Universiteit
Leiden
The Netherlands

A history of East Baltic through language contact

Jakob, A.

Citation
Jakob, A. (2023, July 6). *A history of East Baltic through language contact*. Retrieved from https://hdl.handle.net/1887/3630131

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3630131

Note: To cite this publication please use the final published version (if applicable).

a History of East Baltic through Language Contact

Anthony Jakob

Cover illustration: Trackway across Männikjärv bog in Jõgeva County, Estonia (photograph: author's own work)

Copyright © 2023: Anthony Jakob. All rights reserved.

A History of East Baltic through Language Contact

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 6 juli 2023
klokke 11.15 uur

door

Anthony Misha Jakob

geboren te Londen (Verenigd Koninkrijk)

in 1992

Promotores

Prof.dr. G.J. Kroonen

Prof.dr. A.M. Lubotsky

Promotiecommissie

Dr. P.J. Kallio (Helsingin yliopisto)

Prof.dr. C.C. Levelt

Dr. T.C. Pronk

Prof.dr. P.C.H. Schrijver (Universiteit Utrecht)

This research was carried out with funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 716732).

Contents

List of tables 7

Acknowledgements 9

Symbols and abbreviations 13

Data sources and conventions 19

Introduction 25

Contacts with known languages 31

1.1. Baltic–Slavic contacts

1.1.1. Early Slavic → Baltic loans 33

1.1.2. Early Baltic → Slavic loans? 66

1.2. Early Germanic → Baltic loans 79

1.3. Baltic → Finnic borrowings

1.3.1. Preliminaries 91

1.3.2. Baltic loanwords with an IE etymology 100

1.3.3. Analysis of sound substitutions 116

1.3.4. Loans from Proto-Finnic to Proto-Baltic? 150

1.3.5. Common loans from unknown sources? 160

1.3.6. Analysis of contact relationship 193

1.4. Loanwords into other Uralic languages

1.4.1. Sámi 201

1.4.2. Mordvin 217

1.4.3. Mari 233

1.4.4. Permic 240

1.4.5. Conclusion 243

Contacts with unknown languages 247

2.1. Introduction

2.1.1. Research history 249

2.1.2. Methodological considerations 258

2.1.3. Excursus: illegal root structures 270

2.2. Consonantism

2.2.1. 'Nasalization', *-*VNT*- ∞ *-*VT*- 279

2.2.2. Voicing alternations 298

2.2.3. Sibilant clusters 343

2.2.4. Other irregularities 355

2.3. Vocalism

2.3.1. Initial vowels 367

2.3.2. Alternations between front and back vowels 384

2.3.3. Alternations between low and high vowels 402

2.3.4. Alternations between monophthongs and diphthongs 413

2.3.5. Length alternations 419

2.3.6. IE **a* 436

2.4. Analysis

2.4.1. Semantics 441

2.4.2. Stratification 467

Conclusion 477

Bibliography

Abbreviated titles 481

List by author 488

Word index

Baltic languages 563

Slavic languages 572

Germanic languages 580

Other Indo-European languages 584

Uralic languages 589

Other languages 599

Samenvatting 601

Curriculum vitae 607

List of tables

Table 1. Words similar in Lithuanian and Vedic Sanskrit 29
Table 2. Archaic features in Slavic loanwords 57
Table 3. *i*-diphthongs in Early Proto-Finnic 124
Table 4. Probable shared substrate words in Baltic and Finnic 191
Table 5. Baltic loanwords in Sámi and Mordvin 244
Table 6. Suggested criteria for identifying substrate borrowings 257
Table 7. Comparison of *deus* and θεός 263
Table 8. Possible examples of nasal alternations 297
Table 9. Possible examples of $*g^{(h)} \infty *k$ alternations 317
Table 10. Possible examples of alternations involving dentals 327
Table 11. Possible examples of labial alternations 336
Table 12. Possible examples of an alternation $*ž \infty *š$ 342
Table 13. Possible examples of sibilant metathesis 352
Table 14. Possible examples of 'prefixation' 382
Table 15. Possible examples of front ∞ back alternations 401
Table 16. Possible examples of high ∞ low alternations 411
Table 17. Possible examples of long ∞ short alternations 430
Table 18. Distribution of borrowed animal names 445
Table 19. Distribution of borrowed plant names 449
Table 20. Distribution of borrowed agricultural terms 454
Table 21. Chronology of cereal terms in Baltic 458
Table 22. Distribution of borrowed apicultural terms 464
Table 23. Distribution of borrowed terms for structures 467
Table 24. Distribution of borrowed metallurgical terms 468
Table 25. Alternations showing a geographical patterning 470
Table 26. The **VND* substrate 473

Acknowledgements

The first seed for this dissertation was planted back in 2017 when I told Tijmen Pronk I'd be curious to study the whole subject of "substrate words" in Indo-European. This topic had interested me ever since I first flicked through Derksen's etymological dictionaries. The idea that some of the words in the modern languages could have been borrowed from undocumented sources was something that seemed very appealing to me, especially in northern Europe where the historical record is so shallow. Tijmen put me in touch with Guus Kroonen, and a couple of years on, I was accepted onto Guus' project ('The Linguistic Roots of Europe's Agricultural Transition').

It was amazing to have been able to conduct my research as part of a team, and the weekly group meetings, daily lunches and constant exchange of ideas (and idiocy) were a reliable source of inspiration and the perfect antidote to the mundane aspects of research. This team mentality led to three joint papers (Palmér et al. 2021; Kroonen et al. 2022; Thorsø et al. 2023), the workshop *Sub-Indo-European Europe*, which we hosted in August 2021, and a forthcoming volume based on the results of this workshop. Not only that, but the methodology expounded and applied in the second half of this work is as much a team effort as it is my own. I would like to sincerely thank Yvonne van Amerongen, Paulus van Sluis, Cid Swanenvleugel, Rasmus Thorsø, Andrew Wigman and honorary team member Axel Palmér for making these four years so special.

My individual project took me in a number of unexpected directions. Sergejus Tarasovas has taught me the ways of philological pedantry, and shown me that there is light at the end of most rabbit holes. Countless insights in this work, both big and small, ultimately derive from our discussions (some more are to be published as Jakob forthc. d.). Alexander Savelyev has shown incredible patience in giving me a crash course in Turkic linguistics. Practically all discussions of Turkic languages in this dissertation have greatly benefited from his input.

Most significant, however, was my journey to becoming a Uralicist (which perhaps I could now call myself, having spoken at the Finno-Ugricist Conference in Vienna, August 2022) — something I'd never considered when starting out on this project. A significant catalyzer for me has been the work of Ante Aikio, whose accessible writing style was the gateway I needed, but I would also like to thank Petri Kallio, Sampsa Holopainen, Juho Pystynen, Abel Warries and Mikhail Zhivlov for giving me heaps of constructive criticism and ideas which have made this dissertation considerably better, and especially Santeri Junttila, not only for the same, but also for providing me with an enormous amount of literature I would not otherwise have been able to access.

Dozens of other people have, at various moments, given me answers, made me question the answers I had, or made me realize I was asking the wrong questions. Without them, this dissertation would not be half what it is, and I will mention just one dozen who happen to spring to mind: Lucien van Beek, Chams Bernard, Eugen Hill, Martin Kümmel, Giedrė Motuzaitė Matuzevičiūtė, Michaël Peyrot, Tijmen Pronk, Peter Schrijver, Zsolt Simon, David Stifter and Rémy Viredaz.

This project would not have existed at all, of course, if it wasn't for its initiator Guus Kroonen and Professor (now Emeritus) Sasha Lubotsky,

who have supervised me throughout these four and a half years. And I know all too well that hasn't been the easiest of tasks. I am eternally grateful to them both for keeping their faith in me and helping me grow as a person. It has been an eventful few years, with the Covid-19 pandemic and war in Ukraine both massively impacting my personal life. The former — for the better, since it meant me and my wife could spend an unexpected and magical year together in Leipzig, the latter — decidedly for the worse, although for us it has been rather overshadowed by a serious struggle with my mother-in-law's health. I am incredibly thankful to Guus for being so understanding and allowing me to take the steps I needed.

Finally, let me mention my parents, Dan and Barbara, and my adoptive parent, Natalya, who are always categorically supportive and who I can always rely on to be on my side, no matter what kind of pickle I'm in. Last and most of all, I am thankful to my wife Sasha. I know she has had to put up with a lot, including a lot of my absence, but she has brightened up every single day and given me a reason to carry on. My only hope now is that we finally get a chance to find out what "normal" means!

Symbols and abbreviations

For abbreviated titles, see the relevant subsection of the bibliography.

Symbols

*	1. (before lemma) reconstructed form; 2. (after lemma) unattested form of an attested lexeme
**	counterfactual form
#	(in reconstructions) word boundary
Ø	zero
·	introduces a gloss
=	'(historically) identical to'
~	'belongs with (in unspecified capacity)', 'in variation with'
/	1. mutually exclusive alternative, 2. (in citations) demarcates co-authors
:	1. paradigmatically alternates with; 2. correlates with (providing an analogical model)
∞	'irregularly alternates with'
<	'developed from', 'derived from'
≪	'analogically replaced by' (also used of semantic developments)
←	'borrowed from'
/ /	phonological transcription
[ ]	1. phonetic transcription; 2. alteration to quoted text or form
⟨ ⟩	orthographic transcription
C	consonant
H	laryngeal
N	nasal
R	resonant
T	dental stop
V	vowel

General

arch. – archaic
BCE – before Common Era
c. – century
CE – Common Era
cf. – confer, see
dial. – dialectal
diss. – dissertation
E – east
ed., eds. – editor(s)
e.g. – *exempli gratia*, for example
et al. – *et alii*, and others
etc. – *et cetera*
fn. – footnote
forthc. – forthcoming
i.e. – *id est*, that is
in prep. – in preparation
lit. – literature
loc. cit. – *loco citato*, at the cited location
N – north
N.B. – *nota bene*, please note
NE – north-east
No. – number
NW – north-west
obs. – obsolete
op. cit. – *opere citato*, in the work cited
p. – page, pages
p.c. – personal communication
poet. – poetic
refs. – references
S – south
SW – south-west
sp. – (unspecified) species
s.v. – *sub verbo*, under the corresponding entry
usu. – usually
vel sim. – *vel similia*, or the like
viz. – *videlicet*, namely, to wit
vol., vols. – volume(s)
W – west

Languages

N.B. Abbreviations which only occur as second elements are given in [square brackets].

Alb. – Albanian
Arm. – (Old) Armenian
Aukšt. – Aukštaitian
Av. – Avestan
[B] – Baltic
Bel. – Belarusian
Bg. – Bulgarian
Bret. – Breton
Gaul. – Gaulish
Čak. – Čakavian (dialect of Serbo-Croatian)
Celt. – Celtic
Chuv. – Chuvash
CLuw. – Cuneiform Luwian

[Co.] – Cornish
CS – Church Slavic
  Bes. – *Besědy na evangelije*
  Ps. Sin. – *Psalterium Sinaiticum*
  Supr. – *Codex Suprasliensis*
Cz. – Czech
Du. – Dutch
E – Estonian
[E] – English
Eg. – Egyptian
F – Finnish
Fr. – French
[Fri.] – Frisian
[HG] – High German
HLuw. – Hieroglyphic Luwian
Gaul. – Gaulish
Gm. – Germanic
Go. – Gothic
Gr. – (Ancient) Greek
  H. – Hesychius
Hitt. – Hittite
HLv. – High Latvian
Ic. – Icelandic
It. – Italic
IE – Indo-European
Ingr. – Ingrian
[Ir.] – Irish
Iran. – Iranian
K – Karelian
Kajk. – Kajkavian (dialect of Serbo-Croatian)
Kash. – Kashubian
Khot. – Khotanese
Lat. – Latin
[LG] – Low German
Li. – Livonian
LSrb. – Lower Sorbian
Lt. – Lithuanian
[Luw.] – Luwian
Lv. – Latvian
M… – Middle, Mediaeval
  [MBel., MCo., MDu., MGr., MIr., MR, MUk., MW – see second element]
Ma. – Mari
  ~ E – Eastern (Meadow) Mari
  ~ NW – Northwestern Mari
  ~ W – Western (Hill) Mari
Mac. – Macedonian
Md. – Mordvin
  ~ E – Erzya
  ~ M – Moksha
ME – Middle English
MHG – Middle High German
ML – Mediaeval Latin
MLG – Middle Low German
MP – (Manichaean) Middle Persian
Mo… – Modern
  [MoHG, MoGr., MoIr., MoLG, MoW – see second element]
MoE – Modern English
Myc. – Mycenian
Ngan. – Nganasan
NP – New (= Classical) Persian
Nw. – Norwegian
O… – Old
  [OAv., OBret., OCo., OCz., ODu., OFr., OFri., OIr., OPl., OR, OSw.,

OTur., OW – see second element]
OCS – Old Church Slavic
OE – Old English
OHG – Old High German
ON – Old Norse
ONovg. – Old Novgorodian (dialect of Old Russian)
OS – Old Saxon
Osc. – Oscan
Oss. – Ossetic
~ I – Iron
~ D – Digor
P… – Proto-
[PGm., PMa., PMd., PSá – see second element]
Parth. – Parthian
PIE – Proto-Indo-European
Pl. – Polish
Pr. – (Old) Prussian
~ E – Elbing Vocabulary
~ Gr. – Grunau's vocabulary
~ III – Third Catechism
~ TC – Trace of Crete
PrG – Prussian German
PrLt. – Prussian Lithuanian
PU – Proto-Uralic
R – Russian
RCS – Russian Church Slavic
Sá. – Sámi (Saami)
~ I – Inari (Aanaar) Sámi
~ K – Kildin Sámi
~ L – Lule Sámi
~ N – North Sámi
~ S – South Sámi
~ Sk. – Skolt Sámi
SCr. – Serbo-Croatian
SCS – Serbian Church Slavic
Skt. – Sanskrit
Sl. – Slavic
Slk. – Slovak
Sln. – Slovene
Sw. – Swedish
Tat. – Tatar
To. – Tocharian
Tur. – Turkic
Turk. – Turkish
Uk. – Ukrainian
Udm. – Udmurt
Umbr. – Umbrian
Ural. – Uralic
USrb. – Upper Sorbian
Vp. – Veps
Vt. – Votic
Võ. – Võro (South Estonian)
[W] – Welsh
YAv. – Young Avestan
Žem. – Žemaitian

Grammatical

1… (2…, 3…) – first (second, third) person
ABL. – ablative
ACT. – active
ADJ. – adjective
ADV. – adverb

ACC. – accusative
COLL. – collective
DAT. – dative
DU. – dual
ELAT. – elative
ESS. – essive
GEN. – genitive
F. – feminine
ILL. – illative
INST. – instrumental
LAT. – lative
M. – masculine
N. – neuter
OBL. – oblique
PL. – plural
PART. – partitive
PRES. – present
PRET. – preterite
PRT. – participle
SG. – singular

Data sources and conventions

Reconstructions

In general, following in part the practice adopted by OED[3], I have avoided providing reconstructions unless this is necessary for the argument. The rationale is to maximize the emphasis given to attested data, and also reduce the need for me as an author to make a clear stand with regard to theoretical aspects of reconstruction where these are not strictly relevant to an argument. For instance, while one might object to my reconstruction of Lithuanian *avìs* 'sheep' as PIE *h_3*eui*-, it is unlikely anyone will object to its equation with Latin *ovis* 'sheep'. I have made an exception in the case of evidence from Uralic languages, for which I have provided reconstructions quite systematically. This is partly a means to provide additional clarity for readers more familiar with Indo-European than with Uralic, but is also a reflection of my own process in dealing with these languages.

In the following cases, my reconstruction differs from the established norm and/or requires certain clarification:

East Baltic — Acute intonation is marked with the circumflex or caret ⟨ ˆ ⟩, following the use of this symbol as an indication of the broken tone in Latvian tonal orthography.

Slavic — My reconstruction of the Slavic vowel system differs from the OCS-based one conventionally used (for instance, in the dictionaries of Berneker, Vasmer, Derksen, and ЭССЯ). The use of the standard reconstruction leads to anachronisms, such as an apparent claim that the Early North Russian source of e.g. Finnish dial. *akkuna* 'window' is

more archaic than its Proto-Slavic ancestor. In general, I consider the Slavic vowel shift to be a Common Slavic, not Proto-Slavic development (cf. Olander 2015: 59–67), and therefore use a reconstruction with pre-vowel shift values.[1] The reconstruction used in this work is as follows:

This work			**Traditional**		
ī	ȳ	ū	i	y	u
i̯		u̯	ь		ъ
	e	a		e	o
ē		ā	ě		a

Finnic — In line with the "Leiden" tradition of not marking allophonic features, such as syllabic resonants, in reconstructions, I do not mark consonant gradation, as this is entirely predictable in Proto-Finnic (except for **s* between unstressed vowels, where I have preserved the alternation with **h*). Therefore, I reconstruct Finnish *hammas* (GEN.SG. *hampaan*) 'tooth' as **hampas* rather than **hambas*, **hamβas* or **hamp̆as*.

Data sources

In compiling this work, I have endeavoured to check all the forms cited in primary sources. I have generally avoided citing data which I was unable to independently verify, unless this is crucial to an argument. A selection of literature used to source the forms from the most

[1] Differently from Olander, I do not operate with a Proto-Slavic predating the monophthongization of diphthongs. One reason for this is practicality: for instance, it is often impossible to decide whether Common Slavic *i* derives in any particular case from an earlier **ei* or **ī*. However, I also do not consider it likely that the monophthongization was a post-Proto-Slavic development: the absence of the second palatalization in North Russian (cf. also Holzer 2001: 39–40) does not necessarily imply that it branched off before the development **ai* > **ē* since there is no reason to exclude an intermediate stage */kæ:/.

important languages for this work is presented below. The orthography follows the cited sources unless otherwise indicated. References not provided in full can be found in the bibliography. All web links are valid as of 31 May 2023.

Lithuanian	DLKŽ; LKŽ; *Bendrinės lietuvių kalbos žodynas* (ekalba.lt)
Latvian	*Latviešu literārās valodas vārdnīca* (tezaurs.lv); ME; LVPPV (tonal orthography follows ME; differences with LVPPV have been noted)
Prussian	PKEŽ and facsimile copies hosted at www.prusistika.flf.vu.lt; Trautmann 1910
Russian	*Большой академический словарь русского языка* [*а-продел*]; *Толковый словарь русского языка* (ed. Дмитрий Н. Ушаков); СРНГ; СДРЯ 11–14; СРЯ 11–17; СДРЯ. I have followed pre-revolutionary orthography in the use of the symbol ⟨ѣ⟩ (italics ⟨*ѣ*⟩), where this is etymologically relevant.
Ukrainian	*Словник української мови* (sum.in.ua); ЕСУМ
Belarusian	*Тлумачальны слоўнік беларускай мовы* (1977–1984); ЭСБМ
Polish	*Słownik języka polskiego* (ed. Witold Doroszewski); SSP
Czech	*Příruční slovník jazyka českého* (psjc.ujc.cas.cz); Gebauer; digitized resources at vokabular.ujc.cas.cz
Slovak	*Slovník slovenského jazyka* (slovnik.juls.savba.sk/?d=peciar); SSN
Sorbian	Schuster-Šewc; dolnoserbski.de; hornjoserbsce.de
Slovene	Pleteršnik (tonal orthography follows Pleteršnik; additional data from SSKJ^2 is given as in the source)
Serbo-Croatian	RJA; PCA; Skok (additional dialect data cited after Derksen 2008, 2015)
Church Slavic	*Старославянский словарь (по рукописям X-XI веков)*; SJS; Miklosich 1865

Bulgarian	*Речник на български език* (ibl.bas.bg/rbe/); БЕР
Macedonian	*Дигитален речник на македонскиот јазик* (drmj.eu)
Gothic	Streitberg, Wilhelm. *Gotisch-Griechisch-Deutsches Wörterbuch* (1910) and interlinear texts at wulfila.be/gothic/
Scandinavian	Cleasby/Vigfusson 1874; *Dictionary of Old Norse Prose* and other digitized resources at onp.ku.dk; Blöndal 1989, SAOB, *Norsk ordbok* (no2014.uib.no)
English	DOE; MED; OE. Middle English allophonic lengthening is not marked
Dutch	Philippa et al.; digitized resources at gtb.ivdnt.org
Low German	Tiefenbach 2010; MndWb; Schiller/Lübben. MLG orthography generally follows the latter: allophonic lengthening is not marked; however, I have, after MndWb, distinguished *ö* from *o*
High German	AWb; Schützeichel 2004; MWb; *Mittelhochdeutsches Handwörterbuch* (ed. Matthias Lexer); DWb
Celtic	eDIL; GPC; *Le dictionnaire diachronique du breton* (ed. Martial Menard, devri.bzh)
Latin	TLL; Lewis/Short; Walde/Hoffman; Ernout/Meillet
Albanian	Mann, Stuart E. *An Historical Albanian-English Dictionary* (1948); Demiraj 1997
Greek	LSJ; additional data from Frisk; Beekes 2010.
Armenian	Martirosyan 2009
Finnish	SMS [*a–mähistyä*]; VKS [*a–pitäytä*]; SSA; SKES
Karelian	KKS
Veps	Зайцева/Муллонен 1972; Зайцева 2010. Orthography follows the latter
Votic	VKS
Estonian	EMS [*a–puisklema*]; VMS; *Eesti keele seletav sõnaraamat* (2009); other digitized resources at portaal.eki.ee/sonaraamatud.html; South Estonian: synaq.org

Livonian	Kettunen 1938; Viitso/Ernštreits 2012. Courland Livonian orthography follows the latter. Salaca Livonian: Winkler/Pajusalu 2009
Western Sámi	Digitized resources at kaino.kotus.fi/algu/. Orthography after Bergsland & Magga Mattson (1993). *Sydsamisk-norsk ordbok* [South]; Korhonen (2007). *Lulesamisk-svensk ordbok* [Lule]; Sammallahti & Ocejohka (1989). *Saamelais-suomalainen sanakirja* [North]. Data from older sources is presented in updated orthography
Eastern Sámi	Digitized resources at kaino.kotus.fi/algu/. Orthography follows Sammallahti & Morottaja (1993). *Inarinsaamelais-suomalainen sanakirja* [Inari]; Sammallahti & Mosnikoff (1991). *Suomi-koltansaame sanakirja* [Skolt]. Kildin Sámi data is updated from Lehtiranta 2001 on the basis of Rießler 2022
Mordvin	MdWb; Серебренников et al. (eds.), *Эрзянско-русский словарь* (1993), *Мокшанско-русский словарь* (1998). Phonemic orthography simplified from MdWb, and updated to reflect the modern standard
Mari	TschWb. Phonemic orthography after e.g. Aikio 2014
Permic	Digitized resources at dict.fu-lab.ru, Лыткин/Гуляев 1970
Khanty	OstWb
Mansi	Kannisto (comp.), *Wogulisches Wörterbuch* (2014). Phonemic orthography follows secondary literature