancestors of modern Khorchin and Sibe.

4.3.2 Replacement of personal pronouns with demonstratives

Grammars of written Manchu give the 3rd person pronouns as *i* (3SG) and *ce* (3PL) which are regularly inflected for case. In Manchu texts, especially in the more “natural” ones such as historical narratives the demonstrative plural forms *ese* ‘these’ (singular *ere* ‘this’) and *tese* ‘those’ (singular *tere* ‘that’) are used more frequently than *ce*. As plural forms²⁹ they are generally reserved for human or human-like beings, thus being in fact personal pronouns. In the oral Manchu varieties (Sanjiazi, Aihui and Yibuqi) the 3rd person plural pronoun *ce* has been

²⁹In Manchu only nouns denoting people, deities or ghosts are marked for number.

completely replaced by an oral form of *tese* (Wang 2005: 52 *tetse*, Zhao 1989: 123 *ts'etse*, etc.). A form derived from the 3rd person singular pronoun *i* is, however, attested in all three varieties: Sanjiazi: *yin*, Aihui *i* (Wang 2005: 52), Yibuqi *ji* (Zhao 1989: 189). These forms are noted as used along with the demonstrative *tere/tele* 'that'.

In Mongolic, already in the Middle Mongol period the Proto-Mongolic 3rd person pronouns *i* (singular) and *a* (plural) have been generally replaced by the demonstratives *ene/tere* for singular and *ede/tede* for plural (Rybatzki 2003: 72).

In Sibe the 3rd person pronouns are not attested at all, even though knowledge of the literary language and thus also of the forms *i* and *ce* was widespread till the 20th century.

Hence, the tendency towards replacement of 3rd person pronouns by demonstratives exists not only in Mongolic, but also in Manchu. Systematic usage of personal pronouns in written Manchu may be regarded as a conservative feature and is being abandoned in less canonical Manchu writing. The process, however, is on half-way in Manchurian Manchu while it has been completed in Sibe.

Admittedly, this is a cross-linguistically common process and does not tell anything about the Khorchin-Sibe contacts. However, it is still possible that a direct influence of a Mongolic vernacular on Sibe has accelerated the change that was already underway in the spoken Manchu varieties – the complete loss of the Manchu pronominal form and its replacement with demonstratives which are, moreover, almost homophonous in Mongolian and Manchu.

4.3.3 Possessive clitics and Sibe phrasal possession

Sibe has a system of possessive clitics which resemble the Mongolian possessive clitics and do not occur in any other Manchu variety. Their function is similar, specifically, to Khorchin. Much in the same way as in most modern Mongolian languages including Khorchin, the 3rd person possessive clitic functions as a definite marker or a topicalizer (cf. Hölzl 2017).

Furthermore, Sibe uses the 3rd person possessive clitic to express possession in a way which resembles the prototypical Tungusic head-marked possessive phrases (cf. Gorelova 2002: 45).

4.3.3.1 Phrasal possession and definite marking in Manchu

In written Manchu the principal way to express possession and association is marking on the dependent which then takes the genitive (or genitive-instrumental) suffix, e.g. *min-i bithe* [1SG-GEN book] 'my book'; *morin-i uju* [horse-GEN

head] ‘horse’s head’ or ‘horse head’; *tacikū-i sefu* [school-GEN teacher] ‘teacher of the school/school teacher’. Written Manchu has no possessive clitics.

In the spoken Manchurian varieties possession may be dependent-marked, which is obligatory if the possessor is a pronoun. In other cases juxtaposition is common. However, while no possessive clitics are attested in the available materials, Sanjiazi uses the genitive marker *-ning* (< written Manchu marker of independent definite form *=ningge*) as a possessive and definite marker in the same way as Sibe uses the 3rd person possessive clitic *=ni*, e.g.

- (19) *ame-ning yawe-xei.*
 father-3SG.POSS go-PST
 ‘His father/the father left.’

4.3.3.2 Possessive markers in Mongolian

Most Mongolian varieties have a set of possessive markers which go back to reconstructed genitive forms of the Proto-Mongolic personal pronouns (Table 1).

Table 1: Proto-Mongolic personal pronouns (Janhunen 2003b: 18)

	Singular	Plural
1 st person	*mi.n-U	*bida.n-u
2 nd person	*ci.n-U	*ta.n-u
3 rd person	*i.n-U	*a.n-u

While in some Mongol varieties such as Buryat and Oirat these pronouns have been grammaticalized into possessive suffixes, others, like Khalkha and Khorchin, use slightly modified forms of the 1st and 2nd person possessive pronouns as clitics. Since the 3rd person possessive pronouns have been replaced by demonstratives, the system of possessive clitics has been supplemented with a “neutralized reflex of the original pronominal genitives” (Janhunen 2003a: 92) – the form *ni*. Consequently, the Khalkha possessive clitics are the ones shown in Table 2.

In Khalkha, all the enclitics are alternatively used to express possession along with the basic dependent-marked noun phrases. The choice of a clitic instead of a pronoun in genitive form may have semantic, stylistic or modality reasons, e.g. *min-ii eej* [1SG-GEN mother] ‘my mother (neutral)’ vs. *eej=miny* [mother=2SG.POSS] ‘my mother (expressing emotional attachment)’. The enclitics may be used instead of pronominal genitives in all functions of the latter, i.e.

Table 2: Khalkha possessive enclitics, possessive pronouns and personal pronouns (Svantesson 2003: 164)

	Possessive enclitic	Possessive pronoun	Personal pronoun
1SG	miny	minii	bi
2SG	ciny	cinii	ci
3SG	ny		-(<i>*i</i>)
1PL	maany	manai	bid
2PL	tany	tanai	ta
3PL	ny		-(<i>*a</i>)

possession, association, whole-part relationship (cf. Dixon 2010: 262). They also determine postpositions or indicate the agent in relative clauses. In Khorchin the frequency of clitics slightly differs from other Mongolian varieties: the 3rd person enclitic *=ni* [en] is frequent, closely followed by the 2nd person singular enclitic *šini* [ʃin]. In contrast, the rest, 1st person and 2nd person plural enclitics, are rare.

Examples of possessive enclitics in Khorchin:

- (20) *ger=ni dalan šagaan nohoi-tee.*
 house=3SG.POSS seventy white dog-COM
 ‘In their house there were seventy white dogs.’
- (21) *ger=šini bol bain aa šii?*
 house=2PL.POSS TOP wealthy EMPH Q
 ‘Was your house(hold) wealthy?’

In most modern Mongol varieties, possessive clitics are used in functions whose common denominator is probably best described as definiteness (Janhunen: “deictic determinants connected with the category of definiteness”). In some cases they “refer to the discourse situation” (Janhunen 2003a: 93). The 3rd person and 2nd person singular possessive clitics are the most common in this function. In Khorchin, only the latter two seem to be used as definite markers, e.g.:

- (22) 3rd person possessive enclitic *=ni* (22)
ter olson yum=ni ertnii, uldsen=ni bol suulernii.
 that bamboo thing=DEF ancient, the.rest=DEF TOP later
 ‘The one made of bamboo is ancient, the rest of them is more recent.’

- (23) 2nd person singular possessive clitic =*šini* (23)
ter uise-d=šini iim terg gue.
 that times-DAT.LOC=DEF such cart NEG.EX
 ‘In those times there were no such carts.’

4.3.3.3 Sibe possessive clitics

In Sibe a set of possessive clitics exists which for the 1st and 2nd persons are almost identical with possessive pronouns. In the 3rd person the form *ni* is used which can be interpreted either as having evolved from the Manchu 3rd person possessive pronoun *ini* or as a Mongolian borrowing. However, while the 3rd person clitic is frequent and the 2nd person singular clitic occurs sporadically, the rest of the forms is rather rare.

Table 3: Possessive enclitics in Sibe

	Singular	Plural
1	mini	moni
2	sini	soni
3	ni	ni

Examples of possessive clitics in Sibe:

- (24) *bo=ni ambu.*
 house=3SG.POSS big
 ‘His house is big.’
- (25) *jaqe-we=sini bierxe!*
 thing-ACC=2SG.POSS collect
 ‘Take your belongings!’

In Sibe only the 3rd person possessive clitic is used as a DEFINITE MARKER, e.g.

- (26) *nane=ni ji-xe na?*
 person=DEF come-PTCP.PFV Q
 ‘Has the person arrived?’

Besides the function of definite marker the Sibe marker *ni* is also used as a kind of TOPIC MARKER, e.g.

- (27) *Tana=ni terang baite icxia-qu.*
 Tana=TOP such matter arrange-NEG
 ‘Tana would not do such things. (As for Tana, she would not do such things.)’

4.3.3.4 The case of ‘head-marked’ possession in Sibe

In Sibe, the Manchu-type marking on the dependent is obligatory when the possessor is referred to by a pronoun, e.g. *sin-i bo* [2SG-GEN house] ‘your house’. In other cases it is used alternatively with simple juxtaposition (e.g. *tašqu sewe* [school teacher] ‘teacher of the school/school teacher’), the latter being more frequent. However, the head of possessive phrases is very often (additionally) marked by the 3rd person possessive clitic *=ni*. In such cases the clitic may be interpreted either as a topic marker (28) or/and as emphasizing definiteness (29), the boundaries between the two meanings being rather vague.

- (28) *min-i age=ni tese-maqe yavu-qu.*
 1SG-GEN brother=TOP 3PL-INS go-NEG
 ‘As for my brother, he does not maintain contacts with them.’
- (29) *honin uju=ni yecin.*
 sheep head=DEF black
 ‘The head of the sheep is black.’

This type of constructions, which has no correspondence in any Manchu variety, is so frequent and remarkable in Sibe that it resembles the head-marked possessive phrases in the non-Jurchenic Tungusic languages. In contrast to the latter, however, the marker *=ni* is always optional in Sibe.

While such type of phrases occurs neither in written Manchu nor in the Manchurian oral varieties, in Mongolian we find structurally similar constructions. Possessive phrases often have additional marking on the head which at the same time implies greater definiteness, e.g.

- (30) Khalkha
Ganaa.g-iin eej=ni emch.
 Ganaa-GEN mother=3SG.POSS doctor
 ‘Ganaa’s mother is a doctor.’

In Mongolian, simple juxtaposition is marginal in expressing possession which makes ‘head-marked’ possessive constructions of the Sibe type rare. However, constructions with similar structure still occur:

- (31) Khalkha
eej bie=ni muu baina.
 mother body=3SG.POSS bad COP
 ‘Mother is sick (literally: Mother her body is bad).’

The existence of possessive clitics in Sibe constitutes a remarkable typological difference from written Manchu. The clitics are formed and used in a way that is almost identical with that of Khorchin. On the first sight, ‘head-marked’ possession does not exist in Mongolian. In fact, however, structurally similar possessive phrases occur in colloquial Mongolian. No such possessive phrases seem to have been attested in any other Manchu variety.

4.3.4 The limiting clitic =*li*

In Sibe, the main means for expressing limitation is the clitic =*li*.³⁰ It can follow any sentence member, e.g.

- (32) *bi=li gene-m.*
 1SG=LIM go-NPST
 ‘Only I will go.’
- (33) *eme=li nane ji-xei.*
 one=LIM person come-PST
 ‘Only one person arrived.’
- (34) *eme nane dudu-r=li orun bi-xei.*
 one person lie-PTCP.IPFV=LIM place be-PST
 ‘There was space for only one person to lie.’
- (35) *uculu-m bana-qv, qaici-m=li.*
 sing-CVB.IPFV be.able-NEG shout-NPST=LIM
 ‘They cannot sing, they only shout.’

In most modern Mongolic languages including Khalkha and Khorchin the clitic *IV* (< Classical Mongolian *la/le*) is used in much the same way, but typically does not determine the predicate, e.g.

- (36) Khalkha
bi=l yav-na.
 1SG=LIM go-NPST
 ‘Only I will go.’

³⁰The Mongolic origin of the Sibe limitation marker was suggested by Norikazu Kogura (2020).

- (37) *neg=l xun ir-sen.*
 one=LIM person come-PST
 ‘Only one person arrived.’

In written Manchu, postpositions such as *-i teile*, e.g. *emu niyalma-i teile* ‘only one person’, are used as means of postnominal³¹ limitation, and no clitic with similar meaning seems to be attested. Likewise, any similar clitic does not seem to be attested in the Manchurian spoken Manchu varieties, wherefore the Sibe clitic *=li* is likely to be a borrowing from a Mongolic language.

4.4 Absence of the Manchu directional (itive and ventive) suffixes *-nV-* and *-nji-*

Written Manchu has a large set of deverbal suffixes, most of which have lost their productivity in the spoken varieties. However, in Sanjiazi, Aihui and Yibuqi two of the deverbal suffixes are highly productive – the suffix *-nji-* ‘to come to do something’ and *-nV-* ‘to go to do something’, e.g. written Manchu *ala-na-ha*, Sanjiazi *ale-na-xe* ‘went to tell’.

In Sibe these suffixes have completely lost their productivity. Instead, multi-verb expressions are used to convey similar meanings, e.g. *ale-me gene-xei* [tell-CVB.IPFV go-PST] ‘went to tell’, or *gene-me ale-xei* [go-CVB.IPFV tell-PST] ‘went and told.’

Mongolian has no directional deverbal suffixes and the meanings ‘go to do’ and ‘come to do’ are expressed by multiverb constructions, e.g. *hele-heer ir-sen* [tell-CVB.PURP come-PST].

Multiverb constructions are frequent and preferred in many languages in the area. A tendency towards replacing deverbal suffixes by multiverb chains in Sibe is not surprising. Perhaps more surprising is the retention of productivity of the deverbal suffix in Manchurian Manchu. Still, however, the different developments may have been prompted by the different language environment.

³¹Besides postnominally used expressions, both Manchu varieties and Mongolian employ adverbs to express limitation. These adverbs (e.g. written Manchu *damu*, Sibe *dame*, Khalkha Mongolian *zövxön*) usually stand in the beginning of a sentence, and always come before the noun which they determine, e.g. written Manchu *damu emu niyalma* ‘only one person’, Mongolian *zövxön neg xün* ‘only one person’. These adverbs are often used together with postnominal limitation as means of emphasis, e.g. written Manchu *damu emu niyalma-i teile* ‘only one person’, Khalkha *zövxön neg l xün* ‘only one person’.

5 Lexical borrowings

In addition to the possibly contact-induced features in Sibe grammar, there is a small-scale but interesting evidence of direct contacts with Mongolic languages in the Sibe lexicon.

The vocabulary of modern spoken Sibe is almost identical with that of written Manchu, the main difference being a larger number of Chinese loanwords. In addition, several Russian, Uyghur and Kazakh loanwords are used. Although colloquial Sibe contains a large amount of Mongolian loanwords, most of them are also found in written Manchu and therefore do not testify to any specific Sibe-Mongolian contacts.³²

Several lexical items such as *kurwo* for ‘bridge’ (written Manchu *doohan*) from Mongolian *xöörög* (written Mongol *kögerge*) ‘bridge’ seem to be restricted to Sibe.

While the modern colloquial language hardly yields any lexical evidence of Sibe-Mongolic contact, in more archaic layers of the lexicon there exist Mongolian loanwords related to Buddhism, shamanism and what may be called “folk religion” which are not found in other Manchu varieties. Some of these terms are still in use while others are only found in written sources.

5.1 Buddhist terminology and the language of Buddhist monks

Historical sources mention the adoption of Tibetan Buddhism by the Sibe during the period of their vassalage to the Khorchins. Until the 1930s a Buddhist monastery existed in Chabchal with approximately forty monks. The language of recitation was Classical Mongolian. The language of the monks contained many Mongolian Buddhist terms for which nowadays Manchu words or Chinese loanwords are used. Examples of such pairs are *sumu* (< written Mongol *süme*) vs. *miao* (< Chinese *miao*) ‘Buddhist temple, monastery’, or *burkan baksi* (< written Mongol *burqan baḡsi*) vs. *fišk* (Manchu *fucihi*³³) ‘Buddha’. However, judging,

³²In general, any search for lexical borrowings is complicated by the nature of Manchu-Mongolian language contacts which involved not only interactions of spoken varieties, but also the sphere of written translations between Manchu and Mongol, which were often done by native speakers of Mongolic varieties. There exist many bilingual texts written in the form of interlinear translations. The Manchu parts of these bilingual texts usually contain a greater portion of Mongol(ic) loanwords than other types of Manchu texts, which are mostly synonyms to original Manchu words or Chinese loanwords. Once used in written documents, these Mongolic loanwords also entered Manchu dictionaries, even though their actual use may have been limited.

³³The Manchu word *fucihi* has been interpreted as a borrowing from Korean by Vovin (2006: 259).

among others, from the recording of recitation of a Buddhist text by a Sibe monk (Zhuangsheng 2018), the local Oirat Mongol tradition of Mongolian recitation preserved among the Öölöds of Ili should also be considered as a possible source of the use of Mongolian in Sibe Buddhist tradition.

5.2 Shamanic terminology

Modern Sibe in Xinjiang consider shamanic traditions to be their ‘original’ religion. In the construction of their ethnic culture, ‘shamanism’ is assigned key importance. Several influential publications give detailed and normative descriptions of the pantheon, system of rituals and main types of ritualists considered to belong to the concept of ‘shamanism’.³⁴ The descriptions were accomplished based on fieldwork among family members of shamans, accounts of eyewitnesses and texts written by shamans since the 19th century. These texts, intended as handbooks for shaman disciples and containing mostly invocation texts with few comments and explanations, are the main source of Mongolic loanwords which seem to be found exclusively in Sibe (cf. Zikmundová 2013).

The so far identified Mongolic loanwords in Sibe shamanic texts are the following:

elci, a ritualist specialized in healing children’s diseases, in particular smallpox < Mongolian *elc(in)* ‘messenger’, in Khorchin ritual practice also a type of ritualist

deoci, a ritualist specialized in exorcist rituals connected to the ‘ghost disease’³⁵ < Khorchin *duuci* a person assigned a role of ‘singer’ in healing the ‘ghost disease’

deole-, to perform the exorcist ritual in healing the ‘ghost disease’ < Khorchin *duul-* ‘to sing’; to assist the exorcist ritual by singing

kuri, a ritual implement in the shape of a building used during the exorcist ritual < Mongol *xüree* ‘circle, temple, monastery’

altan kuri, refrain of a song used during the exorcist ritual in healing the ‘ghost disease’ < Khorchin *altan xüree* ‘golden circle/golden temple’, name of a ritual procedure used during the exorcist ritual in healing the ‘ghost disease’.

³⁴For descriptions of Sibe shamanic traditions see e.g. Sárközi & Somfai-Kara (2013) or Harris (2005).

³⁵‘ghost disease’, Sibe *yivaxen niungku*, Khorchin *ad uvšin*, is a term for a specific type of spirit possession occurring mainly in women (cf. Zikmundová 2013)

All but one of the above Mongolic loanwords pertain to a single type of shamanic ritual – healing a certain type of spirit possession. The ritual was apparently borrowed by Sibe from the eastern Mongols, most probably Khorchins, where it existed in several elaborated variants until the Cultural Revolution. The original Mongolian ritual, known as *andai*, is unique for Khorchins and their immediate neighbors. The Sibe version of the ritual is simplified and shortened.

6 Conclusions: The “reality” of Sibe-Khorchin contacts

For reasons that may be called political, the ethnic history of the Sibe – speakers of a Manchu (Tungusic-Jurchenic) variety – has been a much discussed topic in China. As part of the official narrative, the pre-Qing contacts of the Sibe with Khorchin Mongols are being mentioned – a fact recorded by a few brief notes in historical documents. The Sibe are said to have been vassals of the Khorchins before the 1690s. After 1764, when the ancestors of modern Sibe speakers were moved to Xinjiang, no more contacts between Sibe and Khorchin Mongols took place. The Sibe-Khorchin contact narrative has been used, together with popular views with some background in oral tradition, to argue for a non-Jurchen, possibly Mongol-related origin of the Sibe. It has gradually become part of the self-consciousness of modern Xinjiang Sibe. The question has also triggered academic discussion on this topic.

In this paper I tried to select shared features in Sibe and Khorchin which are not, or marginally, documented in other varieties of spoken Manchu and therefore may testify to a specific contact history. Since no diachronic data for either Sibe or Khorchin are available, modern spoken Sibe and modern Khorchin materials were used. Additionally, lexical data from a written source are mentioned that testify to certain cultural exchange between the Sibe and the Khorchins.

The collected features mostly apply to morphology and one of them, the emphatic prefix *me-* is typical for spoken Sibe and eastern Mongolian. The latter, together with the shared shamanic terminology, and possibly also the shared phonetic features, seem to testify to a direct and lively linguistic and cultural exchange between the Sibe and the ancestors of modern Khorchin. The rest of the mentioned analogies have less clear implications: Being more or less typical for all modern Mongolic languages, they may be features of a linguistic area where multiple Mongolic and Tungusic languages influenced each other.

A short overview of historical facts with connection to the linguistic situation in the Qiqihar region during the Ming is given as a broader context of the documented Sibe-Khorchin contacts. These facts show that the main contact language

of the Sibe was not the language of the Khorchins which arrived from the Mongolian plateau in the mid-16th century but rather the language of the Ujiiyed. The Ujiiyed were a Mongolized Tungusic group whose presence in the Qiqihar region dates back to early Ming, or even Yuan, times. Together with two other groups – the Uriangkhan and the Ongniud – these local Mongols may have already spoken a distinct dialect with “eastern” features when the Khorchins arrived and merged with them. The described shared morphological features and lexical borrowings, however scanty, seem to point towards a Mongolic influence that was stronger and longer-lasting on the ancestor of modern Sibe than on the ancestors of the other spoken Manchu varieties. In this context, another important and rather early Mongolic contact language of the Sibe – the Daur – needs to be examined in the future.

Another question posed in this paper is the significance of the shared linguistic features in imagining Sibe history. The areas around modern Qiqihar and Fuyu, where the Sibe lived, were bordering the homeland of the Hūlun Jurchens who are thought to have spoken a Mongolic-influenced Jurchen variety during the Ming period. The whole area was controlled by the Mongolized Ujiiyed and the Hūlun Jurchens were even referred to as Mongols by other Jurchens. This suggests an image of the Sibe as linguistic representatives of this broader Mongolic-influenced Jurchenic community.

The linguistic developments of Sibe during the Qing period fall out of the scope of this paper. It is, however, important to mention that the period of linguistic diversity during the Ming was effectively ended by the subsequent standardization processes, which, for the Sibe, begun with their incorporation into the Manchu Eight Banners in 1692. The latter affected both Mongolic and Jurchenic languages. Introduction of Buddhism to the Khorchin Mongols, accompanied by spread of literature in general, brought about literacy in Classical Mongolian. For the Jurchenic part, standardization efforts of the Manchu ruling strata is a generally acknowledged fact. Both Literary Mongolian and Classical Manchu enjoyed high prestige. Spread of Classical Mongolian may be one of the factors that brought Khorchin vocabulary and grammar closer to the central Mongolian varieties. The local Jurchenic varieties probably became extinct after the incorporation of the speakers, including Sibe, into the Manchu military units where their spoken varieties were gradually replaced by forms of Standard Manchu.

The question remains whether the described features of Mongolic origin in Sibe may be considered remains of traditional diglossia in a standard Manchu language and an older, Mongolic-influenced Jurchenic variety. Information received from Sibe speakers (e.g. Guo Qing, p.c. August 2009) suggests that in the colloquial language of some elderly speakers Mongolic synonyms to Manchu

lexemes are frequent and some of them seem not to be found in Manchu dictionaries, such as the verb *amere-* ‘to rest, to sleep’ (cf. written Manchu *erge-*, Sibe *erxe-*, Khorchin *amer-* ‘to rest, to sleep’). It is worth mentioning that most of the studies of Sibe were conducted on the basis of material gathered from speakers with high level of literacy in Manchu. Any research of the reported non-standard features has not yet been conducted.

Abbreviations

1SG	1st person singular	LIM	limitation marker
1PL	1st person plural	NEG	negation
2SG	2nd person singular	NMLZ	nominalizer
2PL	2nd person plural	NPST	non-past tense
3SG	3rd person singular	POSS	possessive clitic
3PL	3rd person plural	PN	place name
COM	comitative	PFV	perfective
COP	copula	PRIV	privative
CVB	converb	PROG	progressive
DAT.LOC	dative-locative	PURP	purposive
DEF	definite marker	PST	past tense
EMPH	marker of emphasis	PTC	particle
FOC	focus marker	PTCP	participle
GEN	genitive	TOP	topic marker
IPFV	imperfective	Q	question marker

Acknowledgements

I am indebted to Bai Xiaomei, then a student at the Inner Mongolian National University of Tongliao, for the collection of narratives in the Khorchin Northern Banner in 2015. I am grateful to Andreas Hölzl for the idea on §4.2.1. I further thank Andreas Hölzl, Benjamin Brosig, two anonymous reviewers and Jichang Lulu for their helpful suggestions and comments.

References

- Atwood, Christopher P. 2004. *Encyclopaedia of Mongolia and the Mongol empire*. New York: Facts on File.
- Atwood, Christopher P. 2005. State service, lineage and locality in Hulun Buir. *East Asian History* 30. 5–23.

- Bayančoγtu. 2002. *Qorčin aman ayalyun-u sudulul* [Study of the Khorchin dialect]. Kökeqota: Öbür mongγul-un yeke surγayuli-yin keblel-ün qoriy-a.
- Brosig, Benjamin. 2014a. *Aspect, evidentiality and tense in Mongolian: From Middle Mongol to Khalkha and Khorchin*. Stockholm: Stockholm University.
- Brosig, Benjamin. 2014b. The tense-aspect system of Khorchin Mongolian. In Pirkko Suihkonen & Lindsay Whaley (eds.), *On diversity and complexity of languages spoken in Europe and North and Central Asia* (Studies in Language Companion Series 164), 3–65. Amsterdam: Benjamins.
- Caidengduerji, Sayinjiya. 2014. *Caractéristiques phonologiques, morphologiques et syntaxiques d'un dialecte Mongol du nord-est de la Chine: le Khortchin*. Paris: INALCO. (Doctoral dissertation).
- Chengzhi. 2012. *Shibe shūdan no yijū to sono go no seikatsu* [The migration of the Sibe people and life thereafter], vol. 2 (Chūō yūrashia kankyō shi [Central Eurasian environmental history]). Tokyo: Rinsen shoten.
- Chengzhi. 2021. The “Cisan” incident as an example of “Direct Legal Appeals” in the Qing period. *Studia Orientalia Slovaca* 20(1). 179–215.
- Crossley, Pamela K. 2006. Making Mongols. In Pamela K. Crossley, Helen F. Siu & Donald F. Sutton (eds.), *Empire at the margins: Culture, ethnicity, and frontier in Early Modern China*, 58–82. Berkeley: University of California Press.
- Dixon, Robert M. W. 2010. *Basic linguistic theory, vol. 2: Grammatical topics*. Oxford: Oxford University Press.
- Donjina. 1989. *Donjina i sabuha donjiha ejebun* [A record of things Donjina saw and heard]. Zhijian Yong (ed.). Urumchi: Xinjiang renmin chubanshe.
- Elliott, Mark C. 2001. *The Manchu way: The Eight Banners and ethnic identity in Late Imperial China*. Leiden: Stanford University Press.
- Gorelova, Liliya M. 2002. *Manchu grammar*. Leiden: Brill.
- Harris, Rachel. 2005. *Singing the village: Music, memory and ritual among the Sibe of Xinjiang*. Oxford: Oxford University Press & the British Academy.
- He, Ling & Keli Tong. 1989. *Sibe uksurai an tacin* [Sibe ethnography]. Urumchi: Xinjiang renmin chubanshe.
- Heissig, Walther. 1980. *The religions of Mongolia*. Berkeley: University of California Press.
- Hözl, Andreas. 2017. *A typology of possession in Tungusic*. Paper presented at SLE 50, 2017.09.10-13, Zurich.
- Hözl, Andreas & Yadi Hözl. 2019. A wedding ceremony of the Kyakala in China: Language and ritual. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 16. 87–144.
- Jang, Taeho (Zhang Taihao). 2008. *Xibo yu yufa yanjiu* [Sibe grammar]. Kunming: Yunnan Minzu Chubanshe.

- Janhunen, Juha. 1996. *Manchuria: An ethnic history* (Mémoires de la Société Finno-Ougrienne 222). Helsinki: The Finno-Ugrian Society.
- Janhunen, Juha. 2003a. Khamnigan Mongol. In Juha Janhunen (ed.), *The Mongolic languages*, 83–101. London: Routledge.
- Janhunen, Juha. 2003b. Proto Mongolic. In Juha Janhunen (ed.), *The Mongolic languages*, 1–29. London: Routledge.
- Janhunen, Juha. 2010. Reconstructing the language map of prehistorical North-east Asia. *Studia Orientalia* 108. 281–303.
- Janhunen, Juha. 2012. *Mongolian* (Oriental and African Language Library 19). Amsterdam: Benjamins.
- Kałużyński, Stanisław. 1987. *Die Sprache des mandschurischen Stammes Sibe aus der Gegend von Kuldscha*. Warszawa: Państwowe wydawnictwo naukowe.
- Kogura, Norikazu. 2018. *Shibe-go no modariti no kenkyū [Study of modality in Sibe]*. Tokyo: Tōkyō gaikokugo daigaku ajia afurika gengo bunka kenkyūsho.
- Kogura, Norikazu. 2020. *A brief note on the development of verbal morphology in Manchu*. Paper presented at International Symposium “Endangered languages in Northern Asia”, Freie Universität Berlin, 2020.11.29, Berlin.
- Kradin, Nikolay. 2018. Who was the builder of Mongol towns in Transbaikalia? *Golden Horde Review* 6. 224–237.
- Lattimore, Owen. 1935. *The Mongols of Manchuria*. London: George Allen & Unwin Ltd.
- Mu, Yejun. 1985. Alechuka manyu yuyin jianlun [Short discussion about Alechuka Manchu phonemes]. *Manyu yanjiu [Manchu Studies]* 1985(1). 5–15.
- Mu, Yejun. 1986a. Alechuka manyu de shuci yu gezhuci [Numerals and case endings in Alechuka Manchu]. *Manyu yanjiu [Manchu Studies]* 1986(1). 2–17.
- Mu, Yejun. 1986b. Lalin manyu yuyin gailun [A survey of Lalin Manchu phonetics]. *Manyu yanjiu [Manchu Studies]* 1986(2). 2–30.
- Mu, Yejun. 1987. Balayu [The Bala language]. *Manyu yanjiu [Manchu Studies]* 1987(2). 2–31, 128.
- Mu, Yejun. 1988. Alechukayu yuanyin fasheng de yinbian tedian [Phonetic change in Alechuka Manchu vowels]. *Manyu yanjiu [Manchu Studies]* 1988(2). 1–24.
- Norman, Jerry. 1974. A sketch of Sibe morphology. *Central Asiatic Journal* 18. 159–174.
- Poppe, Nicholas. 2006. *Grammar of written Mongolian*. 5th edn. Wiesbaden: Harrassowitz.
- Porter, David C. 2018. Manchu racial identity on the Qing frontier: Donjina and early twentieth-century Ili. *Modern China* 44(1). 3–34.

- Rybatzki, Volker. 2003. Middle Mongol. In Juha Janhunen (ed.), *The Mongolic languages* (Routledge Language Family Series), 57–82. London: Routledge.
- Sárközi, Ildikó Gyöngyvér. 2019. *From the mists of martyrdom: Sibe heroes and ancestors on the altar of Chinese nation-building*. Berlin: LIT Verlag.
- Sárközi, Ildikó Gyöngyvér & Dávid Somfai-Kara. 2013. Legitimizing the shamanic calling among the Sibe. *Shaman* 21. 67–78.
- Su Deshan. 1984. Guanyu “Jifuxi” yu de tantao – jianlun xiboyu yu manyu de guanxi [A probe into the “Jivš” language – to the discussion about the relationship of Sibe and Manchu]. *Xinjiang shehui kexue* [Xinjiang Social Science] 1984(2). 152–154.
- Svantesson, Jan-Olof. 2003. Khalkha. In Juha Janhunen (ed.), *The Mongolic languages* (Routledge Language Family Series), 154–176. London: Routledge.
- Ujeed, Uranchimeg. 2013. A Mongolian source on the Manchu manipulation of Mongolian. Buddhism in the seventeenth century: The biography of the second Neichi Toyin. *Inner Asia* 15. 225–242.
- Vovin, Alexander. 2006. Why Manchu and Jurchen look so non-Tungusic? In Alessandra Pozzi, Juha Janhunen & Michael Weiers (eds.), *Tumen jalafun jecen akū. Manchu studies in honour of Giovanni Stary*, 255–266. Wiesbaden: Harrassowitz.
- Wang, Qingfeng. 2005. *Manyu yanjiu* [Study of the Manchu language]. Beijing: Minzu chubanshe.
- Wu, Yuanfeng, Anjun & Zhiqiang Zhao. 1985. *Sibe uksura-i šolokon suduri* [Short history of the Sibe nationality]. Urumchi: Xinjiang renmin chubanshe.
- Yamakoshi, Yasuhiro. 2015. Mongoru-go horuchin hōgen tekisuto: nichijō kaiwa o daizai ni shita kihon bunrei-shū [Texts of the Khorchin dialect of Mongolian: A collection of texts focused on everyday conversation]. *Hoppō gengo kenkyū* 5. 281–317.
- Zhao, Jie. 1989. *Xiandai manyu yanjiu* [Study of modern Manchu]. Beijing: Minzu chubanshe.
- Zhuangsheng. 2019. The development of Sibe ethnic awareness: With special consideration of the Sibe people of the Ili River Basin. *Written Monuments of the Orient* 5(1). 50–76.
- Zhuangsheng, (Kicentai). 2018. *The descendants of the Eight Banners and the inheritance of Manchu culture in the 20th century*. Presentation at the 2nd international conference of Sibe studies, 2018.10.04, Prague.
- Zikmundová, Veronika. 2013. Several observations concerning the Sibe practice of the Deoci and Andai rituals of the Khorchin Mongols. *Studia Orientalia Electronica* 113. 217–225.

Name index

- Aalto, Pentti, 112
Abbi, Anvita, 99, 103
Aixinjueluo, Yingsheng, 9, 10, 130, 136
Alonso de la Fuente, José Andrés, 2, 7, 94, 112, 113, 232, 265
An, Jun, 4, 114
Anderson, Gregory D. S., 83, 280
Androsova, Svetlana V., 205
Angina, S. V., 110, 125
Anthony, David W., 101
Aoi, Hayato, 100
Aralova, Natalia, 2, 3, 11, 26, 27, 29, 31–34, 49, 112, 115, 120, 121, 138, 217
Arsen'ev, Vladimir K., 95, 231, 234
Atknine, Victor, 25, 151, 156
Atwood, Christopher P., 301, 306
Avrorin, Valentin A., 113, 122, 124, 126, 232, 236, 245, 250

Balakrishnan, R., 91
Banzhibomi, 157–187, 190–192
Barðdal, Jóhanna, 93
Bayančoytu, 308, 312
Belikov, Vladimir I., 231
Benzing, Johannes, 31, 39, 89, 92, 110–113, 116, 265
Bisang, Walter, 85
Black, Jeremy A., 102
Blagden, C. O., 98
Blust, Robert A., 90

Boldyrev, Boris V., 26, 28, 117, 245
Bouckaert, Remco, 265, 266, 279
Brailovskij, Sergej N., 231, 232
Brosig, Benjamin, 304, 326
Brylkin", A., 112, 114, 123
Bulatova, Nadezhda, 28, 111, 116, 151, 156, 202, 203, 205, 208, 213, 219
Bybee, Joan, 66, 73

Caidengduoerji, Sayinjiya, 305, 307
Castrén, M. Alexander, 13, 111, 150, 154, 156–193
Chaoke, D. O., 8, 9, 111, 117, 119, 120, 124, 126, 128, 136, 157–181, 183–191, 193
Cheng, Mingyuan, 128
Chengzhi, 301, 302
Cincius, Vera I., 7, 25, 26, 111, 157–192, 214, 217, 219, 236, 237, 239, 248, 265, 276
Coler, Matt, 90, 138
Comrie, Bernard, 22, 24, 35
Cotrozzi, Stefano, 111, 116
Coupe, Alexander Robertson, 103
Creissels, Denis, 24
Croft, William, 66, 99
Crossley, Pamela K., 300, 301, 306–308
Cui, Yinqiu, 279, 283
Czerwinski, Patryk, 2, 11, 66, 84, 125, 138, 285

- Dai, Guangyu, 9, 97
Damdinov, D. G., 194
Derevyanko, Anatoliy P., 272
Diessel, Holger, 101
Dixon, Robert M. W., 97, 138, 317
Doerfer, Gerhard, 3, 4, 7, 8, 92, 111,
112, 115, 133, 134, 136, 181–
183, 185–192, 265
Dong, Xingye, 114, 124, 126
Donjina, 302
Dorji, Do, 157–187, 190–192
Drennan, Robert D., 279
Dryer, Matthew S., 101, 120
Duggan, Ana T., 281, 283
Duo, Limei, 8
Dybo, Anna V., 265, 266, 269, 270

Elias, David, 98
Elliott, Mark C., 302
Emeljanov, Alexandr A., 232
Enhebatu, Merden, 9, 113, 127, 128
Epps, Patience, 4
Evans, Nicholas, 63, 64
Everett, D. L., 98

Faehndrich, Burgel R. M., 98
Fillmore, Charles, 92, 106
Fox, Barbara A., 199
Fuller, Dorian Q., 279

Galdanova, G. R., 154
Georg, Stefan, 2, 7, 92, 93, 265
Gil, David, 90, 95, 98, 100, 131
Girfanova, Albina Kh., 236
Golden, Peter et al., 194
Goldsmith, John A., 238, 244
Gorelova, Liliya M., 270, 271, 300,
301, 315
Grenoble, Lenore, 28, 202, 203, 205,
208, 213, 219

Grube, Wilhelm, 112
Gruntov, Ilya A., 184
Gu, Changchun, 10
Guo, Qing, 325

Haenisch, Erich, 134, 194
Haiman, John, 34
Hammarström, Harald, 40
Han, Youfeng, 8, 9, 119
Harris, Rachel, 323
Harrison, David K., 83
Hasibate'er, 8, 116, 117
Haspelmith, Martin, iii, v, 21–24, 35,
40, 43, 44, 47, 57
Hauer, Erich, 127, 157, 158, 160, 162–
173, 175–187, 189–191
Hayashi, Makoto, 207
He, Ling, 302
He, Rongwei, 129
Heissig, Walther, 300
Helimski, Eugene, 272
Hilpert, Martin, 92
Holman, Eric W., 269, 270
Hözl, Andreas, iii, 1–4, 7, 8, 10, 11, 49,
85, 90, 92, 101, 106, 108, 112,
113, 125, 131, 134, 167, 178,
194, 228, 229, 231, 233, 236,
296, 299, 315, 326
Hözl, Yadi, 1, 10, 113, 296, 299
Hough, Carole, 90
Hovav, Malka Rappaport, 24
Huang, Lillian M., 91
Hudson, Mark J., 280

Idiatov, Dmitry, 91, 92, 95, 99, 100,
116, 131, 208, 214, 219
Ikegami, Jirō, 2, 4, 10, 64, 68, 73, 75,
78, 80, 81, 110, 125, 265
Ivanovskiy, A. O., 112, 118, 119, 121

Name index

- Jang, Taeho, 2, 9, 31, 296, 299
Janhunen, Juha, 1–4, 7, 8, 10, 13, 25,
92, 93, 111, 129, 150–194, 256,
265, 269–273, 278, 296, 299,
301, 303, 305, 309, 310, 314,
316, 317
Jeong, Choongwon, 283
Jin, Ning, 106, 107
Jin, Qizong, 9
Johanson, Lars, 280
Johnson, Allen, 94
Johnson, Mark, 108

Kajia, 8, 9, 119
Kalina, 8, 9, 120
Kałużyński, Stanisław, 128, 302
Kandakova, Galina I., 49
Kane, Daniel, 10, 111, 130
Kapišovská, Veronika, 194
Kara, György 1990, 194
Kaufman, Terence, 280
Kazama, Shinjirō, 2–4, 7, 8, 31, 111,
112, 114, 119, 131, 133
Kazarova, Antonina V., 49
Kazenin, Konstantin I., 31
Keevallik, Leelo, 216
Kern, B., 98
Khabtagaeva, Bayarma, 2–4, 8, 9, 13,
111, 113, 149–151, 154, 156,
177, 178, 181–183, 185–193
Khasanova, Marina, 25, 112
Kialundziuga, Valentina, 236, 258
Kim, Juwon, 9, 113, 115, 129
King, J. R. P., 105
Kittilä, Seppo, 22, 23, 31
Kiyose, Gisaburo N., 9, 11, 111, 112, 130
Klamer, Marian, 103
Kluge, Angela, 101
Klyachko, Elena, 3, 285

Knüppel, Michael, 8, 111, 112, 115
Ko, Dongho, 124, 125
Kobayashi, Masato, 102
Kogura, Norikazu, 296, 320
Koile, Ezequiel, 285
Konstantinova, Olga A., 28, 116, 203,
205, 219
Koontz-Garboden, Andrew, 22
Kopecky, Felix, v
Kormušin, Igor V., 227, 232, 233, 236,
242, 251, 265
Korovina, Evgeniya, 265, 266, 269,
270, 272, 273, 278
Kradin, Nikolay, 305
Kuznetsova, Natalia, 259

Lahiri, Aditi, 24
Lakoff, George, 108
Langacker, Ronald W., 92, 110
LaPolla, Randy, 98
Lar'kin, Viktor G., 232, 233
Lattimore, Owen, 301
Lavrillier, Alexandra, 94
Lebedeva, Elena P., 113, 122, 232, 236,
245, 250
Lefebvre, Claire, 92
Leipe, Christian, 276, 279
Leontovič, Sergej, 232
Lessing, Ferdinand D., 194
Levin, Beth, 22, 24
Levshina, Natalia, 34
Li, Fengxiang, 29, 31, 200
Li, Guojun, 10
Li, Paul Jen-kuei, 99
Li, Shulan, 98
Li, Tao, 276, 279, 283
Lichtenberk, Frantisek, 91
Lie, Hiu, 111, 117
Lock, Arnold, 102, 138

- Lopatin, Ivan A., 122
- Ma, Mingchao, 9, 10
- Ma, Wenye, 10
- Maak", R., 116
- Magata, Hisaharu, 68
- Majewicz, Alfred F., 68, 112
- Makarova, K. I., 116
- Malchukov, Andrej L., 32, 64–67, 69,
73, 77, 83–85
- Mallory, James, 279
- Malyarchuk, Boris, 281, 283
- Margaritov, Vasiliy P., 231, 232
- Maslinskij, Kirill, 259
- Matić, Dejan, 218
- Mazo, Olga M., 184
- McMahon, April, 4
- Meng, Huiying, 10
- Meng, Shuxian, 8, 9, 119
- Menges, Karl Heinrich, 265, 272
- Merlo, Paola, 23
- Miller, Matthew, 285
- Miyamoto, Kazuo, 279
- Morozova, Olga N., 205
- Mostaert, Antoine, 194
- Moutu, Andrew, 107–109
- Mu'ercha, Yiling'a, 10
- Mu, Yejun, 9–11, 112–114, 127, 130,
133, 299
- Mühlhäusler, Peter, 91, 101
- Munshi, Sadaf, 97, 138
- Mushin, Ilana, 90
- Myl'nikova, Klavdija M., 25
- Myreeva, Anna N., 26, 28
- Nadarov, Ivan P., 231, 232, 236, 243
- Nagaraja, K. S., 103
- Najia, 8, 9
- Nakanome, Akira, 68, 112, 125
- Nau, Nicole, 99
- Nedjalkov, Igor, 28–33, 136, 200, 203
- Nelson, Sarah, 279
- Nichols, Johanna, 22, 24, 38, 42, 44,
47
- Nikolaeva, Irina, 111, 113, 114, 122, 133,
214, 229, 236–238, 244, 246,
256
- Ning, Chao, 283
- Nordhoff, Sebastian, v
- Norman, Jerry, 111, 181, 296
- Oskolskaya, Sofia, 2–4, 7, 13, 49, 64,
265, 266, 268–272, 274, 280,
284
- Östman, Jan-Ola, 92
- Ozolinja, Larisa V., 68, 125
- Pakendorf, Brigitte, 2, 3, 11, 26, 27, 29,
31–34, 217, 280
- Pallas, Peter Simon, 111, 112
- Palmer, Frank R., 31
- Paperno, Denis, 104, 138
- Payne, Doris L., 102
- Payne, Thomas E., iii, 9, 31, 49, 94,
102, 138
- Perekhval'skaya, Elena, 3, 10, 11, 49,
122, 138, 228, 229, 231, 234,
257, 268
- Peterson, Christian E., 279
- Petrova, Taisija I., 68, 82
- Pevnov, Alexander M., 7, 25, 30, 33,
64, 78, 82, 112, 229, 269–273
- Plank, Frans, 24
- Podlesskaya, Vera I., 204, 213
- Poppe, Nicholas, 194, 219, 314
- Porter, David C., 302
- Prins, Maria Clazina, 106, 138
- Protodjakonov, Prokopij, 231, 232

Name index

- Przevalskij, Nikolaj M., 231
Pugach, Irina, 272, 281
- Qin, Ling, 279
- Radčenko, Galina L., 253, 256
Rinčen, Byambin, 194
Rišes, Ljubov' D., 26
Robbeets, Martine, 2, 3, 7, 66, 84, 265,
269–272, 274, 277–279, 281–
283, 285, 287
Robbek, Maria E., 26, 218
Robbek, Vasiliy A., 26, 218
Rozycki, William, 183, 185–187, 189–
192
Rybatzki, Volker, 315
- Sa, Xirong, 118
Samardžić, Tanja, 23
Sameng, Yierhanzhi, 127
Sandman, Erika, 98
Sárközi, Ildikó Gyöngyvér, 298–300,
302, 323
Sarvasy, Hannah, 97
Savelyev, Alexander, 265
Schäfer, Florian, 22
Schebesta, Pater P., 98
Schiefner, Anton, 116
Schluessel, Eric, 129
Schmidt, Peter, 111, 112, 121, 122, 124,
232, 236
Schulze, Wolfgang, 95, 138
Seifart, Frank, 44, 46
Sem, Lidija I., 113, 114, 123
Senft, Gunter, 101
Sergusheva, Elena A., 270, 276, 279
Shirokogoroff, Sergei Mikhailovich,
9
Simonov, Mikhail D., 234, 236–238,
243, 245, 246, 248, 256–258
- Sirenbatu, 8, 136
Siska, Veronika, 281, 283
Šneider, Evgenij R., 232, 236–238,
242, 248, 256, 258
Sohn, Ho-Min, 102
Somfai-Kara, Dávid, 323
Song, Jae Jung, 102
Šrenk, Leopold I., 231
Starikovskaya, Elena B., 281
Starostin, Sergei, 276
Startsev, Alexandr F., 233
Stary, Giovanni, 9, 157, 160, 161, 171,
181, 185, 189
Stevens, Chris J., 279
Su Deshan, 303
Sukernik, Rem I., 281, 283
Suliandziga, Rodion V., 230
Sundueva, E. V., 194
Sunik, Orest, 31, 37, 39, 113, 114, 123,
235, 236, 258, 265
Svantesson, Jan-Olof, 317
- Tal'ko-Grynčevič, Yu. D., 154
Thomason, Sarah G., 280
Tirkey, Bablu, 102
Titov, E., 154
Tolskaya, Inna, 221
Tolskaya, Maria, 111, 113, 114, 122, 133,
214, 221, 229, 236, 238
Tong, Keli, 302
Tournadre, Nicolas, 98
Trousedale, Graeme, 92
Trudgill, Peter, 229, 234
Tsuchida, Shigeru, 99
Tsumagari, Toshirō, 8, 68, 73, 111, 117,
119, 121, 122, 200, 246, 257
Tugolukov, V. A., 152, 153, 272
Tumurtoogo, Domiin, 194

- Ujeed, Uranchimeg, 300
Uray-Kóhalmi, Katalin, 194
Uvarova, T. B., 152
- Van Coetsem, Frans, 280
Van de Velde, Mark, 90, 107
Van Langendonck, Willy, 90, 107
Vasilevič, Glafira M., 151, 154
Vasilevich, Glafira M., 154, 157–186,
188–192, 200, 201, 265, 271,
272
Venukoff, M., 112, 114
von Goethe, Johann Wolfgang, 90
von Klaproth, Julius, 111
Vostretsov, Yury E., 270, 276, 279
Vovin, Alexander, 7, 129, 131, 265, 322
- Wang, Chuan-Chao, 272, 274, 279,
281–283
Wang, Jing, 9
Wang, Qingfeng, 9, 127, 299, 314, 315
Weng, Jianmin, 8
Whaley, Lindsay, 2, 7, 9, 29, 31, 64,
200, 265
Wichmann, Søren, 272, 274
Winford, Donald, 280
Witsen, Nicolaas, 111
Wu, Yuanfeng, 302
Wuge, Shouping, 128
Wurigexiletu, 8
Wurm, Stephen A., 91, 101
Wylie, Alexander, 128
- Yamada, Yoshiko, 64, 67, 68, 71, 75, 82
Yamakoshi, Yasuhiro, 304
Yan, Shi, 283
Yanušević, Zoya, 280
Yoon, Kyung-Eun, 207
Yu, Wonsoo, 117
- Yurn, Gyudong, 124, 125
- Zgusta, Richard, 280
Zhang, Yanchang, 118, 133
Zhao, Jie, 9, 127, 299, 315
Zheng, Zhonghua, 9
Zhuangsheng, 298–303, 323
Zikmundová, Veronika, 2, 9, 10, 128,
133, 136, 138, 157, 158, 160–
169, 171–173, 175–184, 186,
187, 191, 194, 296, 323
Zinder, Lev R., 237, 251
Zúñiga, Fernando, 22, 23, 31

Language index

- Abau, 101, 102
Afroasiatic, 98
Ainu, 41, 278
Alchuka, 9–11, 92, 113, 130, 132, 133,
135
Alchukaic, 10
Altaic, 150, 278
Amdo-Tibetan, 41
Amuric, 3, 272, 278
Amuric group of Tungusic, 2, 151
Arabic, 41, 171, 184
Arman, 8, 92, 110, 112, 115, 121, 132,
135
Armenian, 41
Athabaskan, 278
Austroasiatic, 98
Austronesian, 90, 91, 99, 101
Avar, 41
Aymara, 90
Aymaran, 90
Azerbaijani, 41, 42

Baarin, 305, 306, 308
Badaga, 90, 91
Bala, 9–11, 93, 110, 111, 113, 114, 130,
132, 135
Balaic, 10
Bannermen Manchu, 10, 11, 299, 300
Batsbi, 41
Beng, 104
Bulgar Turkic, 188
Burmese, 41

Burushaski, 40, 41, 97
Buryat, 149, 150, 152, 177, 178, 181–
183, 185–193, 308, 316

Central Asian Turkic, 42
Central Eastern Tungusic, 3
Central Western Tungusic, 3
Chapacuran, 98
Chimakuan, 278
Chinese, 1, 3, 8, 10, 94, 110, 113, 119,
126, 130, 131, 152, 153, 232,
256, 257, 269, 270, 279, 281,
282, 302, 305–308, 310, 322
Chinese Kyakala, 10, 11
Chukotko-Kamchatkan, 66, 278
Chuvash, 11
Common, *see* Evenki
Common Tungusic, 157
Czech, 1

Dagur, 117, 119, 153, 154, 177, 178, 181–
183, 185–193, 297, 301, 302,
304, 308, 325
Daur, *see* Dagur
Domaaki, 40, 41
Dravidian, 91, 102
Dula'er, 8, 9

early Mongolic, 149, 325
Elunchun, *see* Oroqen
English, 1, 2, 13, 22, 23, 41, 90–92, 96,
98, 100, 101, 104

Language index

- Eskimo-Aleut, 66
- Even, iii, 2, 3, 7, 8, 11, 12, 24–28, 30–32, 35–41, 43, 44, 46–48, 57, 58, 64, 65, 93, 110–112, 114–118, 132, 134–136, 150, 218, 222, 231, 264, 265, 269, 274, 277, 280, 282, 283, 286, 304, 307
- Beryozovka, 115, 218
- Bystraja, 25, 27, 30, 31, 33, 40, 41
- East, 115, 218
- Lamunkhin, 25, 27, 28, 30–33, 40, 41, 46, 115
- Standard, 25, 40, 151
- Evenic, *see* Ewenic
- Evenki, iii, 3, 8, 9, 11–13, 24–28, 30, 32, 35–41, 48, 64, 93–95, 111, 112, 114, 116–119, 121, 132, 135, 136, 150–152, 154–156, 200, 203–207, 212–214, 217, 218, 220, 222, 229–231, 235, 239, 240, 242, 247, 248, 264, 265, 269, 271, 274, 280, 282, 283, 286
- Aldan, 159–162, 164–170, 172, 174–177, 179–183, 186, 188, 191, 192
- Aoluguya, 8, 9, 116, 117, 136, 200
- Ayan, 159–162, 168, 170, 172–177, 180, 181, 185, 191
- Barguzin, 150, 158–162, 166–168, 170, 172–177, 180, 181, 183, 185, 186, 188–192
- Baunt, 150, 154, 172, 183
- Chulman, 157
- Chumikan, 159–163, 166–168, 170–177, 180, 182, 183, 186, 189, 214, 217, 218
- Eastern, 116, 151, 200, 203, 214, 217, 271
- Ilimpeya, 158–162, 164–170, 172–177, 179, 180, 182, 183, 191, 214, 217, 221
- Khamnigan, 3, 8, 13, 111, 119, 122, 134, 150, 152–194, 200
- Khingán, 170, 183, 188
- Mankovo, 157, 160, 166, 174, 176, 192
- May, 134, 157–160, 180
- Morigele, 8, 9
- Nepa, 159–163, 165–169, 171, 172, 174–177, 179, 181–183
- Nercha, 3, 13, 111, 150, 152–154, 156–193
- North-Baikal, 150, 159, 161–164, 166–168, 170, 172–177, 179–181, 183, 188
- North-Baikal Evenki
- Tungir, 170
- Northern, 151, 200, 203, 214, 217, 221, 271
- Podkamennyj, 157–162, 165–177, 179–183, 189, 191
- Poligus, 221
- Sakhalin, 64, 111, 116, 157, 159–163, 166–177, 179–183, 185, 186, 190, 191, 214, 281
- Siberian, 151, 157–192, *see* Evenki
- Siberian Common, *see* Evenki
- Southern, 151, 200, 203, 214, 216, 217, 219, 221, 271
- Sym, 159, 161, 162, 166, 168–170, 172–174, 176, 177, 179, 181, 214, 221, 222
- Tokko, 157–159, 183

- Tokma, 162, 163, 165–170, 172, 174–177, 179, 180, 182
Tokmin, 160, 162
Tommot, 157, 158, 160, 180, 183
Tungir, 159–163, 165–177, 179–181, 183, 191, 192
Tunguska, 151, 214, 217
Uchir, 165, 166, 181, 182
Uchur, 157–163, 166–170, 172–177, 179, 180, 182, 183, 185, 186, 188, 190–192
Upper Lena, 159, 162, 167, 172–177, 179, 182, 183, 185, 186, 189–192
Urmi, 157, 159–163, 166–170, 172–177, 179, 181–183, 185, 186, 189–191
Urulga, 154, 157, 158, 162, 166–170, 174–176, 181, 187, 188, 191, 192
Urulyungui, 8, 153–155, 157, 158, 170, 174–183, 187–189, 192
Vanavara, 219
Vilyuy, 161
Vitim, 162, 172, 175, 182, 185, 190
Yerbogachyon, 158–162, 165–177, 179, 181–183, 191, 208, 217, 219, 221, *see* Yerbogochon
Zeya, 158–162, 164–170, 172–177, 179–183, 185, 188, 189, 191, 192
- Ewen, *see* Even
Ewenic, 2–4, 6–9, 11–13, 92, 101, 112, 114, 115, 117, 119–123, 125–127, 131, 136, 265
Ewenke, 8
Ewenki, *see* Evenki
- Finnish, 41, 42, 101
French, 1, 41, 93, 97, 100
- Georgian, 41
German, 1, 41, 95, 96, 101, 103–105, 184
Germanic, 2, 101
Gilyak, *see* Nivkh
Gold, *see* Nanai
Great Andamanese, 99, 103
Greek, 41
Gyalrong, 41, 105, 106
- Hebrew, 41
Hezhe, 3, 264, 268, 275
Hezhen, 4, 281
Hindi-Urdu, 41
Hungarian, 1, 41, 42
- Iatmul, 108
Icelandic, 41
Indo-European, 2, 42, 91, 97, 100, 101, 104, 105
Italian, 1, 93, 100, 101
Italic, 2
- Jahai, 98
Japanese, 1, 31, 41, 42, 101
Japonic, 42, 47, 100
Jibsi language, 304
Jinghpaw, 41
Jungarian Chakhar, 297
Jurchen, 2, 3, 9–11, 64, 93, 110, 111, 130, 132, 135, 150, 264–266, 268, 270, 271, 277, 286, 296, 298, 300, 314, 324, 325
Jurchenic, 2–4, 6–14, 92, 110, 112–114, 117, 120, 126–131, 133, 134, 136, 137, 296, 297, 299, 301, 308, 319, 324, 325

Language index

- Karlong Mongghul, 98
Kashmiri, 41
Kazakh, 41, 322
Kekar, 232, 233, 235, *see* Russian
 Kyakala
Khakas, 41, 42
Khalaj, 11
Khalkha, 188, 191, 298, 304, 305, 309,
 310, 314, 316, 319, 320
Khamnigan Mongol, 150, 154–156,
 177, 178, 181–183, 185–194
Kharachin, 300, 305, 306
Khitano, 306
Khitano-Mongolic, 98, 117, 119, 131
Khor, 122
Khorchin, 14, 296–298, 300, 301, 304–
 310, 312–318, 320, 323–326
Kilen, 3, 4, 11, 92, 113, 114, 118, 123, 124,
 126, 130, 132, 133, 135, 229,
 247, 248
Kili, 4, 93, 110, 113, 114, 123, 132, 135,
 264, 268, 277
Kilivila, 101
Kita Akita Japanese, 41
Korean, 1, 31, 41, 102, 105, 127, 306,
 308
Koreanic, 7, 102, 105, 129, 131
Koryak, 41
Kurux, 101, 102
Kyakala, 9, 11, 93, 110, 113, 132, 135,
 234, 235, 241, 249, *see* Rus-
 sian Kyakala
Kyrgyz, 41

Lamut, *see* Even
languages of Europe, 24, 42, 91, 100,
 101, 120, 131
languages of North America, 38, 42
languages of Papunesia, 44, 47

languages of Siberia, 38
languages of South America, 44, 47
languages of the Caucasus, 42
Latin, 1, 184
Lezgian, 41
Literary Mongolian, 177, 178, 181–
 194, 299, 325
Lithuanian, 41

Macro-Altaic, 66, 83, 84
Maltese, 41, 42
Manam, 91
Manchu, 1–3, 9–11, 14, 39, 41, 42,
 49, 64, 93, 111–114, 119, 124,
 126–136, 150, 151, 157–187,
 189–191, 229, 232, 239, 240,
 242, 248, 264–266, 268–271,
 274, 277, 284, 286, 296–304,
 308–310, 314–316, 318–322,
 324–326
 Aihui, 9, 10, 113, 127, 299, 309,
 314, 315, 321
 Jing, 10, 110, 112, 130, 299
 Lalin, 9, 10, 110, 112, 130, 299
 Sanjiazi, 9, 10, 97, 113, 127–129,
 136, 298, 299, 309, 314–316,
 321
 Yanbian, 9, 10, 127
 Yibuqi, 9, 10, 113, 127, 299, 309,
 314, 315, 321
Manchu Kyakala, *see* Chinese
 Kyakala
Manchuic, 10
Manchuric, 3, 93, 263, 265, 268, 269,
 280, 285, *see* Jurchenic
Mandarin, 41, 96, 97, 100, 103–105,
 299, 308
Mande, 104
Manegir, 118, 121

- Marathi, 41
Matsigenka, 94
Meche, 41
Middle Mongol, 177, 178, 181–193, 315
Mongolian, 31, 41, 129, 131, 133,
134, 185–191, 194, 296, 298–
300, 302, 304–306, 308–310,
313–325
Mongolic, 7, 13, 14, 110, 117, 118, 120,
129, 149, 150, 153, 155, 156,
171, 177, 178, 181–193, 272,
283, 296–298, 300–303, 305,
308, 314–316, 320–325
Mongsen Ao, 102, 103
Nanai, 2–4, 11, 39–41, 49, 66, 93, 111–
114, 124–126, 130, 132, 135,
151, 157–187, 189–191, 229–
231, 239, 240, 248, 264, 268,
269, 274, 275, 277, 280, 281,
283, 286
Ussuri, 4, 93, 111, 113, 114, 123,
124, 132, 135, 229, 264, 268,
275, 286
Nanaic, 2–4, 6, 7, 10–12, 64, 66, 77, 83,
92, 112, 117, 120, 122, 125, 126,
131, 264, 265, 268, 269
Negidal, iii, 3, 8, 11, 12, 22–41, 43, 44,
46–49, 57, 58, 93, 119–121,
125, 126, 132, 135, 150, 157–
187, 189, 190, 217, 218, 222,
239, 240, 247, 248, 264, 265,
274, 277, 280, 281, 283, 286
Lower Negidal, 25, 112, 119
Upper Negidal, 25, 26, 112, 120,
121
Nehe, 8, 9
Nepali, 41
Ngaju Dayak, 90, 91
Nihali, 102, 103
Nivkh, 3, 14, 64, 66, 125, 231, 276, 278,
281, 283
Northeastern Tungusic, 3
Northern Tungusic, 2, 11, 24, 25, 27–
29, 34, 38–42, 47, 48, 64, 65,
92, 101, 112, 114, 123, 130, 150,
205, 218, 222, 268, 269
Northwestern Tungusic, 3
Nungon, 97
Oirat, 316, 323
Old Turkic, 177, 182, 185, 187, 188, 191
Öölöd, 297, 302
Oroch, iii, 2, 3, 11, 92, 112, 113, 121,
122, 125, 132, 135, 151, 157–
187, 189–191, 229–237, 239,
240, 242, 246–251, 257, 264,
265, 268, 269, 274, 277, 280,
286
Koppi, 232, 234, 235, 237, 241,
242, 246, 249–251, 257
Tumnin, 230, 234–236
Xadi, 232, 234–236, 241, 249
Orochen, *see* Oroqen
Orochic, 2, 3, 93
Orochon, 231
Orok, 264, 268, 269, 274, 277, 278,
286, *see* Uilta
Oroqen, 8, 9, 93, 112, 117–120, 122,
126, 130–132, 135, 136, 156–
181, 183, 184, 186, 188, 191,
193, 200, 264, 274, 278, 280,
281, 283
Gankui, 118
Nanmu, 120
Shengli, 119
Xunke, 118
Paleosiberian, 66, 84, 278, 279

Language index

- Pama-Nyungan, 90
Pan-Tungusic, 12
Panare, 94, 101, 102
Papuan Malay, 101
Para-Mongolic, 7
Pazih, 98, 99
Persian, 101
Polish, 2, 91
Portuguese, 93
Proto-Amuric, 276, 278–280, 283, 284
Proto-Tungusic, 12–14, 94, 110, 112, 131, 133, 134, 136, 137, 264, 269–272, 274–276, 278–280, 283–287
Punjabi, 41
Romance, 93
Romanian, 22, 41, 43, 93
Russian, 2, 3, 13, 22, 26, 41, 48, 94, 95, 100, 101, 119–121, 123, 124, 126, 131, 152–154, 156, 180, 184, 193, 200, 206, 211, 212, 220, 223, 227, 231, 235, 258, 259, 270, 276, 279, 280, 283, 284, 322
Russian Kyakala, 10
Sakha, 41, 42, 48
Salishan, 278
Samar, 93, 112, 124, 126, 132, 135
Samoyedic, 272
Shuri Okinawan, 41, 42
Sibe, iii, 9, 10, 14, 93, 98, 106–108, 113, 126–128, 132, 133, 135, 151, 157–187, 189–191, 264–266, 268, 269, 271, 274, 277, 281, 286, 296–304, 307–312, 314–316, 318–326
Sinitic, 7
Sino-Tibetan, 97, 98, 103–106
Slavic, 2
Slovene, 41
Solon, 8, 9, 13, 93, 112, 117–120, 125, 130, 132, 135, 136, 151, 152, 155–193, 200, 247, 264, 265, 269, 275, 277, 280, 286, 301, 302, *see* Solon
Arong, 8, 9, 120
Hailar, 151, 155
Huihe, 8, 9, 112, 117, 118
Nonni, 275
Ongkor, 112
Southeastern Tungusic, 66
Southern Tungusic, 3, 7, 64, 92, 112, 114, 131, 150, 157–187, 189, 190, 205, 218, 222
Spanish, 93, 94
sub-Saharan African languages, 24
Sumerian, 102
Swedish, 41
Tajik, 41
Tarama Miyako, 100
Tazy, 231
Teiwa, 102, 103
Thai, 41
Tiddim Chin, 41
Tigre, 98
Tocharian, 111, 188
Tok Pisin, 91, 101
Tokharian, *see* Tocharian
Trans-New Guinea, 97, 103
Transeurasian, *see* Altaic
Tungus, *see* Evenki, Khamnigan, *see* Evenki
Turkic, 11, 42, 47, 103, 150, 171, 177, 182, 185, 187, 191, 279, 282,

- 283
- Turkish, 22, 41, 42, 102, 103
- Turkmen, 41
- Tzez, 41
- Udeghe, *see* Udihe
- Udegheic, 2–4, 6, 7, 10–13, 64, 66, 77, 83, 92, 112, 117, 120–123, 125, 126, 131, 265
- Udehe, 286, *see* Udihe
- Udeheic, *see* Udegheic
- Udihe, iii, 2, 3, 11, 13, 39–42, 49, 66, 93, 111, 113, 114, 121–123, 125, 126, 131–133, 135, 151, 157–184, 186, 187, 189, 190, 214, 221, 222, 227–243, 246–249, 252, 256–259, 268, 269, 274, 277, 280, 283
- Anjuj, 228, 230, 233–235, 241, 242, 249
- Bikin, 112, 122, 228–230, 233–235, 237, 238, 241, 245–247, 249, 251–253, 256, 257
- Iman, 228, 233–235, 237, 238, 241, 246, 249, 257
- Kur-Evenki
- Urmi, 228, 229, 234, 235
- Urmi Nanai, *see* Kili
- Samarga, 122, 228, 230, 233–235, 237, 246
- Xor, 228, 233–235, 237, 238, 241–247, 249, 251–253, 256
- Xungari, 228, 230, 234, 235, 242
- Udmurt, 41, 42
- Uilta, iii, 10–12, 63, 64, 66, 67, 69–77, 79–85, 93, 100, 111, 112, 114, 125, 126, 131, 132, 134, 135, 214, 222
- Northern Uilta, 64, 67, 77–79, 82–84, 125
- Southern Uilta, 64, 67, 75, 77, 78, 82, 125
- Ulch, *see* Ulcha
- Ulcha, 10, 93, 110, 112, 113, 124, 125, 132, 135, 151, 157–187, 189, 190, 229, 233, 248, 264, 268, 269, 274, 277, 281, 283, 286
- Ulchaic, 10
- Uralic, 66, 84, 101
- Uyghur, 282, 322
- Uzbek, 41
- Vietnamese, 41
- Viljuj, 116
- Wakashan, 278
- Wari', 98
- written Manchu, 10, 11, 93, 110, 111, 128, 130, 266, 298, 299, 302, 304, 309, 310, 314–316, 319–322, 326
- written Mongol, 309, 314, 322
- written Sibe, 10
- Wulai Atayal, 91
- Wutun, 98
- Xianbei, 303
- Xibe, *see* Sibe
- Yakut, *see* Sakha
- Yankunytjatjara, 90
- Yeniseian, 66, 272, 278
- Yukaghir, 66

Tungusic languages

Tungusic is an endangered language family that encompasses approximately twenty languages located in Siberia and northern China. These languages are distributed over an enormous area that ranges from the Yenisey River and Xinjiang in the west to the Kamchatka Peninsula and Sakhalin in the east. They extend as far north as the Taimyr Peninsula and, for a brief period, could even be found in parts of Central and Southern China.

This book is an attempt to bring researchers from different backgrounds together to provide an open-access publication in English that is freely available to all scholars in the field. The contributions cover all subbranches of Tungusic and a wide range of linguistic features. Topics include synchronic descriptions, typological comparisons, dialectology, language contact, and diachronic reconstruction. Some of the contributions are based on first-hand data collected during fieldwork, in some cases from the last speakers of a given language.