

Language Dispersal Beyond Farming

*edited by Martine Robbeets
and Alexander Savelyev*

John Benjamins Publishing Company

Language Dispersal Beyond Farming

Language Dispersal Beyond Farming

Edited by

Martine Robbeets

Alexander Savelyev

Max Planck Institute for the Science of Human History, Jena

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/z.215

Cataloging-in-Publication Data available from Library of Congress

ISBN 978 90 272 1255 9 (HB)

ISBN 978 90 272 6464 0 (E-BOOK)

© 2017 –John Benjamins B.V.

The electronic edition of this book is Open Access under a CC BY-NC-ND 4.0 license.
<https://creativecommons.org/licenses/by-nc-nd/4.0>

This license permits reuse, distribution and reproduction in any medium for non-commercial purposes, provided that the original author(s) and source are credited. Derivative works may not be distributed without prior permission.

This work may contain content reproduced under license from third parties. Permission to reproduce this third-party content must be obtained from these third parties directly.

Permission for any reuse beyond the scope of this license must be obtained from John Benjamins Publishing Company, rights@benjamins.nl

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

List of tables	VII
List of figures	IX
List of contributors	XI
Acknowledgements	XIII
CHAPTER 1	
Farming/Language Dispersal: Food for thought <i>Martine Robbeets</i>	1
CHAPTER 2	
Proto-Quechua and Proto-Aymara agropastoral terms: Reconstruction and contact patterns <i>Nicholas Q. Emlen and Willem F. H. Adelaar</i>	25
CHAPTER 3	
Subsistence terms in Unangam Tunuu (Aleut) <i>Anna Berge</i>	47
CHAPTER 4	
Lexical recycling as a lens onto shared Japano-Koreanic agriculture <i>Alexander Francis-Ratte</i>	75
CHAPTER 5	
The language of the Transeurasian farmers <i>Martine Robbeets</i>	93
CHAPTER 6	
Farming-related terms in Proto-Turkic and Proto-Altaiic <i>Alexander Savelyev</i>	123
CHAPTER 7	
Farming and the Trans-New Guinea family: A consideration <i>Antoinette Schapper</i>	155

CHAPTER 8	
The domestications and the domesticators of Asian rice	183
<i>George van Driem</i>	
CHAPTER 9	
Macrofamilies and agricultural lexicon: Problems and perspectives	215
<i>George Starostin</i>	
CHAPTER 10	
Were the first Bantu speakers south of the rainforest farmers?	
A first assessment of the linguistic evidence	235
<i>Koen Bostoen and Joseph Koni Muluwa</i>	
CHAPTER 11	
Expanding the methodology of lexical examination in the investigation of the intersection of early agriculture and language dispersal	259
<i>Brian D. Joseph</i>	
CHAPTER 12	
Agricultural terms in Indo-Iranian	275
<i>Martin Joachim Kümmel</i>	
CHAPTER 13	
Milk and the Indo-Europeans	291
<i>Romain Garnier, Laurent Sagart and Benoît Sagot</i>	
Language index	313
Subject index	321

List of tables

Chapter 2

- Table 1. Crops and plant parts
- Table 2. Agricultural techniques, tools, structures, and materials
- Table 3. Food products derived from agriculture, and associated tools and techniques
- Table 4. Domesticated animals
- Table 5. Herding techniques, structures, locations, and materials
- Table 6. Weaving techniques and technology

Chapter 3

- Table 1. Unangan subsistence terms by gendered activity and semantic domain
- Table 2. Percentages of Eskimo-Aleut cognates in semantic domains pertaining to subsistence terminology
- Table 3. Percentages of Eskimo-Aleut borrowings in semantic domains pertaining to subsistence terminology
- Table 4. Relative levels of borrowings of subsistence terminology between neighboring Yupik languages and Unangam Tunuu

Chapter 4

- Table 1. Agricultural vocabulary (1)
- Table 2. Agricultural vocabulary (2)
- Table 3. Non-trivial consonant correspondences
- Table 4. Non-trivial vowel correspondences
- Table 5. Etymologies with pre-MK **po*
- Table 6. Etymologies with pre-MK **po* and Japanese cognates
- Table 7. Etymologies with pre-MK **po* and pKJ reconstructions
- Table 8. Lexical recycling of pre-rice vocabulary

Chapter 5

- Table 1. Bayesian time estimates for the primary splits in the Transeurasian family

Chapter 6

- Table 1. Proto-Turkic pastoralist vocabulary
- Table 2. Proto-Turkic agricultural vocabulary

Chapter 7

- Table 1. Proto-Trans-New Guinea free pronouns
- Table 2. Proto-Trans-New Guinea object prefixes
- Table 3. Posited Proto-Trans-New Guinea **jaBu* ‘sugarcane’
- Table 4. Posited Proto-Trans-New Guinea **mungo[l]* ‘banana’

Chapter 10

- Table 1. Reflexes of **-kòndò* ‘banana’ in present-day West-Coastal Bantu languages
- Table 2. Reflexes of **-gómbo* ‘okra’ in present-day West-Coastal Bantu languages
- Table 3. Reflexes of **-kúmà* ‘kapok tree’ in Bantu languages belonging to distinct major branches
- Table 4. Reflexes of **-dódò* ‘*Annona* sp.’ in Bantu languages belonging to distinct major branches
- Table 5. Reflexes of **-pòmí/*-kaca* ‘*Erythrophleum guineense/suaveolens*’ in West-Coastal Bantu
- Table 6. Reflexes of **-pánji* ‘*Pentaclethra macrophylla*’ in West-Coastal Bantu

Chapter 13

- Table 1. Languages where both ‘to milk’ and the name for ‘milk’ are from **h₂melǵ-*
- Table 2. Languages where only the verb ‘to milk’ is from **h₂melǵ-*

List of figures

Chapter 1

Figure 1. Distribution of the homelands proposed in this volume

Figure 2. Range of time depths estimated for the language families discussed in our volume

Figure 3. A continuum-distribution for agricultural lexicon discussed in this volume

Chapter 2

Figure 1. Simplified history of the Quechuan and Aymaran lineages

Chapter 3

Figure 1. Late Prehistoric population movements, ca. 1000–400 BP

Chapter 5

Figure 1. The Transeurasian languages (generated with WALS tools)

Figure 2. The ethnic groups of prehistorical Manchuria in the first millennium BC according to Janhunen (1996: 216)

Figure 3. Densi Tree of the Transeurasian family (Robbeets & Bouckaert forthc.)

Figure 4. The Xinglongwa culture and the establishment of millet agriculture

Figure 5. The Hongshan culture and the eastward spread of millet agriculture

Figure 6. The Yayoi culture and the integration of rice and millet agriculture

Figure 7. Mapping the agricultural development in Northeast Asia on the language tree of Transeurasian

Chapter 7

Figure 1. Papuan language families (shaded)

Figure 2. Posited Trans-New Guinea language families (after Ross 2005)

Figure 3. Posited homeland and direction of Trans-New Guinea expansions

Chapter 8

Figure 1. The relative position of early Hmong-Mien (Miáo-Yáo) tribes and early Kra-dai (T'ai) tribes (reproduced from Forrest 1948: 129)

Figure 2. The geographical distribution of Austroasiatic language communities

Figure 3. The family tree of Austroasiatic (Diffloth 2012). Unlike the Khasi-Aslian branch, the internal phylogeny of the Munda branch has not been established

Figure 4. The geographical ranges for the possible domestication of (A) *ghaiyā* or upland rice, (B) wet *indica* rice and (C) the *japonica* cultivar (adapted from Londo et al. 2006)

Figure 5. The region of overlap of the geographical ranges of megafaunal species for which Proto-Austroasiatic etyma are reconstructible

Figure 6. The 2012 Benares recension of Stanley Starosta's 2001 Périquex East Asian linguistic phylum (Starosta 2005; van Driem 2014b)

Figure 7. After the last glacial maximum, the Y chromosomal haplogroup O (M175) split into the subclades O1 (F265, M1354) and O2 (M122)

Figure 8. A male-biased linguistic intrusion introduced both Austroasiatic language and a paternal lineage, haplogroup O1b1a1a (M95), into the indigenous population of the Choṭā Nāḡpur

Figure 9. Branching of the paternal lineage O into new subclades

List of contributors

Willem Adelaar

Leiden University Centre for Linguistics
Postbus 9515
NL-2300 RA Leiden
The Netherlands
w.f.h.adelaar@hum.leidenuniv.nl

Anna Berge

Alaska Native Language Center
University of Alaska Fairbanks
P.O. Box 750119
Fairbanks, AK 99775
USA
amberge@alaska.edu

Koen Bostoen

BantUGent – UGent Centre for Bantu
Studies
Department of Languages and Cultures
Ghent University
Rozier 44
9000 GENT
Belgium
koen.bostoen@ugent.be

George van Driem

Institut für Sprachwissenschaft
Universität Bern
Länggassstrasse 49
CH 3000 Bern 9
Switzerland
george.vandriem@isw.unibe.ch

Nicholas Q. Emlen

Leiden University Centre for Linguistics
Postbus 9515
NL-2300 RA Leiden
The Netherlands
n.q.emlen@hum.leidenuniv.nl

Alexander T. Francis-Ratte

Furman University
Furman Hall 235-H
3300 Pionsett Hwy
Greenville, SC 29613
USA
alexander.francis-ratte@furman.edu

Romain Garnier

Université de Limoges
Faculté des Lettres et des Sciences Humaines
39E rue Camille Guerin
87036 Limoges cedex
France
garromain@gmail.com

Brian Joseph

Department of Linguistics
The Ohio State University
Columbus Ohio 43210
USA
joseph.1@osu.edu

Joseph Koni Muluwa

I.S.P. de Kikwit, Democratic Republic of the
Congo
BantUGent – UGent Centre for Bantu
Studies
Department of Languages and Cultures
Ghent University
Rozier 44
9000 GENT
Belgium
Joseph.KoniMuluwa@UGent.be

Martin Kümmel

Friedrich-Schiller-Universität Jena
Philosophische Fakultät
Institut für Orientalistik, Indogermanistik
Ur- und Frühgeschichtliche Archäologie
Seminar für Indogermanistik
Zwätzengasse 12a
D-07743 Jena
Germany
martin-joachim.kuemmel@uni-jena.de

Martine Robbeets

Max Planck Institute for the Science of
Human History
Department of Linguistic and Cultural
Evolution
Kahlaische Strasse 10
07745 Jena
Germany
robbeets@shh.mpg.de

Laurent Sagart

Centre de Recherches Linguistiques sur l'Asie
Orientale
INALCO
2 rue de Lille
75007 Paris
France
sagart@ehess.fr

Benoît Sagot

Institut National de Recherche en
Informatique et en Automatique
2 rue Simone Iff
CS 42112
75589 Paris Cedex 12
France
benoit.sagot@inria.fr

Alexander Savelyev

Max Planck Institute for the Science of
Human History
Kahlaische Strasse 10
07745 Jena
Germany
savelyev@shh.mpg.de

Antoinette Schapper

a_schapper@hotmail.com

George Starostin

Institute for Oriental and Classical Studies
Russian State University for the Humanities
Miusskaya Square, 6
125993 Moscow
Russian Federation
gstarst@rinet.ru

Acknowledgements

The incentive for this collective volume came from a symposium entitled “The language of the first farmers”, organized by Martine Robbeets at the 49th Annual Meeting of the Societas Linguistica Europaea in Naples, September 2–3, 2016. We thank all participants who contributed their papers to the symposium: Alexander Francis-Ratte, George Starostin, Anna Berge, Koen Bostoen, Joseph Koni Muluwa, Tom Güldemann, Anne-Maria Fehn, Antoinette Schapper, Nicholas Emlen, Willem Adelaar, Brian Joseph, Martin Kümmel, Laurent Sagart, Romain Garnier, Adam Hyllested and Russell Gray.

We are grateful to the authors in the volume for submitting and revising their papers and for respecting our strict schedule in spite of their busy agendas. A heartfelt word of thanks also goes to Yanjun Liu, who is currently involved in the eurasia3angle project as a student assistant for her dedicated work as an Assistant Editor. We further acknowledge our other team members of the eurasia3angle project for their critical but supportive comments on Farming/Language Dispersal. Our gratitude also goes to our colleagues at the Max Planck Institute for the Science of Human History in Jena, notably Annette Günzel and Moritz Zauleck for their graphic support and to Anne Gibson for her English editing.

The realization of this volume has received generous funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 646612), granted to Martine Robbeets.

We would also like to thank Kees Vaes for his enthusiasm in publishing this volume and for his help in seeing this project through. Finally, we hope that many readers will click the URL, providing open access to this volume, as Kees promised to consider the publication of an interdisciplinary linguistic journal if our volume is widely consulted online.

Jena, June 2017

Farming/Language Dispersal

Food for thought

Martine Robbeets

Max Planck Institute for the Science of Human History

1. Agriculture-driven language spread

Just as plant and animal lineages are not uniformly distributed around the world, the same is true for the distribution of language families. As of 2017 the Ethnologue list includes around 50 distinct language families covering 7099 living languages, some of which, like Austronesian, have spread over a huge geographical range while others, like Amuric, have only a single living member (i.e., Nivkh) and are geographically very restricted. The uneven geographical distribution of language families across the world calls for an explanation of why some languages wither and die, while others prosper and spread. A major reason proposed to explain the spread of many of the world's large language families is agriculture. This proposal, advanced by Renfrew (1987), Bellwood & Renfrew (2002), Diamond & Bellwood (2003) and Bellwood (2005, 2011) is known under the label "Farming/Language Dispersal Hypothesis". The hypothesis posits that many of the world's major language families owe their dispersal to the adoption of agriculture by their early speakers. In this context, farming or agriculture is generally understood in its restricted sense of economic dependence on the cultivation of crops and does not usually include the raising of animals as livestock.

Since farming can unquestionably support far greater population densities than hunting and gathering, the basic logic behind this hypothesis is that population growth steadily pushed the early farmers and their language into wider territories, displacing the languages of preexisting hunter-gatherer populations. Indeed, agriculture is argued to be one of the major factors causing dispersal in families such as Indo-European (Renfrew 1987; Comrie 2002; Gray & Atkinson 2003) in Europe, Bantu (Philipson 2002) and Semitic (Diakonoff 1998) in Africa, Austronesian (Blust 1995, 2013; Pawley 2002; Bellwood & Dizon 2008), Sino-Tibetan (Janhunen 1996: 222; LaPolla 2001; Sagart 2008, 2011), Tai-Kadai (Ostapirat 2005: 128), Austroasiatic (Higham 2002; Diffloth 2005; Sidwell & Blench 2011; Sagart 2011) and

Dravidian (Fuller 2002) in Asia and Tupian, Arawakan (Aikhenvald 1999:75) and Otomanguean (Kaufman 1990; Brown et al. 2013a/b, 2014a/b) in the Americas.¹

In this volume, we would like to investigate to what extent the economic dependence on plant cultivation impacted language spread in various parts of the world, reassessing some of the above proposals and paying attention to language families that cannot unequivocally be regarded as instances of Farming/Language Dispersal, even if subsistence may have played a role in their expansion.

In the contribution on Eskimo-Aleut by Anna Berge, it is clear that the expansion could not have been driven by agriculture because this widely spread language family never developed farming in the first place. Nevertheless, a hunter-gatherer subsistence strategy that provided access to relatively rich food resources had linguistic effects equivalent to those brought by agriculture.

There are also contributions on widely spread language families, for which the ancestral vocabulary at best provides only a glimpse of agriculture, such as Trans-New Guinea by Schapper, Transeurasian by Robbeets, Turkic and Altaic by Savelyev and various macrofamilies in Eurasia by Starostin.

Moreover, we find widespread families, for which an agricultural lexicon can be confidentially reconstructed, but where it remains unclear whether agriculture is indeed the reason for their spread. This is, for instance, the case for the Quechuan and Aymaran languages discussed by Emlen and Adelaar and for the Hmong-Mien languages discussed by van Driem. It is arguable that proto-Hmong-Mien had rice agricultural vocabulary and its homeland was situated in the Mid-Yangtze Valley where *japonica* rice was first domesticated. However, the prevalent view (Ratliff 2004: 158–159; Sagart 2011: 127–128) that most of its rice vocabulary has been borrowed from Sinitic and that it has a relatively shallow time-depth (500 BC) is in conflict with the direction of borrowing and time depth suggested by van Driem. Uncertainty about agriculture-driven expansion despite the reconstruction of some agricultural vocabulary also marks the debate in Indo-European between the Anatolian hypothesis, suggesting that farmers migrated out of the Middle East around 7000 BC, on the one hand, and the Steppe hypothesis, suggesting that herders migrated out of the Eurasian steppe around 4000 BC, on the other. Whereas the former hypothesis is in accordance with Renfrew's (1987) traditional view of Farming/Language Dispersal, the contributions by Joseph, Kümmel and Garnier et al. supporting the latter hypothesis should not necessarily be in conflict with the model of subsistence-driven linguistic expansion in general.

1. Brown (2015) now challenges his earlier proposal that agricultural vocabulary can be reconstructed back to proto-Otomanguean, arguing that the Otomanguean languages are not yet conclusively demonstrated to descend from a common ancestor.

Next, there is the Bantu spread discussed by Koen Bostoen and Joseph Koni Muluwa, previously claimed to be “one of the most dramatic examples of language/farming dispersal in world history” (Bellwood 2005: 222). However, as the authors show, Bantu turns out to be a less convincing case of agriculture-driven spread than initially anticipated.

Finally, this volume also includes a discussion of a language family for which there seems to be a relative consensus about Farming/Language Dispersal, notably Austroasiatic. Regardless of the controversy about the location of the homeland, be it in the Mekong Valley (Sidwell & Blench 2011: 318) or as van Driem suggests in his contribution, in the Brahmaputra Valley, there seems to be a consensus that the dispersal of the Austroasiatic languages could have been motivated by the spread of rice agriculture.

As such, the contributions to this volume differ from the influential works mentioned above in that they do not perfectly fit into a framework of agriculture-driven language spread, but invite us to relativize the importance of the factor of agriculture, without completely rejecting it. Taken together, our case studies make it clear that farming is neither a necessary nor sufficient condition for language spread and that we need to abandon one-factor explanations and consider many other causes that may have influenced linguistic expansion. Moreover, this volume shows that a dualistic concept of a proto-language either having or lacking agricultural vocabulary is untenable and urges us to think in terms of a continuum-distribution of agricultural proto-lexicon.

2. Data and questions

The language families discussed in this volume are very diverse and widely distributed across continents, from Africa to Europe, Asia and Oceania to the Americas. In Africa, we find the homeland of West-Coastal Bantu, situated between the Bateke Plateau and the Bandundu region in Congo and that of Afroasiatic, situated in the Eastern Mediterranean by Militarev (2002) but in the western Red Sea Coast by Ehret (2003). In Eurasia, the location of the assumed homelands ranges from the Pontic Steppe north of the Black Sea for Indo-European, the region south of the Caucasus for Nostratic and the area around the Aral Sea for proto-Indo-Iranian, over to the Brahmaputra Valley area for Austroasiatic, the mid-Yangtze River Basin for Hmong-Mien to the West Liao River Basin for Transeurasian and the Liaodong Peninsula for Japano-Koreanic. In Oceania, the homeland of Trans-New Guinea is situated in the central highlands of Papua New Guinea. In the Americas, we find the original location of Eskimo-Aleut on the North American Pacific Coast and the

homelands of Quechua and Aymara in central Peru. Figure 1 shows the proposed locations for the homelands of the language families discussed in this volume.

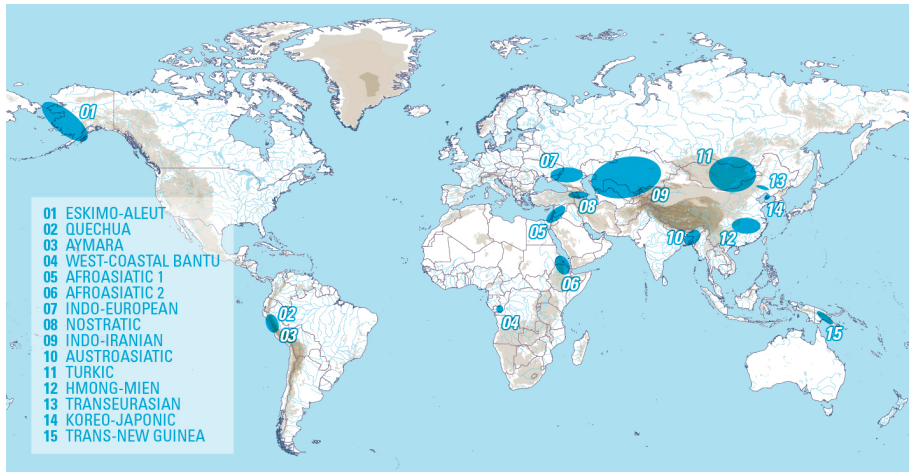


Figure 1. Distribution of the homelands proposed in this volume

Not only the presumed locations but also the estimated time-depths of the ancestral languages under discussion show much variety. The shallowest time-depths are situated around the beginning of our era with Quechua, Aymara, West-Coastal Bantu and Hmong-Mien. Other families such as Indo-Iranian, Japano-Koreanic and Eskimo-Aleut go back to between 2000 and 3000 BC, while Indo-European, Austroasiatic, Transeurasian and Trans-New Guinea lie between 4000 and 6000 BC. Long-range families under discussion, situated around 10,000 BC and beyond include Sino-Caucasian, Afroasiatic and Nostratic.

The questions we address in this volume are in the first place linguistically oriented, investigating language in order to draw inferences about early subsistence strategies and causes of dispersal. However, we are also interested in how our knowledge about early subsistence and demography can help us to draw inferences about language. The following questions are related to the use of language as a window on early subsistence in individual case studies.

1. What was the subsistence component of a given ancestral language like? What words did the ancestral speakers use to designate the environment they lived in, the plants they cultivated, the animals they raised, the food they consumed and the technology they used in their daily lives?
2. Can we estimate the time depth and the location of a given ancestral language?
3. What kind of linguistic evidence is required to conclude that a proto-language was spoken by farmers?

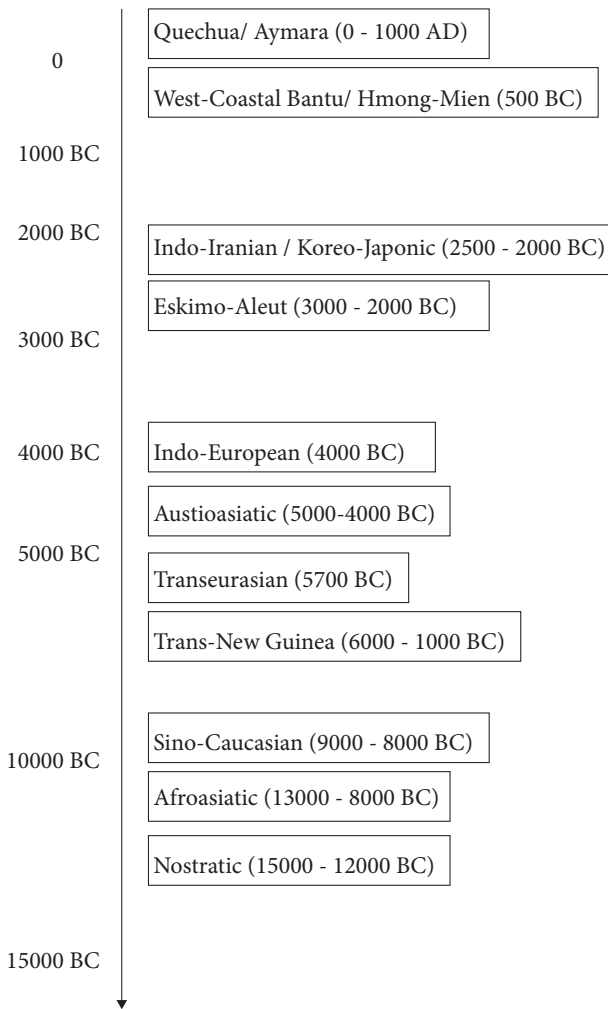


Figure 2. Range of time depths estimated for the language families discussed in our volume

4. Does the reconstruction of agricultural vocabulary to the proto-language of a widespread language family necessarily imply that the language spread was driven by agriculture?
5. Are there any linguistic traces of interactions between the ancestral speakers of a given proto-language and other groups? Who was involved? What was their relationship like? Did the relationship involve the transfer of subsistence strategies or technologies?

By contrast, the following questions draw on what we know about prehistoric demography and subsistence and use this information as a window on language.

1. Does the archaeological information about early subsistence at the proposed time and location of the linguistic homeland tie in with the reconstructed terms for subsistence, technology and natural environment?
2. Are there any indications for a switch from a less successful subsistence style to a lifestyle based on more successful subsistence strategies, e.g., from hunting and gathering to agriculture? Is there any evidence that this change was mirrored by language replacement?
3. Which demographic transitions have occurred at the estimated time and in the homeland of the ancestral language and are these changes mirrored in linguistic effects such as splits and spreads of the language family? Can they be attributed to a change in subsistence style?
4. Are there indications that relativize the importance of agriculture as a factor behind the expansion of language families? What other processes can account for early language spread?

3. Methods

The tools that can help us to find an answer to our questions are situated at the interface between linguistics and other disciplines, such as archaeology and genetics. Such tools include, notably, the diversity hotspot principle, phylolinguistics, mapping demographic dispersal on linguistic phylogeny, cultural reconstruction and contact linguistics. The integration of these different methods and principles will result in a clearer window on the past than would the individual application of one or another method. Each approach has its own pitfalls, but we can gain more from applying and integrating the various methods than we can lose from disregarding them.

3.1 The diversity hotspot principle

The “diversity hotspot principle” is not so much a method, but rather a principle that can help us in locating the original homeland of a language family. The notion was originally Edward Sapir’s (1916: 87), who referred to it as the “centre of gravity principle”, but it is also known as the “focus of diversity” principle (Heggarty 2015: 612–613). Assuming that the deepest splits within a family reflect the greatest age, the location of these splits on the map is thought to point to the area where the proto-language began to diversify. The principle is thus 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. A schoolbook example is the Austronesian family, that extends across a huge geographical range, all the way from Madagascar to Easter Island, but the deepest subgroups are found on just one Island, Taiwan. In Chapter 7, Schapper applies the principle to the Trans-New Guinea family, indicating that the eastern highlands of Papua New Guinea is the best candidate for a homeland because it has the highest concentration of primary subgroups.

Although the diversity hotspot principle can provide some clues about the homeland of a language family, it must also contend with certain limitations. First, the identification of the homeland depends on the location of the deepest subgroups and therefore, on how robustly the internal structure of a given family has been established. In the case of Austroasiatic, for instance, van Driem finds that the concentration of the deepest phylogenetic divisions in the family tree points to the northern Bay of Bengal littoral, but if Sidwell and Blench (2011) are correct in establishing a “flat array” structure of Austroasiatic, in which Munda would not be a primary branch, this would shift the center of gravity of the family towards the Mekong Valley, as they suggest.

A second limitation of this principle is that the contemporary hotspot of linguistic diversity may diverge from the earlier one. Looking at the present map of Indo-European with the Balkan Peninsula hosting the highest diversity of deep subgroups, we might conclude that the homeland is there, instead of the Pontic Steppe or Anatolia. A possible way out is to return to the earliest language distributions we know of. In this volume, van Driem, for instance, uses the historically attested distribution of the early Hmong-Mien tribes during the Eastern Zhōu dynasty (770–256 BC) to push the homeland of the family further north, towards the middle Yangtze and Robbeets proposes a location for the Transeurasian homeland on the basis of records of ethnic and linguistic diversity in Chinese historical sources. However, earlier diversity may also have been lost long before recorded historical times. This observation is at the basis of Starostin’s discussion of the various homeland theories for the Afroasiatic stock. Some scholars such as Ehret (2003) favor a homeland in the Horn of Africa on the grounds that, except Semitic, all subgroups occur only in Africa, while others, such as Militarev (2002), support it having originated in the Levant, where earlier diversity may simply have been lost. This example makes it clear that the application of the diversity hotspot principle at profound time-depths is highly speculative because the elapse of time may have erased earlier diversity and the proposed genealogical relationships are not reliably established.

Finally, linguistic diversity is a function not only of time but also of other factors such as environmental change and disease. These may have made the original homeland unsuitable for human habitation at a certain point in time. In this way

original linguistic diversity may have been erased and it may no longer be possible to pinpoint the homeland using the diversity hotspot principle. However, even if the principle is not foolproof, it offers valuable clues for the location of a homeland at less remote time-depths.

3.2 Phylolinguistics

A second tool that is useful for our linguistic window on the past is “phylolinguistics”, a cover term for all quantitative approaches to language change, based on the historical behavior of cognate sets. This includes distance-based approaches, such as the lexicostatistic method mentioned by Starostin as well as character-based approaches, such as the Bayesian method, which became widely applied to linguistics since Gray and Jordan (2000) and is here applied by Robbeets to the Transeurasian languages. These methods estimate the relationship between two languages, the former from the amount of difference in their shared cognate proportion and the latter by inferring the pathways by which each developed from their common ancestor (Dunn 2015). Such computational techniques can be useful in double-checking the internal structure of a linguistic family previously established on the basis of classical historical linguistics, providing us with absolute dates for the nodes in a given family and by giving us an idea of the robustness of our inferences. The assumptions are, first, that the amount of language change between two related languages is in relation to their divergence time and, second, that we can calibrate the divergence time against known cases of language divergence over attested timespans. Among the challenges of phylolinguistics for classical historical linguists, we can first mention the “garbage in, garbage out” principle, meaning that our inferences will depend on the quality of the inserted data and how we interpret their coding. Second, the “mathemagic” these methods involve is at times difficult to access for classically trained historical linguists. In order to evaluate the quality and reliability of these methods, many linguists would like more transparency about what the algorithm is really doing.

3.3 Mapping demographic dispersal on linguistic phylogeny

Mapping demographic dispersal on linguistic phylogeny, we try to correlate expansive processes revealed by archaeological or genetic research with language split and spread, visible in language classifications and current linguistic geography. It can be expected that formative processes in population prehistory, such as those motivated by successful subsistence strategies, will shape language relationships. The prehistoric population movements out of Taiwan and through Island South-East

Asia into the Pacific discussed by Gray et al. (2009), for instance, display pulses and pauses that closely match the stages of splits and spreads in the phylogenetic tree of Austronesian languages.

Several chapters in this volume draw connections between demographic and linguistic processes. Schapper proposes to correlate the active population dynamics pulsing out of New Guinea at a time before the Austronesian migrations with the dispersal of the Trans-New Guinea languages. Van Driem associates the spread of the paternal lineage O in human genetics with the linguistic ancestors of the so-called “East Asian linguistic phylum”, which unites the Sino-Tibetan, Hmong-Mien, Austroasiatic, Austronesian and Kra-dai families. Robbeets proposes a scenario that links the developmental stages of agriculture and its effects on demographic transitions in southern Manchuria to the dispersal of the Transeurasian languages. Garnier et al. suggest that the strong population expansion of the Yamnaya culture around 4000 BC can be connected with the spread of the Indo-European languages through the favorable demography of herders having the unique capacity to digest animal milk in adulthood.

When mapping demographic dispersal on linguistic phylogeny, there is the pitfall of drawing a straightforward relationship between material culture, ethnic groups and language. However, instead of the conservative, static approach linking one monolithic archaeological culture to one mono-ethnic and mono-lingual group, this volume attempts to develop a more dynamic framework of inference whereby demographic processes are mapped on change in the archaeological record and these become in their turn associated with linguistic dispersals.

3.4 Cultural reconstruction

Cultural reconstruction, the investigation of the cultural vocabulary revealed in the reconstructed vocabulary of a proto-language is a major tool to investigate the correlations between language and farming and, therefore, frequently applied in this volume. It is a subfield of comparative historical linguistics that enables us to study human prehistory by correlating our linguistic reconstructions with information from archaeology about the possible cultural and natural environment of the speakers of the proto-language. As explained by van Driem in Chapter 7, the method was first introduced under the label “linguistic paleontology” by Adolphe Pictet (1859), who was inspired by Julius von Klaproth’s (1830: 112–113) pioneering work.

In addition to “cultural reconstruction” (Crowley & Bown 2010: 299; Epps 2015; Heggarty 2015) and “linguistic paleontology” (Hock 1991: 573–578), this method is also known as “Wörter und Sachen” (Campbell 2004 [1998]: 367–368) or “linguistic archaeology” (Southworth 2005). We also find terms such as “linguistic

ethnobiology” (Hunn & Brown 2011) or “paleobiolinguistics” (Brown 2015) in the literature, but this approach is more specifically directed at correlating linguistic reconstructions with archaeobotanical insights about plants.

Cultural reconstruction relies on two assumptions, specifically, first, that words and their meanings can be confidentially reconstructed to the proto-language and second, that reconstructed words allow us to make direct inferences about the nature of the ancient speech communities that used these words. Related to these assumptions is the inference that cultural items that have cognates widely spread across the languages in the family have existed in the associated cultures longer than items that lack such a wide distribution. In Chapter 6, for instance, Schapper observes striking linguistic similarities in terms for ‘sugarcane’ and ‘banana’ across widely dispersed groups of the Trans-New Guinea family. This enables her to reconstruct the terms back to proto-Trans-New Guinea and to infer that sugarcane and banana must have been part of the agricultural package possessed by early Trans-New Guinea populations. This situation contrasts with the distribution pattern of the word for ‘taro’, which can only be reconstructed to some low-level families and shows clear signs of later cultural diffusion.

Inventorying the reconstructed vocabulary in its entirety can contribute to a fuller picture of prehistory than the study of individual cultural reconstructions. Much information about the culture and society of the speakers of the proto-language can be recovered by paying attention to the clustering of different cultural items in a specific semantic domain or the unequal distribution of cognates in different semantic domains. In Chapter 3, for instance, Berge draws inferences on the basis of a gender difference in the distribution of Eskimo cognates in Aleut.

Among the limitations and challenges of cultural reconstruction are the potential lack of accuracy in semantic reconstruction, the occurrence of lexical recycling, the deception of a single item not backed up by a semantic domain and the shakiness of inferences made on the basis of absence.

3.4.1 *The accuracy of semantic reconstruction*

It is a fact that semantic reconstruction is less precise than phonological reconstruction. Therefore, we should be cautious not to be semantically overpermissive in our reconstructions. In Chapter 8, George Starostin suggests that a layer of agricultural lexicon may be reconstructable to the Sino-Caucasian macrofamily. However, some Sino-Caucasian agricultural reconstructions have rather ambiguous semantics. The Sino-Caucasian root $*\lambda w\check{z}wV$ ‘millet, rice’ reconstructed by Sergei Starostin (2005 <<http://starling.rinet.ru>>), for instance, is based on Sino-Tibetan $*liwH \sim *liwH$ ‘rice grain’ and North Caucasian $*\lambda w\check{z}wV$ ‘millet’, which in its turn involves a speculative semantic reconstruction as it is based on comparing the meaning ‘grain’ in Nakh, ‘mown crops’ in Lak, ‘bread’ in Lezghian and ‘millet’ in West Caucasian.

Since the meaning assigned to a reconstructed form can be no more specific than the meaning shared by all the cognate forms, the common denominator here is at best as concrete as ‘any plant used for consumption’. Building semantic reconstructions upon semantic reconstructions, our hypotheses risk collapsing like a house of cards.

By contrast, comparisons enjoying a high degree of semantic stability across different subgroupings of a language family may be particularly telling. The point is that when a particular meaning did not get replaced by a new meaning in the daughter languages, it is likely that the corresponding item or activity likewise did not get substituted by a newly introduced one. Such stable semantics appear in Schapper’s study of the meanings ‘banana’ and ‘sugarcane’, as well as in Starostin’s discussion of some North Caucasian agricultural reconstructions such as the verb ‘to thresh’.

3.4.2 *Lexical recycling*

“Lexical recycling” is a process whereby words with a general, non-cultural meaning become repurposed as words with a specific, cultural meaning after the importation or invention of the corresponding innovation. As a result, reconstructions with an agricultural meaning could have existed before the agricultural inventions with a non-agricultural meaning. In Aleut, for instance, the agricultural verbs ‘to plant’ and ‘to sow’ are recycled from hunter-gatherer terminology such as ‘to drop a fishing line’ and ‘to distribute sea-catch’, while in Proto-Quechua the verbs ‘to irrigate’ and ‘to sow’ are derived from ‘to fall (water), wet’ and ‘to hit, knock, push’. Names for domesticated crops often derive from their wild predecessors, as Bostoen and Koni Muluwa show for West-Coastal Bantu. Moreover, the names of agricultural imports may be derived from native domesticates, such as the development of rice agricultural vocabulary from dry crop vocabulary in Korean, discussed by Francis-Ratte. Savelyev finds that many pastoralist terms in Turkic are derived from non-pastoralist vocabulary in the proto-Turkic period, such as the derivation of ‘kid’ from ‘son, child’, or ‘dried quark, cheese’ from ‘to dry’. The same may be true for pastoralist terms in Indo-European as indicated by the reanalysis of a noun meaning ‘one who collects (liquids)’ into the Indo-European verb ‘to milk’, studied by Garnier et al. Joseph takes it one step further, not just analyzing the particular derivation or reanalysis of a single word, but trying to detect derivational patterns in the creation of agricultural vocabulary as a whole. He suggests that reduplication is commonly used as a strategy to extend previously non-agricultural vocabulary into agricultural vocabulary in proto-Indo-European.

3.4.3 *The deception of a single item*

If we can only reconstruct a single cultural item that is not backed up by other members of the semantic domain to which it belongs, there is reason for suspicion.

In this volume, Kümmel warns us of the deception of single items, pointing out that only very few grain terms in Indo-Iranian can be shown to be inherited from Indo-European, while pastoralist vocabulary is clearly inherited. This is taken as an indication that the spread of Indo-European can be motivated by pastoralism rather than by farming. In contrast, Schapper strengthens the argument that the spread of the Trans-New Guinea languages is driven by agriculture adding ‘sugarcane’ and ‘banana’ to the reconstructed package of crops, which so far consisted only of ‘taro’.

3.4.4 *The shakiness of inferences made on the basis of absence*

As the traditional aphorism goes, “absence of evidence is not evidence of absence.” The observation that an agricultural lexicon cannot be reconstructed for a certain proto-language may be explained by the fact that the proto-speakers simply were not familiar with farming, but it could also be due to the lack of the necessary exhaustive research or to the attrition of agricultural cognates over time. Therefore, inferences made on the basis of absence are not necessarily wrong, but they should not be taken as absolute proof for an argument.

In this volume, for instance, Robbeets maintains that common rice vocabulary is completely absent from Japano-Koreanic, while Francis-Ratte suggests a cognate for ‘dry rice’ on the basis of internal segmentation of some Middle Korean words. Given the presence of agricultural cognates in Transeurasian, Savelyev argues that the near absence of agricultural cognates found only in Altaic (i.e., Turkic, Mongolic and Tungusic) may be explained by the loss of agricultural terms, which may have swept away by or recycled as pastoralist terms. He supports this by contrasting the secondary or areal nature of pastoralist vocabulary with the absence of identifiable borrowings and the primary nature of agricultural terms in proto-Turkic. Similarly, assuming the presence of some agricultural vocabulary in Nostratic, Starostin proposes that traces of an earlier agricultural lexicon may have been lost in Uralic together with the practice itself, as former agriculturalists switched back to hunting-gathering.

3.5 Contact linguistics

A final set of tools at our disposal to determine the correlation between language and subsistence is offered by contact linguistics, the study of the ways in which languages influence each other when their speakers interact. The study of prehistoric borrowing and diffusion can be useful to shed light on past interactions and help us

determine the chronology of our data. If there is an exchange of loanwords between two or more languages, the assumptions are, first, that the speakers of the languages in question were directly in contact with each other either directly or indirectly through mediation of an intermediate population, and, second, that the loanwords cannot be dated to a time earlier than the established time of transmission of the relevant concept. In this volume, Savelyev argues that the borrowing of terms relating to horse pastoralism from proto-Turkic into proto-Mongolic must have taken place after 1200 BC, when horse-ridden pastoralism first appeared on the eastern steppes. Two of the challenges of contact linguistics are distinguishing between borrowing and inheritance, and determining the direction of the borrowing.

3.5.1 *The distinction between borrowing and inheritance*

If a word and its meaning correspond across various daughter languages, this does not necessarily imply that the word ultimately originated in the common ancestral language. It is quite possible that the word entered the relevant family by way of borrowing, either at the proto-stage or in a chain of transmissions after its break-up. The distinction between borrowing and inheritance in common subsistence vocabulary is therefore a serious concern, which is taken up in the chapters by Emlen and Adelaar, Berge, Savelyev, Schapper, van Driem and Kümmel. Criteria used in this volume to distinguish between borrowed and inherited items include the following.

1. When a given root corresponds beyond the presumed language family or a probable donor word exists in an unrelated language, borrowing is the most likely explanation. For instance, Schapper argues for diffusion of the term ‘taro’ across the Trans-New Guinea languages because reflexes of the form are found in numerous non-Trans-New Guinea languages as well.
2. The distributional pattern of borrowing is typically linear, progressing from one contact language into the other. Genealogical divergence, however, may be pictured as the rings formed when a stone is thrown into the water: innovations start in the center and push the older forms towards the periphery. Therefore, a distributional pattern whereby cognates leave traces in remote, unconnected areas is consistent with inheritance, but inconsistent with borrowing. In contrast to the term for ‘taro’, for instance, reflexes of the term for ‘sugarcane’ extend from the extreme east to the extreme west, with a gap in central New Guinea.
3. Correspondences between cultural items that show a remarkable semantic stability, whereby all reflexes of a certain protoform appear with exactly the same meaning as the protoform, are likely to be inherited. Borrowed items display more frequent semantic changes and substitutions than inherited cultural items do. This recalls Starostin’s findings about the semantic stability of the cognate verb ‘to thresh’ across the North Caucasian languages.

4. Borrowing is a likely explanation in cases when the similarity concerns a morphologically complex form in one language that cannot be analyzed as such in the other language. For example, Berge argues that Unangam Tunuu (Aleut) *angaaġu-x̂* ‘single-bladed paddle for skin boat’ is a borrowing from the Alutiiq (Eskimo) word *anguarun* ‘single bladed paddle’ because only the latter can be derived from *anguar-* ‘to row’.
5. Irregular sound correspondences are indicative of borrowing, an argument used by Kümmeel in his demonstration that the agricultural lexicon of Indo-Iranian is not inherited from Indo-European, but rather points to borrowing.
6. Correspondence sets that refer to innovations post-dating the proto-language split are arguably borrowings. For example, current findings that the kayak may have been a recent technological advance that reached the Aleutians within the past 1500 years supports Berge’s suggestion that all nominal correspondence sets related to the kayak, including the very term ‘kayak’ itself, are borrowings from neighboring Yupik languages to Unangam Tunuu (Aleut), rather than being inherited from Eskimo-Aleut.

3.5.2 *The directionality of the borrowing*

Especially in cases of prehistoric contact, it may be a challenge to determine the direction of the borrowing. One objection against van Driem’s proposal to regard Hmong-Mien as the source of borrowing for Sinitic rice agricultural vocabulary, for instance, comes from the observation that some of the alleged loans include characteristic Chinese morphology (Sagart 2011). Berge re-examines a list of probable borrowings of uncertain direction in Bergsland (1994: 655), supporting borrowing from Unangam Tunuu (Aleut) into Alutiiq or Yupik (Eskimo), rather than the other way around.

4. Organization of this volume

This volume is organized into 13 chapters, mostly case studies, reflecting on subsistence-based language spread on various continents around the world.

In Chapter 2, Nicholas Emlen and Willem Adelaar reconstruct proto-Quechua and proto-Aymara lexical items related to cultivation and herding to draw some inferences about the relationship between language and subsistence in the ancient Andes. Stripping away the many layers of Quechua-Aymara lexical borrowings, they find that the early speakers of both lineages were engaged in sophisticated cultivating and herding economies before their initial contact. Since both proto-languages exhibited terms for cultivation and herding at a wide range of ecological and elevational zones before their first contact, the early speakers appear to have

sustained contact across elevations and engaged in various subsistence practices. In spite of the presence of ancestral agropastoral vocabulary in both proto-languages, the authors question whether these families really owe their wide geographical range to the adoption of agriculture, pointing to the fact that the languages replaced the languages of pre-existing small-scale cultivators, rather than those of hunter-gatherers.

In Chapter 3, Anna Berge studies the motivation for the spread of Eskimo-Aleut languages after their split around 2000 BC. She pays special attention to the advance of Alutiiq (Eskimo) and the retreat of Unangam Tunuu (Aleut) in the Aleutian and Kodiak Islands around 500–1000 AD. To this end, she analyzes the distribution of Eskimo-Aleut cognates and Alutiiq borrowings in the subsistence terminology in Unangam Tunuu. She finds that agriculture was responsible neither for the original spread of Eskimo-Aleut, nor for the more recent instance of borrowing from and shift to Alutiiq in the previously Aleut region. Rather, the comparison of borrowing versus inheritance patterns suggests an influx of Alutiiq men, resulting in borrowing as well as language replacement as a result of warfare. Interestingly, in support of subsistence-driven language spread, prestige-triggered wars seem to have led to borrowing, while wars involving a struggle for insufficient resources seem to have led to replacement.

In Chapter 4, Alexander Francis-Ratte examines agricultural vocabulary shared between Japanese and Korean. In spite of the presence of various etymologies for ‘field’, Japanese and Korean share barely any words relating to rice agriculture. Proposing cognate sets for ‘rice’, ‘buckwheat’ and ‘millet’, Francis-Ratte suggests that Japanese and Korean may have diverged at a time when field rice was already being cultivated in Northeast Asia alongside millet, while paddy rice was not introduced yet. He further proposes that pre-rice vocabulary has undergone a process of lexical recycling in Korean to refer to later rice-related practices.

In Chapter 5, Martine Robbeets investigates to what extent agriculture impacted the dispersal of the Transeurasian language family, i.e. the genealogical grouping consisting of the Turkic, Mongolic, Tungusic, Koreanic and Japonic languages. In addition to disagreeing on their genealogical relatedness, previous scholarship has called into question the claim of agriculture-driven language spread for these languages. Applying techniques such as the diversity hotspot principle, phylolinguistics, mapping demography on linguistic phylogeny and cultural reconstruction, Robbeets finds indications that proto-Transeurasian was spoken by people gradually adopting farming and that its dispersal was indeed driven by agriculture.

In Chapter 6, Alexander Savelyev compares the origin of farming-related and pastoralism-related vocabulary across the Altaic (i.e., Tungusic, Mongolic and Turkic) languages with special attention to the developments in Turkic. He finds that in proto-Turkic, pastoralist vocabulary can often be shown to result from

secondary derivation or borrowing, whereas agricultural terms include more primary roots and cannot easily be identified as borrowings. On the basis of this observation, he explains the limited reconstructability of agricultural vocabulary in Altaic as opposed to Transeurasian, by a loss of agricultural terms after the break-up of Altaic, whereby pastoralist terms were borrowed or recycled from preexisting agricultural terms.

In Chapter 7, Antoinette Schapper investigates whether the Trans-New Guinea Phylum, a language family comprising a large number of the languages of New Guinea that remains largely untested by the traditional methods of historical comparative linguistics, can be considered to be an instance of Farming/Language Dispersal. In addition to previous comparative research focussing on taro, she compares the terms for two different crops, sugarcane and banana across the Trans-New Guinea languages. Stressing the great cultural and economic importance of these crops throughout the Papuan language area, she proposes linguistic evidence that not taro but rather banana and sugarcane were associated with the expansion of the Trans-New Guinea languages.

Challenging the traditional view of a single domestication of rice in the Yangtze River Basin in Chapter 8, George van Driem brings together linguistic, archaeobotanical and genetic evidence supporting three separate domestication events. He associates the paternal lineage O in human genetics with the linguistic ancestors of the so-called “East Asian linguistic phylum”, which unites the Sino-Tibetan, Hmong-Mien, Austroasiatic, Austronesian and Kradai families. He suggests that at least two of these families, Austroasiatic and Hmong-Mien, owe their wide distribution to their involvement in rice domestication events, the former in the Brahmaputra Valley area and the latter located further east, south of the Yangtze River.

In Chapter 9, George Starostin surveys some of the more developed hypotheses on Eurasian macrofamilies such as Nostratic, Sino-Caucasian and Afroasiatic and examines whether agricultural vocabulary can be reconstructed back to the ancestral languages. He concludes that the most convincing case of an early linguistic stock with a reconstructible layer of agricultural lexicon is the Western subdivision of Sino-Caucasian. This follows from his observation that agricultural terminology can be convincingly reconstructed to proto-North Caucasian and from the existence of plausible Euskaro-Caucasian connections in the agricultural lexicon, which suggests that the original speakers of Basque once dwelled in close proximity to speakers of North Caucasian languages. In this connection, he points to the possible Caucasian origins of some of the substrate lexicon, found in various branches of the Indo-European languages across Europe. He further finds that evidence of ancient agricultural lexicon in the Afroasiatic stock remains at best circumstantial, whereas evidence of early agricultural vocabulary in Nostratic is completely lacking.

In Chapter 10, Koen Bostoen and Joseph Koni Muluwa question the plausibility of agriculture as the main driving force behind the initial Bantu Expansion. Instead, they propose that the early language spread was facilitated through climate-induced openings of the Central African rainforest block. The first Bantu-speaking populations that, following savannah corridors, arrived south of the rainforest were the West-Coastal Bantu speakers. Bostoen and Koni Muluwa review subsistence-related plant-vocabulary that can be reconstructed in Proto-West-Coastal Bantu to assess the question of whether the Bantu speakers had become farmers by the time that they reached the area south of the rainforest. They find that even if the first Bantu speakers south of the rainforest knew how to cultivate certain crops, they were still largely dependent on plant resources that they could collect in their natural environment. As the West-Coastal Bantu speakers were only gradually moving from foraging to plant cultivation to domestication, the emergence of agriculture in early Bantu speech communities is characterized as a slow revolution.

Using examples from Indo-European historical comparison in Chapter 11, Brian Joseph reviews the methods by which we infer that the lexicon of a certain proto-language contains agricultural items. In addition to paleolinguistics, including cultural reconstruction, etymological derivation and loanword detection of lexical items relating to agriculture, he proposes two further types of lexically based argumentation. The first type reconstructs derivational processes involved in the creation of agricultural words and their meanings, such as for instance a process of reduplication that is found to be a productive strategy in the derivation of agricultural vocabulary in Indo-European. The second type of argumentation examines the embedding of agricultural vocabulary into the religious practices and mythological tales associated with early Indo-European culture. In this way, he proposes to expand our methodology of examination of agricultural vocabulary to the larger word-formational patterns and cultural context of the words involved.

Comparing pastoralist to agricultural reconstructions in Chapter 12, Martin Kümmel makes inferences about the significance of farming for the spread of the Indo-Iranian languages. He finds that pastoral terminology, such as words for cattle, horses, sheep and goats is clearly inherited from Indo-European. This is in contrast to the lack of genealogical continuity for plant cultivation terms, such as words for cereals, pulses and vegetables, which reflect several layers of loanwords. Observing that the agricultural terminology of Indo-Iranian is largely divergent from that of most European branches of Indo-European, Kümmel argues that the Indo-Iranian languages have mainly spread through pastoralism.

Finally, in Chapter 13, Sagart, Garnier and Sagot reconcile the idea of pastoralist and subsistence-driven language spread by associating the spread of the Indo-European languages with the origins of dairying. To this end, they bring together archaeological, genetic, ethnographic and linguistic evidence. Their observations

give additional support to the Pontic Steppe hypothesis that identifies the ancestral group of proto-Indo-European speakers with pastoralists in the steppes north of the Black Sea around 4000 BCE. Examining reconstructed Indo-European dairying vocabulary in addition to ancient texts, they find evidence that the ancestral speakers of Indo-European were the first in Eurasia to develop the ability to drink milk in adulthood, which conveyed a serious advantage in subsistence. As a result of boosting demography, lactase persistence increased the need for pasture land and is thus thought to have driven the expansion of the Indo-European languages.

5. Findings

Like the three aspects of a crime that must be established to prove guilt, language spread usually involves an opportunity, a means and a motive. The opportunity has to do with the conditions of the time and space in which the proto-language is situated and over which the ancestral speakers have little or no control. Conditions that may invite speakers to spread include outside population pressure, disease, volcanic activity, climate change, vegetation or other ecological change, etc. The initial Bantu expansion, for instance, was facilitated by climate-induced openings of the Central African rainforest and the separation of the Transeurasian languages was triggered by climate change.

The means refers to the force or the instrument that drives the spread. Advantages in transport, weaponry and state organization are what empower speech communities to spread and to dominate other communities. For instance, as discussed in this volume, increased mobility through horse riding was instrumental in the spread of the Turkic, Mongolic and the Indo-European languages, while an advantage in weaponry was a major factor in the spread of Alutiiq (Eskimo).

Finally, language dispersal also requires a motive, a mechanism that causes the dispersal. Among the mechanisms proposed by Renfrew (1987: 123–131) are (a) demography/subsistence, (b) elite dominance and (c) system collapse, but not all of these mechanisms have an equal likelihood of causing language shift and replacement. In fact, elite dominance, whereby the incomers are demographically insignificant relative to the local population, is rarely seen to cause shift. When a dominating group is relatively small in comparison to the dominated speech community, the expected outcome of language contact is instead language maintenance with borrowing (Thomason & Kaufman 1988; Heggarty 2015). This is supported by historical cases of elite dominance, such as the Normans leaving an extensive layer of loanwords in English, without ever replacing English with French in Britain. Similarly, Berge finds that the language of the Alutiiq (Eskimo) elite heavily influenced Unangam Tunuu (Aleut) spoken on the Aleutian Islands but did not replace

it. However, as illustrated in this volume, cases of elite dominance can involve language replacement, especially when the elites benefit from a particularly favorable opportunity or have an acute advantage in means. Examples in our volume include Eskimo replacing Aleut on Kodiak Island or Turkic and Indo-European replacing pre-existing local languages. In addition to a crucial advantage in means for transportation or warfare, these language shifts may also have been facilitated by biological advantages such as immunity to diseases like the plague or lactase persistence. By definition, the elite are a small group of persons who exercise influence over a larger one, but these physical advantages may have allowed the elite to survive an event that decimated the local population, thus providing a favorable demography for language shift. In addition, the resource surpluses on Kodiak Island and dairying among the Indo-Europeans suggest that subsistence played a role as well. Therefore, these cases seem to be situated at the interface of the Subsistence/Demography and the Elite Dominance model.

The contributions to this volume relativize the importance of agriculture as a motive for language spread by showing that Farming/Language Dispersal is just one instantiation of the Subsistence/Demography model and by viewing subsistence regimes and the reconstructed agricultural lexicon in which they are mirrored as a continuum rather than a discrete division.

Some language families such as Eskimo-Aleut have no farming, but subsistence played a role in their development in that the language spoken by the population gaining access to the food resources replaced the pre-existing language spoken by the population losing access. Other families such as Turkic and Indo-European may have been familiar with farming but their spread was caused by food surpluses and mobility associated with horse-ridden pastoralism. Yet other language families such as West-Coastal Bantu and Transeurasian initially occupied a middle ground between farming and foraging. Next, there are families such as Quechua, Aymara, Japano-Koreanic and Trans-New Guinea, which demonstrably had agriculture, but replaced pre-existing languages of populations that were already familiar with farming, be it on a smaller scale. An indisputable case of Farming/Language Dispersal in this volume may be represented by Austroasiatic, but even here controversy remains about the homeland and whether rice indeed was the original crop (Sidwell & Blench 2011). Therefore, the more general Subsistence/Demography model seems to be more widely applicable than the Farming/Language Dispersal Hypothesis. The key issue is an advantage in subsistence strategy and thus expansive potential – be it related to foraging, farming or pastoralism – that eventually makes the incoming population demographically more successful than the local one.

Moreover, considering that the transition to an agricultural lifestyle must have taken place over centuries, if not millennia, including a lengthy pre-domestication stage, we find that a dualistic concept whereby a subsistence regime is either

agricultural or not is not tenable and neither is the characterization of a proto-language as either having or lacking agricultural vocabulary. As shown in Figure 3, our contributions suggest a continuum distribution, whereby some proto-languages such as Eskimo-Aleut completely lack agricultural vocabulary, others like Indo-European languages inserted agricultural vocabulary from the languages they supplanted, while yet others such as Aleut and possibly Hmong-Mien borrowed terms for agricultural innovations in their lexicons. Families such as Transeurasian and West-Coastal Bantu then, represent a transitional stage between foraging and farming, cultivation and domestication. Even if such families as Japano-Koreanic, Quechua, Aymara, Trans-New Guinea and Austroasiatic clearly reflect an agricultural lexicon, this does not necessarily imply that the language spread is driven by agriculture alone.

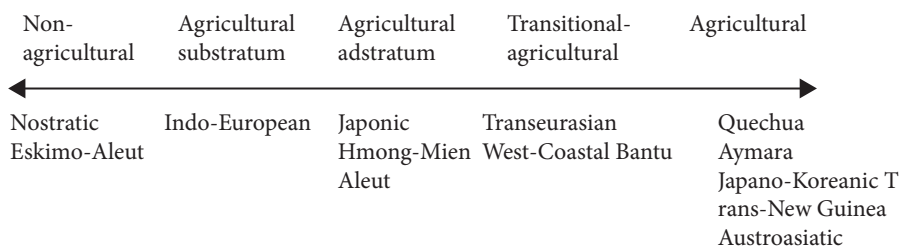


Figure 3. A continuum-distribution for agricultural lexicon discussed in this volume

In sum, farming is not a magic wand that can be waved to explain all instances of language spread, but Farming/Language Dispersal remains a useful working hypothesis because especially in Neolithic times, when human societies tended to be smaller in size and less complex in technology, the transition to farming must have held the promise of a better life. Thinking more broadly of farming as a relatively successful subsistence strategy involving potential for demographic growth and assessing language spread in terms of the three aspects of a crime – opportunity, means and motive – may help us to abandon one-factor explanations and consider many more factors that stimulated linguistic expansion.

References

- Aikhenvald, Alexandra. 1999. The Arawak language family. In *Amazonian Languages*, Alexandra Aikhenvald & Robert Dixon (eds), 65–106. Cambridge: CUP.
- Bellwood, Peter. 2005. *First Farmers: The Origins of Agricultural Societies*. Malden MA: Blackwell.
- Bellwood, Peter. 2011. *First Migrants: Ancient Migration in Global Perspective*. Malden MA: Blackwell.

- Bellwood, Peter & Dizon, Eusebio. 2008. Austronesian cultural origins: Out of Taiwan, via the Batanes Islands, onwards to Western Polynesia. In *Past Human Migrations in East Asia, Matching Archaeology, Linguistics and Genetics*, Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross, Ilia Peiros & Marie Lin (eds), 23–39. London: Routledge.
- Bellwood, Peter & Renfrew, Colin (eds). 2002. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- Bergsland, Knut. 1994. *Aleut Dictionary: Unangam Tunudgusii*. Fairbanks: ANLC.
- Blust, Robert. 1995. The prehistory of the Austronesian-speaking peoples: A view from language. *Journal of World Prehistory* 9: 453–510. doi:10.1007/BF02221119
- Blust, Robert. 2013. *The Austronesian Languages* [Asia-Pacific Linguistics. Open Access Monographs]. Canberra: Australian National University.
- Brown, Cecil. 2015. Paleobiolinguistics of New World crops and the Otomanguean language family. *Ethnobiology letters* 6(1): 189–191. doi:10.14237/eb1.6.1.2015.436
- Brown, Cecil *et al.* 2013a. The paleobiolinguistics of domesticated chili pepper (*Capsicum* spp.). *Ethnobiology Letters* 4: 1–11. doi:10.14237/eb1.4.2013.2
- Brown, Cecil *et al.* 2013b. The paleobiolinguistics of domesticated manioc (*Manihot esculenta*). *Ethnobiology Letters* 4: 61–70. doi:10.14237/eb1.4.2013.5
- Brown, Cecil *et al.* 2014a. The paleobiolinguistics of domesticated maize (*Zea mays* L.). *Ethnobiology Letters* 5: 52–64. doi:10.14237/eb1.5.2014.130
- Brown, Cecil *et al.* 2014b. The paleobiolinguistics of the common bean (*Phaseolus vulgaris* L.). *Ethnobiology Letters* 5: 104–115. doi:10.14237/eb1.5.2014.203
- Campbell, Lyle. 2004[1998]. *Historical Linguistics: An Introduction*. Edinburgh: EUP.
- Comrie, Bernard. 2002. Farming dispersal in Europe and the spread of the Indo-European language family. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 409–420. Cambridge: McDonald Institute for Archaeological Research.
- Crowley, Terry & Bower, Claire. 2010. *An Introduction to Historical Linguistics*. Oxford: OUP.
- Diakonoff, Igor. 1998. The earliest Semitic society. *Journal of Semitic Studies* 43(2): 209–219. doi:10.1093/jss/43.2.209
- Diamond, Jared & Bellwood, Peter. 2003. Farmers and their languages: The first expansions. *Science* 300: 597–603. doi:10.1126/science.1078208
- Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 77–80. London: Routledge Curzon. doi:10.4324/9780203343685_chapter_5
- Dunn, Michael. 2015. Language phylogenies. In *The Routledge Handbook of Historical Linguistics*, Claire Bower & Bethwyn Evans (eds), 190–211. London: Routledge.
- Ehret, Christopher. 2003. Language family expansions: Broadening our understandings of cause from an African perspective. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 373–400. Cambridge: McDonald Institute for Archaeological Research.
- Epps, Patience. 2015. Historical linguistics and socio-cultural reconstruction. In *The Routledge Handbook of Historical Linguistics*, Claire Bower & Bethwyn Evans (eds), 579–597. London: Routledge.
- Fuller, Dorian. 2002. An agricultural perspective on Dravidian historical linguistics: Archaeological crop packages, livestock and Dravidian crop vocabulary. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 191–214. Cambridge: McDonald Institute for Archaeological Research.

- Gray, Russell & Jordan, Fiona. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature*: 1052–1055. doi:10.1038/35016575
- Gray, Russell & Atkinson, Quentin. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426: 435–439. doi:10.1038/nature02029
- Gray, Russell, Drummond, Alexei & Greenhill, Simon. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913): 479–83.
- Heggarty, Paul. 2015. Prehistory through language and archaeology. In *The Routledge Handbook of Historical Linguistics*, Claire Bowern & Bethwyn Evans (eds), 598–626. London: Routledge.
- Higham, Charles. 2002. Languages and farming dispersal: Austroasiatic languages and rice cultivation. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 223–232. Cambridge: McDonald Institute for Archaeological Research.
- Hock, Hans Heinrich. 1991. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter. doi:10.1515/9783110219135
- Hunn, Eugene & Brown, Cecil. 2011. Linguistic ethnobiology. In *Ethnobiology*, Anderson *et al.* (eds), 319–333. Oxford: Wiley-Blackwell. doi:10.1002/9781118015872.ch19
- Janhunen, Juha. 1996. *Manchuria: An Ethnic History* [Mémoires de la Société Finno-Ougrienne 222]. Helsinki: Suomalais-Ugrilainen Seura.
- Kaufman, Terence. 1990. Early Otomanguean homelands and cultures: Some premature hypotheses. *University of Pittsburgh Working Papers in Linguistics* 1: 91–136.
- LaPolla, Randy. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, Alexandra Aikhenvald & Robert Dixon (eds), 225–254. Oxford: OUP.
- Militarev, Alexander. 2002. The prehistory of a dispersal: The Proto-Afrasian (Afroasiatic) farming lexicon. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 135–150. Cambridge: McDonald Institute for Archaeological Research.
- Ostapirat, Weera. 2005. Kra-dai and Austronesians. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 107–131. London: Routledge Curzon. doi:10.4324/9780203343685_chapter_7
- Pawley, Andrew. 2002. The Austronesian dispersal: languages, technologies and people. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 251–274. Cambridge: McDonald Institute for Archaeological Research.
- Philipson, David. 2002. Language and farming dispersals in sub-Saharan Africa, with particular reference to the Bantu-speaking peoples. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 177–190. Cambridge: McDonald Institute for Archaeological Research.
- Pictet, Adolphe. 1859. *Les origines indo-européennes ou les aryas primitifs: Essai de paléontologie linguistique (première partie, seconde partie)*. Paris: Joël Cherbuliez.
- Ratliff, Martha. 2004. Vocabulary of environment and subsistence in the Hmong-Mien protolanguage. In *Hmong/Miao in Asia*, Nicholas Tapp, Jean Michaud, Christian Culas & Gary Yia Lee (eds), 147–165. Chiang Mai: Silkworm Books.
- Renfrew, Colin. 1987. *Archaeology and Language: The Puzzle of Indo-European Origin*. London: Jonathan Cape.
- Sagart, Laurent. 2008. The vocabulary of cereal cultivation and the phylogeny of East Asian languages. *Bulletin of the Indo-Pacific Prehistory Association* 23: 127–36. doi:10.7152/bippa.v23i0.11858
- Sagart, Laurent. 2011. How many independent rice vocabularies in East Asia? *Rice* 4(3): 121–133. doi:10.1007/s12284-011-9077-8

- Sapir, Edward. 1916. *Time Perspective in Aboriginal American Culture: A Study in Method* [Canada Department of Mines, Geological Survey, Memoir 90. Anthropological Series 13.] Ottawa: Government Printing Bureau. doi:10.4095/103486
- Sidwell, Paul & Blench, Roger. 2011. The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, Nick Enfield (ed.), 315–343. Canberra: Pacific Linguistics.
- Southworth, Franklin. 2005. *Linguistic Archaeology of South Asia*. London: Routledge. doi:10.4324/9780203412916
- Starostin, Sergei. 2005. Sino-Caucasian etymological database, <<http://starling.rinet.ru>> (4 March 2017).
- Thomason, Sarah Grey & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- von Klaproth, Julius Heinrich. 1830. Réponse à quelques passages de la préface du roman chinois intitulé: Hao khieou tchhouan, traduit par M.J.F. Davis. *Journal Asiatique* 5: 97–122.

Proto-Quechua and Proto-Aymara agropastoral terms

Reconstruction and contact patterns

Nicholas Q. Emlen and Willem F. H. Adelaar

Leiden University Centre for Linguistics

This chapter presents reconstructed Proto-Quechua and Proto-Aymara lexical items related to cultivation and herding, and draws conclusions about language and subsistence in the ancient Andes. The patterns of lexical borrowing between the two lineages offer a novel empirical perspective on how early Quechuan and Aymaran speakers lived. When the many layers of borrowing are stripped away, it is clear that both were engaged in agropastoral economies before the languages first came into contact. Furthermore, the presence of terms from a wide range of ecological zones, from the high grasslands to (in the case of Quechua) the tropical lowlands, suggests that both languages cross-cut elevations in a manner consistent with the typically Andean system of ecological complementarity.

Keywords: Quechua, Aymara, Andes, agropastoralism, language contact

1. Introduction

The Quechuan and Aymaran languages are spoken by millions of people across a vast expanse of the Central Andean region. Both families are closely associated with agriculture and pastoralism, and the Central Andes is one of the few regions on Earth where these modes of subsistence – as well as the complex social formations that they support – developed independently.

Given these facts, it is of interest to know what the relationship might have been between agropastoralism and the early history of the Quechuan and Aymaran lineages. The wide geographical distribution of both families, for instance, makes them candidates for consideration within the Farming/Language Dispersal Hypothesis, which proposes that language families expand when “farmers and their culture replace neighboring hunter-gatherers and the latter’s culture” (Diamond & Bellwood 2003: 598). However, the Andean case does not constitute a straightforward test of

that hypothesis: the Aymaran and Quechuan families first expanded around one or two millennia BP into landscapes that had already been occupied by herders and cultivators for thousands of years. Indeed, a broad range of domesticated animals and plants, agropastoral practices, and farming and herding technologies were already in place in the Andean highlands well before those expansions, including camelid herding by 5500 years BP (Pearsall 2008; Wheeler 1995); maize by 3600 to 4000 calibrated years BP (Perry et al. 2006; see also Tykot et al. 2006); and irrigation by 3500 years BP (Zimmerer 1995). It is no surprise, then, that many of the languages with which the Quechuan and Aymaran families came into contact during their initial dispersals already had agricultural lexicons. Regarding this poor fit between the time depths of the emergence of agropastoralism (3500–5500 BP) and the Quechuan and Aymaran dispersals (1000–2000 BP), Heggarty and Beresford-Jones (2010) argue that the extreme diversity of Andean environments delayed the intensification of agriculture – and thus, the attendant linguistic expansions – until later.

However, there are other ways of approaching these questions beyond merely correlating the respective time depths of the advent of agropastoralism and the Quechuan and Aymaran dispersals. In this chapter, we use reconstructed Proto-Quechua and Proto-Aymara lexical items related to cultivation and herding to draw some conclusions about the kinds of subsistence activities practiced by speakers of those languages. Indeed, fully developed vocabularies for the crops, animals, techniques, tools, and products associated with cultivating and herding constitute evidence that the speakers of those languages engaged in these practices; thus, reconstructions of these lexical domains afford a perspective on how the early speakers of these languages might have lived.

This endeavor is greatly complicated by the multilayered history of contact between the Quechuan and Aymaran languages, which resulted in intense lexical borrowing and profound structural convergence (for summaries, see Adelaar 2012a, 2012b). This contact began before the respective proto-language stages, which requires us to consider hypothetical periods before the first contact: Pre-Proto-Quechua and Pre-Proto-Aymara. As much as a third of the Proto-Aymara lexicon may have been borrowed from Pre-Proto-Quechua during this first contact (Emlen 2017); thus, before the early lexicons of both linguistic lineages can be adequately characterized, it is first necessary to identify and strip away the layers of borrowing between them. The reconstructions presented in this chapter are part of a larger effort to disentangle these contact influences, and to reveal what Pre-Proto-Quechua and Pre-Proto-Aymara might have been like before their first contact (Adelaar 1986; Emlen 2017; Emlen to appear).

To be sure, the complexity of this language contact situation makes interpreting any aspect of the ancient Andean linguistic panorama a daunting task indeed. However, the patterns of borrowing themselves may offer a novel empirical vantage

point on this issue. For if one or the other linguistic lineage had a privileged association with farming or herding, or with a particular crop or ecological zone, then we would expect that language to be a source of borrowing for terms regarding those practices. This borrowing might have taken place within the Quechua-Aymara relationship itself, as well as with other languages in the region. On the other hand, if the early speakers of both the Quechuan and Aymaran lineages were already engaged in herding and cultivating economies before their first encounter, then we would expect each lineage to exhibit a full range of relatively independent – that is, non-borrowed – terminology related to those practices. Furthermore, if the lexicons of both proto-languages include separate terms for domesticates found in a variety of different ecological zones (along with their associated techniques, tools, products, etc.), then we can be confident that speakers of both languages accessed land in those zones. This would be consistent with the vertically distributed system of land-holding typical of Andean societies, whereby social groups herd and cultivate on land at a variety of elevations – often discontinuously – to support different kinds of crops and domesticated animals (Murra 1972). In fact, as will be shown in this chapter, this is what we find: when the many layers of Quechua-Aymara lexical borrowing are stripped away, it becomes clear that the early speakers of both lineages were engaged in sophisticated cultivating and herding economies from the high, wind-swept grasslands above 4000 meters; to the lush intermontane valleys above 2300 meters; and, in the case of the Quechuan lineage, perhaps into the tropical lowlands below 1600 meters.

In this manner, the examination of the lexicons of each proto-language may also help clarify some unresolved issues regarding the prehistoric linguistic dynamics of the Central Andes. First, if Pre-Proto-Quechua and Pre-Proto-Aymara were both distributed across social networks spanning ecological and elevational zones (perhaps discontinuously), this might suggest a sociolinguistic ecology in which languages were interspersed across the landscape rather than representing blocks on the map (this would be similar to the situation during the Inka period in Southern Peru, described by Mannheim 1991). This scenario would help explain the complex and gradient patterns of historical contact effects among the Andean languages, and it would require conceptualizing linguistic contacts and continuities that straddle different elevations and environments from the highlands to the lowlands (a common pattern in the region; see Emlen 2016).

Second, knowing what kinds of economic activities were practiced by the speakers of the Quechuan and Aymaran lineages before their initial contact might shed light on the sociolinguistic circumstances of that contact. As Muysken (2011) notes, the particular contact effects that emerged between the two lineages must be understood as the outcome of a particular political-economic encounter – involving, for instance, dominance, prestige, language shift or maintenance, or some

other type of sociolinguistic relationship – which may have correlates in the archaeological record. Information about the subsistence activities and elevational distributions of each group before their first contact would certainly be relevant to identifying this scenario.

Third, this approach offers a separate line of evidence regarding the aforementioned proposal of Heggarty and Beresford-Jones (2010) that the intensification of maize cultivation by Aymaran speakers was the ultimate cause of that family's dispersal across the Andes during the Early Horizon. If this was the case, we might expect Aymaran maize terms, and the techniques and products of maize cultivation, to have been borrowed into the languages with which the Aymaran family came into contact during its expansion (including Quechuan languages). However, it appears that the neighboring Andean languages already had vocabularies related to maize cultivation before their contact with Aymaran languages, and in the cases in which such terms are borrowed, they often come from Quechuan languages. These observations do not necessarily contradict Heggarty and Beresford-Jones' proposal, but they do suggest a more complex picture that might be clarified if we examine the kinds of subsistence activities that are encoded in the early lexicons of each linguistic lineage.

This chapter begins with a brief introduction to the history of the Quechuan and Aymaran lineages (Section 2), with a special focus on the multilayered contact between them. Our reconstructions of the agricultural and pastoral lexicons of Proto-Quechua and Proto-Aymara are presented in Section 3, including a brief discussion of the apparently innovative character of some of the Proto-Quechua terms. We conclude with some comments about these findings and their implications for the relationship between agropastoralism and the early Quechuan and Aymaran lineages.

2. The Quechua-Aymara relationship

Before describing the place of agricultural and pastoral terminology within the early history of the Quechuan and Aymaran languages, it is first necessary to present a concise historical summary of those linguistic lineages and the contacts between them. This is a very complex language contact situation, both because of the profound transformations that both lineages underwent as a result of their initial contact, and because various Quechuan and Aymaran languages have subsequently come into contact in other places throughout their long shared history. Thus, any question regarding the early Quechuan and Aymaran lineages must be answered within a framework that accounts for this contact.

The Quechuan and Aymaran families each comprise a group of closely related languages spoken by millions of people across a vast and overlapping expanse of the Central Andean region (for thorough introductions to these families, see Adelaar & Muysken 2004: 179–319; Cerrón-Palomino 1987; Cerrón-Palomino 2000). Varieties of Quechua are found more or less continuously from Southern Colombia in the north to Bolivia, Northern Argentina, and Northern Chile in the south; they are also found far into the Amazonian lowlands east of the Andes, and they were attested on the Peruvian coast until the colonial period. The Aymaran family comprises two surviving branches: the Southern Aymaran languages, spoken in Southern Peru, Bolivia, and Northern Chile, and the Central Aymaran languages, spoken in a few villages in the Department of Lima in Central Peru. Aymaran languages were probably also spoken further north, as attested anecdotally (Hardman 1966: 15), by the ubiquity of Aymaran toponymy in the Central Peruvian highlands, and by post-dispersal Aymaran loans in the Quechuan languages spoken there (see also Cerrón-Palomino 2008b). Furthermore, the Quechuan and Aymaran lineages underwent early contact before their dispersal across the Central Andes; and since Quechua appears to have spread from Central Peru, the ancestor of the Aymaran family must have been spoken there as well (for more, see Adelaar 2012a; Cerrón-Palomino 2000; Emlen 2017).

The Quechuan and Aymaran families are both relatively shallow – perhaps comparable in scope and time depth to the Romance languages, or slightly less (Heggarty & Beresford-Jones 2010: 172). Thus, a reasonable subjective estimate for the Proto-Quechua and Proto-Aymara stages and subsequent dispersals is 1000–2000 years BP. Both families appear to have dispersed from Central Peru. The comparative reconstruction of Proto-Quechua (Cerrón-Palomino 1987) and Proto-Aymara (Cerrón-Palomino 2000) does not present major problems; a more vexing challenge for scholars of Andean linguistics has been accounting for the great number of resemblances between the Quechuan and Aymaran languages. The languages share a substantial proportion of their basic and non-basic lexicons (15–30%, by most accounts); their phonemic inventories are nearly identical; and their heavily agglutinating morphosyntactic structures exhibit notable structural isomorphism (Cerrón-Palomino 2008a), though most of the grammatical morphemes themselves are different in form. Furthermore, some Quechuan and Aymaran varieties that share overlapping territories in Southern Peru and Bolivia exhibit similar series of glottalized and aspirated consonants (e.g. Mannheim 1991), including in many lexical items that are shared by both families (Emlen 2017: 324–332).

These resemblances have led some scholars to advocate for a Quechua-Aymara (or Quechumara) genetic grouping (e.g. Orr and Longacre 1968), a notion that has been entertained since at least the 17th century (see Cerrón-Palomino 2000 for a thorough overview). However, as linguists began to conduct systematic descriptive

and comparative studies of Quechuan and Aymaran languages beginning in the 1960s, consensus emerged that many of the resemblances between the families were better explained as the product of intense language contact. Of course, this does not rule out the possibility that a deeper genetic grouping can eventually be discerned once the contact influences are accounted for (Adelaar 1986; Campbell 1995; Emlen 2017).

2.1 Pre-Proto-Quechua and Pre-Proto-Aymara

One of the biggest problems for interpreting the Quechua-Aymara relationship is the fact that all of the Quechuan and Aymaran languages exhibit the effects of their mutual contact; there are no (known) languages from either family that have developed outside of that contact. In other words, the earliest stages of Proto-Quechua and Proto-Aymara that can be reconstructed through comparison of their respective daughter languages existed after the first contact between the lineages had already taken place. This situation requires that we look even further back in both lineages, to the periods before the initial contact, to what Cerrón-Palomino (2000) and Adelaar (2012a) (among others) call Pre-Proto-Quechua and Pre-Proto-Aymara (note, however, that these were not necessarily static languages, but rather hypothetical periods before the first moment of contact; see Emlen 2017: 308). This also requires that we make a clear distinction between two periods of contact: those that took place between the two pre-proto-languages, before the stages of the proto-languages – what Adelaar (2012b) calls the “initial convergence” – and the subsequent “local convergences” that took place among individual Quechuan and Aymaran languages, after those families ramified and dispersed across the region. These terms will be used throughout this chapter.

The initial convergence probably took place a relatively short time before the proto-language stages of each family, since most of the roots borrowed during this time remained phonologically identical, or nearly identical, in Proto-Quechua and Proto-Aymara. Thus, if the proto-languages can be subjectively dated at one or two millennia BP, the initial convergence between Pre-Proto-Quechua and Pre-Proto-Aymara may have taken place around 1500–2500 years BP. If the Quechuan and Aymaran lineages do in fact descend from a common ancient language, it would have existed earlier than this period (perhaps much earlier); however, little evidence of such a connection remains once the contact influences of the initial convergence are taken into account. Of course, these figures should be taken as ballpark estimates, since the comparative method generates relative rather than absolute chronologies. Figure 1 gives a simplified graphic representation of this

history (dotted lines indicate known instances of language contact). Note that the image in Figure 1 is not to scale.¹

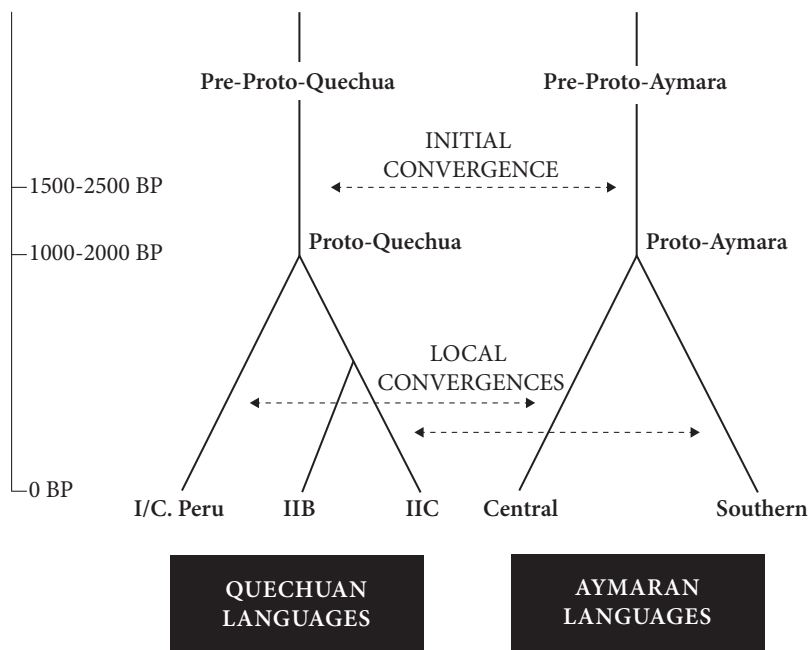


Figure 1. Simplified history of the Quechuan and Aymaran lineages

The directionality of influence during the initial convergence appears to have been asymmetrical: Pre-Proto-Aymara took on a large quantity of Quechuan loans at this point, including non-basic and basic vocabulary such as the numerals **kimsa* ‘three’ and **piçqa* ‘five’. At the same time, the morphosyntax and perhaps the phonology of Pre-Proto-Quechua were reformatted on the Aymaran template (Adelaar 2012b; Emlen to appear; Muysken 2011). Both of these processes suggest a situation of stable, intimate, and possibly long-term multilingualism.

In order to understand the prehistoric dynamics of agriculture and pastoralism in the Andes, we must focus on the earliest discernible stages of each lineage: Pre-Proto-Quechua and Pre-Proto-Aymara. This requires disentangling the history of borrowing between the two lineages – both during the initial convergence and the subsequent local convergences – in order to clarify what their early lexicons

1. In the Quechuan diagram, the terms I, IIB, and IIC refer to branches identified by Torero (1964). C. Peru refers to the Quechuan varieties of Central Peru that do not fit easily into a branching representation of the family.

might have been like. To this end, Adelaar (1986) proposes that three categories of lexical items can be isolated within the Proto-Quechua and Proto-Aymara lexicons: (a) non-shared Proto-Quechua roots, which are attested across the Quechuan family but unattested in Aymaran languages; (b) shared roots, which can be reconstructed in both proto-languages; and (c) non-shared Proto-Aymara roots, which are attested across the Aymaran family but unattested in Quechuan languages. All things being equal, the non-shared roots in categories (a) and (c) are most likely to descend from Pre-Proto-Quechua and Pre-Proto-Aymara (respectively), and to retain the phonological characteristics of those pre-proto-languages. These phonological characteristics can then be used as diagnostic features to determine the provenance of some of the shared roots in category (b). Much of the Proto-Quechua and Proto-Aymara lexicons can be sorted accordingly. Emlen (2017) applied this methodology to a large corpus of reconstructed Proto-Quechua and Proto-Aymara roots, and posited several hundred Pre-Proto-Quechua and Pre-Proto-Aymara roots that descend from a period before the initial convergence. According to that analysis, as much as a third of the reconstructed Proto-Aymara lexicon may have been borrowed from the Quechuan lineage during the initial convergence. For more about these reconstructions, including the data and methodology, see Emlen (2017, to appear).

3. Agricultural and pastoral terminology in the early Quechuan and Aymaran lineages

The question addressed in this chapter is how terminology related to agriculture and herding fits into the history of Quechuan-Aymaran contact outlined above. The early agricultural and pastoral lexicons cannot be understood except with respect to this history; in addition, the borrowing patterns themselves may help answer important questions about the relationship between ancient languages and subsistence practices in the Andes. For instance, consider the following three possibilities: (a) we might find, once all of the borrowing has been accounted for, that only one pre-proto-language had a fully developed agricultural and herding vocabulary. It would be reasonable to conclude from this scenario that the political-economic context of the initial convergence was an encounter between people who were engaged in a mixed agricultural and pastoral economy, and people who were not. Or, we might find (b) that one pre-proto-language was associated with agriculture, and the other with herding, as in the more recent relationship of complementarity between Quechua-speaking cultivators in the intermontane valleys and Aymara-speaking camelid pastoralists in the high grasslands of the Andes (Urton 2012). If such a relationship functioned between the pre-proto-languages, we might expect

to find that asymmetry reflected in the subsistence lexicons. Or, finally, we might find (c) that both pre-*proto*-languages had fully developed agricultural and pastoral vocabularies. This would indicate a political-economic context in which both languages were already spoken by people engaged in mixed agricultural and pastoral economies before the initial convergence. In this scenario, each language would have been distributed across a range of ecological and elevational zones – what John Murra (1972) called a “vertical archipelago” of often discontinuous parcels in which a wide variety of crops and animals could be tended. These three scenarios illustrate how we might interpret the agricultural and pastoral vocabularies of each pre-*proto*-language and the subsequent patterns of borrowing between them. As will be clear from the following discussion, it appears that (c) is the most likely scenario.

The reconstructed Proto-Quechua and Proto-Aymara terms regarding agriculture and herding are presented in Table 1–Table 6 below. The terms are grouped in the following categories: crops and plant parts (Table 1); agricultural techniques, tools, structures, and materials (Table 2); food products derived from agriculture, and their associated tools and techniques (Table 3); domesticated animals (Table 4); herding techniques, structures, locations, and materials (Table 5); and weaving techniques and technology (Table 6). Terms that appear only in the Proto-Quechua or Proto-Aymara column are not shared by the other *proto*-language, and thus descend, according to our analysis, from Pre-Proto-Quechua and Pre-Proto-Aymara (respectively). Reconstructed terms that appear in both columns are shared by both *proto*-languages (e.g. *kuka ‘coca’ in Table 1). These shared items are **outlined**, and in cases in which it is possible to determine their provenances, they are indicated in the center column. There are several diagnostic criteria for identifying such provenances: roots that begin with *w or *y, or that have internal non-resonant codas or final consonants, are likely Quechuan in origin (Emlen 2017). These are marked with ‘Q’. Initial *l is one of few indicators of Aymaran provenance, as in *lampa ‘shovel, hoe’ in Table 2. This is marked with ‘A’. Shared terms that do not exhibit these diagnostic criteria cannot be definitively attributed to one lineage or the other, and are indicated with a question mark in the center column (as with *kuka ‘coca’ below). However, because the directionality of borrowing during the initial period appears to have been overwhelmingly from Quechua to Aymara, it is likely that most of the shared items presented below follow the same pattern.

Table 1 presents terms for Proto-Quechua and Proto-Aymara crops and plant parts. These include (a) tubers; (b) maize; (c) other high-elevation crops; (d) tropical crops; and (e) herbs. The terms in Tables 1–6 are presented alphabetically.

Table 1. Crops and plant parts²

Proto-Quechua	Provenance	Proto-Aymara
<i>(a) Tubers</i>		
*čawča ‘potato variety’		
*mašwa ‘tuber variety’		
		*šuta ‘potato variety’
*ułuku ‘olluco (tuber variety)’		
*uqa ‘oca (tuber variety)’		
*wayru ‘potato variety’		
<i>(b) Maize</i>		
*čułpi ‘maize variety’		
*muruču ‘maize variety’		
*panqa ‘corn husk’		
		*paru ‘toasted, golden-brown, maize variety’
*sara ‘maize’		
		*suq’u ‘corn husk’
		*tunqu ‘maize’
<i>(c) Other high-elevation crops²</i>		
*kinwa ‘quinoa’		
*tawri ~ *tarwi ‘lupine’		
<i>(d) Tropical crops</i>		
*kuka ‘coca’	?	*kuka ‘coca’
*šawintu ‘guava’		
*uču ‘chili pepper’		
*utku ‘cotton’		
<i>(e) Herbs</i>		
*wakatay ‘ <i>Tagetes minuta</i> ’		
*wałwa ‘ <i>Psoralea glandulosa</i> ’		

A few observations can be made about the reconstructions in Table 1. First, despite the great overlap between the Proto-Quechua and Proto-Aymara lexicons, they each exhibit separate terms for tubers and maize. There are more terms for tubers in our Proto-Quechua lexicon, but this may be because the reconstructed Proto-Quechua lexicon is larger (824 roots) than the reconstructed Proto-Aymara lexicon (496 roots). Furthermore, there is reason to suspect that Proto-Aymara in fact had separate terms for many of the Proto-Quechua items listed in Table 1: Southern Aymaran exhibits its own set of such terms, but they are not reconstructable in Proto-Aymara because they do not have Central Aymaran cognates. These earlier Aymaran terms may have been replaced in Central Aymaran by Quechuan terms

2. It is possible that the Southern Aymaran term *hup^ha* ‘quinoa’ is related to Central Aymaran *uhara* [uɸara] ‘maize’.

during the local convergence in Central Peru – indeed, the Central Aymaran lexicon appears to have borrowed around a quarter of its lexicon from neighboring Quechuan languages at this time (Emlen 2017: 337). No terms for tubers and maize crops in our corpus are shared at the level of the proto-languages. This suggests that speakers of Pre-Proto-Quechua and Pre-Proto-Aymara each cultivated these crops before the initial convergence, and that neither language had a special association with either maize or tuber cultivation before that time.

Furthermore, the reconstructions suggest that speakers of both languages cultivated crops at a range of different elevations: the tubers in Table 1 are mostly grown in the high *suni* and *puna* zones from 3500 meters to above 4000 meters (Pulgar Vidal 1987; Sandweiss & Richardson 2008), while most maize is grown in the *qheswa* zone between 2300 and 3600 meters, and in some places as high as 4100 meters (Staller 2016). This is consistent with a scenario in which both pre-proto-languages were distributed across ecological and elevational zones (as described above).

A notable difference between the Proto-Quechua and Proto-Aymara reconstructions in Table 1 is that tropical lowland crops (coca, chili pepper, cotton, guava) can be reconstructed in Proto-Quechua, but not in Proto-Aymara. This may suggest that the geographical range of Pre-Proto-Quechua extended further into the lowlands than that of Pre-Proto-Aymara (for instance, Gade 1975: 194 reports that guava is grown below 1600 meters in Southern Peru). However, this disparity may be due instead to the larger size of the reconstructed Proto-Quechua lexicon. Furthermore, the Aymaran languages that survive today are all found at high elevations – unlike today’s Quechuan languages, which are found across many elevations – so if there were once Aymara terms for lowland crops, they simply might not have been retained among today’s speakers. For example, it may be the case that Proto-Aymara had a term for ‘cotton’ (cf. Southern Aymaran *q^hiya* ‘cotton’), but that its reflex does not appear in Central Aymaran varieties because their distribution today is far from the lowland areas where cotton is grown.

Table 2 presents Proto-Quechua and Proto-Aymara terms for agricultural techniques, tools, structures, and materials.

The patterns of borrowing found in Table 2 confirm those in Table 1: Proto-Quechua and Proto-Aymara each have rich lexicons regarding agricultural techniques, tools, structures, and materials, and only a few of these terms are shared between the two languages. This constitutes further evidence that speakers of Pre-Proto-Quechua and Pre-Proto-Aymara were both sophisticated agriculturalists before the initial convergence.

Unlike in Table 1, however, most of the reconstructed terms in Table 2 do not suggest particular elevations, but rather refer to techniques or tools used for a variety of crops (with the exception of some terms that refer specifically to the harvesting of potatoes). For this reason, these reconstructions tell us that the speakers

Table 2. Agricultural techniques, tools, structures, and materials

Proto-Quechua	Provenance	Proto-Aymara
		*ali 'plant, stem'
*a ^h la- 'to harvest potatoes'		
		*at ^h a 'seed'
*čakma- 'to plow earth'		
*čaqu- 'to clear land for agriculture'		
*tšakra 'agricultural plot'		
*(h)a ^h lma- 'to turn soil'		
		*hipi- 'chaff, to shear, thresh'
		*hunu- 'to dig, harvest potatoes'
*išku- 'to shell (grain)'		
*kantša 'corral'		
*lampa 'shovel, hoe'	A	*lampa 'shovel, hoe, flat'
		*lama- 'to harvest, harvest potatoes, pick'
		*mač'a- 'fallow, dry season, to irrigate'
*muhu 'seed'	?	*muhu 'seed'
*murka- 'to thresh'		
*pa ^h la- 'to harvest, pick'		
*parqu- 'to irrigate'		
*pata 'terrace, platform'	?	*pata 'terrace, platform'
*pirwa 'granary, storage'	?	*pirwa 'granary, storage'
*qu ^h lpa- 'granary; to store'		
		*qurpa 'furrow, ditch, boundary'
*rawma- 'to prune'		
		*sa- 'to sow seeds'
*šikwa- 'to broadcast seeds'		
*šuka 'furrow'		
*tak ^h la- 'foot plow, to plow'		
*tarpu- 'to sow seeds'		
*wanu 'guano (fertilizer)'	Q	*wanu 'guano (fertilizer)'
*yapu- 'to plow'	Q	*yapu 'agricultural plot'
*yura 'plant'		

of Pre-Proto-Quechua and Pre-Proto-Aymara practiced agriculture, but not which crops they cultivated.

Table 3 presents food products derived from agriculture, as well as the tools and techniques used to produce those foods.

The patterns in Table 3 are more difficult to interpret than those in Table 1 and Table 2. Here, we see that Proto-Quechua has a robust lexicon of agriculturally derived food products, as well as terms for the tools and methods used to prepare them. Proto-Aymara also has roots that refer to grinding and flour, but most of

Table 3. Food products derived from agriculture, and associated tools and techniques

Proto-Quechua	Provenance	Proto-Aymara
		*aku ‘flour’
*anka- ‘to toast beans or corn’		
*api ‘a gelatinous porridge’		
*aswa ‘chicha (corn beverage)’		
*čučuqa ‘corn-based dish’		
*čuñu ‘dehydrated potato’	?	*č’uñu ‘dehydrated potato’
*čuqʷu ‘corn on the cob’	Q	*čuqʷu ‘corn on the cob’
*kaʷana ‘pan for toasting grain’		
*kamča- ‘toasted corn, to toast’		
*kaspɑ ‘ear of corn’		
*maʃka ‘toasted grain flour’		
*muti ‘boiled corn kernels’	?	*mut’i ‘boiled corn kernels’
*pʷiqa ‘corn flour’		
*qawi ‘dried oca’		
*tanta ‘bread’**	?	*t’anta ‘bread’ *t’iki- ‘to grind, mix’ *utʃɑ- ‘porridge, mush, to gulp’ *uʷɑ ‘(over)cooked or spoiled potato’
*upi ‘corn juice’		
*wayunkɑ ‘ear of corn hung up to dry’		

** The term *tanta may have had a different meaning in Proto-Quechua.

the Proto-Aymara terms for agriculturally derived food products themselves are shared with Proto-Quechua (and likely come from the Quechuan lineage, since that was the primary directionality of borrowing during the initial convergence). It is not clear why Proto-Aymara terms for food products would be borrowed from the Quechuan lineage, if the crops and techniques used to make them already existed in Pre-Proto-Aymara. This might suggest an Aymaran adoption of Quechuan cultural products, or it may simply be an artifact of the data samples. Note too that maize-related terms in Proto-Aymara come from the Quechuan lineage (e.g. *čuqʷu ‘corn on the cob’ and, probably, *mut’i ‘boiled corn kernels’); this does not support a scenario in which the Aymaran lineage has a privileged association with maize cultivation, at least at this early time.

Table 4 presents the reconstructed Proto-Quechua and Proto-Aymara lexical items that refer to domesticated animals.

The reconstructions in Table 4 show that terms for domesticated animals can be reconstructed in both Proto-Quechua and Proto-Aymara, and that each lineage has largely distinct terms for these animals. Thus, speakers of Pre-Proto-Quechua and Pre-Proto-Aymara likely both had domesticated animals before the initial

Table 4. Domesticated animals

Proto-Quechua	Provenance	Proto-Aymara
		*k'uyi 'guinea pig'
*lama 'llama'		
*paqu 'alpaca'		*qawra 'llama'
*uña 'juvenile domesticate'		
*uywa 'domestic animal'	?	*uywa 'domestic animal'

convergence. While this small sample does not support many generalizations regarding types of domesticates, one conclusion can be drawn: the fact that each lineage has separate terms for domesticated camelids indicates that speakers of both pre-*proto*-languages practiced high elevation camelid pastoralism before the initial convergence. If this is the case, then the two languages not only cross-cut elevational and ecological zones – in this case, extending to the high *puna* grasslands (4000–4800 meters) where camelids are herded – but were also spoken by herders as well as cultivators.

The reconstructed lexical items referring to herding techniques, structures, locations, and materials are presented in Table 5.

Table 5. Herding techniques, structures, locations, and materials

Proto-Quechua	Provenance	Proto-Aymara
		*ana- 'to herd'
		*awati- 'to graze, pasture'
*t̥saqna- 'to hobble an animal'		
		*(h)ik ^h a- 'to herd'
*miči- 'to pasture, feed'		
*puna 'high grasslands'		
*qarqu- 'to expel, drive out of a corral'		
*qati- 'to herd, drive (animals)'		
*qayku- 'to drive into a corral' ³		
*qint̥sa 'corral, enclosure'	?	*qint̥sa 'corral, enclosure'
*qiwa 'fodder, pasture grass'		

3. The Proto-Quechua terms in Table 5 that relate to herding (*qarqu- 'to expel, drive out of a corral', *qati- 'to herd, drive (animals)', and *qayku- 'to drive into a corral') are lexicalizations of an earlier Pre-Proto-Quechua monosyllabic root *qa 'to move, displace, herd (animals)' (Emlen to appear). Proto-Quechua probably also had other terms comprising *qa and the other directional suffixes: *qarku- 'to turn earth, drive animals uphill' and *qarpu- 'to push downward, drive animals downhill'. These terms survive in some Central Peruvian varieties of Quechua.

Proto-Quechua	Provenance	Proto-Aymara
		*uyu ‘corral’

The reconstructions in Table 5 demonstrate that a wide range of techniques, technologies, and materials connected to camelid pastoralism were used by speakers of Proto-Quechua and Proto-Aymara, and that each linguistic lineage has a rich and mostly separate vocabulary related to herding. This constitutes further evidence that speakers of both pre-proto-languages likely engaged in this subsistence activity before the initial convergence, and that the geographical reach of both languages included the high *puna* grasslands.

The use of fibers from alpacas and vicuñas is an important part of Andean domestic production, and it is closely connected to pastoralism. The reconstructed lexical items related to weaving techniques and technology are presented in Table 6.

Table 6. Weaving techniques and technology

Proto-Quechua	Provenance	Proto-Aymara
*awa- ‘to weave’		
*awáí- ‘to warp, weave’		
		*tʂ’anka ‘yarn, woolen thread’
		*tʂ’isa- ‘fuzz, lint, to card, comb wool’
*iáawa ‘shuttle, warp’	?	*iáawa ‘shuttle, warp’
*kaáwa ‘weaving instrument’		
*kurur ‘ball of yarn, clew’		
*miáwa ‘wool’		
*mini- ‘weft, to weave’		
*piruru ‘whorl’	?	*p ^h iáuru ‘whorl’
		*p’ita- ‘to weave’
		*p ^h awi- ‘to wind, spin thread’
*pučka- ‘spindle, to spin thread’		
		*qapu- ‘spinning wheel, to spin thread’
*qaytu ‘strand, thread’		
		*sayu- ‘to weave’
*šukšu ‘part of spinning wheel’		
		*t’apra ‘wool’

Like in the other reconstructions given above, the Proto-Quechua and Proto-Aymara terms that refer to weaving and spinning in Table 6 are mostly distinct. This suggests that speakers of both pre-proto-languages likely produced textiles from camelid fibers. These patterns, along with those found in Table 4 and Table 5, constitute evidence that pastoralism was practiced in the high *puna* grasslands by speakers of both Pre-Proto-Quechua and Pre-Proto-Aymara.

3.1 The innovative character of some Proto-Quechua agropastoral terms

A final observation can be made about Proto-Quechua agricultural and pastoral terms. As part of a process that took place across the whole of the Proto-Quechua lexicon, some of these items appear to be lexicalizations of archaic, monosyllabic Pre-Proto-Quechua roots (see, for instance, Adelaar 1986; Adelaar 2008; Emlen to appear; and Muysken 2011). Significantly, some of these were not originally related to agropastoralism.

For instance, *parqu- ‘to irrigate’ appears to comprise an archaic Pre-Proto-Quechua root *pa- ‘to fall (water), wetten’ and the well-documented directional suffix *-rqu ‘outward motion’, which still exists in some Quechuan languages. The resulting Proto-Quechua root *parqu- would have meant ‘to distribute water outwards’. But while irrigation is central to Andean agriculture, *pa- did not have a specifically agricultural meaning in Pre-Proto-Quechua: it appears to be lexicalized, to give just a few examples, in Proto-Quechua roots such as *paqča ‘waterfall, stream of water’; in Central Peruvian Quechua roots such as *paqa-* ‘to wash, bathe’ and *patška-* ‘to splash water’; and in Southern Peruvian Quechua roots such as *p^hawchi* ‘waterfall’, *p^hapi-* ‘to moisten dry corn to remove husk’, *p^haspay* ‘light irrigation’, and *para-* ‘to rain’ (Academia Mayor de la Lengua Quechua 2005).⁴ Therefore, it appears that speakers of Proto-Quechua innovated this term for irrigation from a non-agricultural root already present in the lexicon.

Similarly, Proto-Quechua *tarpu- ‘to sow seeds’ and *takła- ‘foot plow, to plow’ both contain a Pre-Proto-Quechua root *ta- that refers to hitting, knocking, and pushing (cf. *taka- ‘to punch, knock’; *taqła- ‘to slap, punch’; *tanqa- ‘to push’). *tarpu- ‘to sow seeds’ also includes a well-documented directional suffix *-rpu ‘downward motion’; the resulting bimorphemic construction would have meant ‘to hit or push downwards’. Other examples of roots lexicalized from Pre-Proto-Quechua *ta – just from the Cuzco variety (Academia Mayor de la Lengua Quechua 2005) – include *t^haqta-* ‘to flatten earth’; *t^haqti-* ‘to stomp, especially during dancing’; *t^haqpa-* ‘to throw earth onto’; *t^haya-* ‘to turn earth with plow’; *t^hasta-* ‘to flatten, shorten’; *t^harmi-* ‘to smash, stomp’; *t^halmi-* ‘to dig, scratch, look for leftover tubers’; and *t^hawi-* ‘to dig, looking for roots or tubers’. Such roots, some of which refer to agricultural techniques and some of which do not, are also ubiquitous across the other Quechuan languages.

If speakers of Pre-Proto-Quechua constructed novel pastoral and agricultural terms on the basis of earlier roots (like *pa- and *ta-, among many others) that did not have such meanings, this may suggest that agropastoralism was adopted at this

4. Note that it is not always clear what the adjoining morphology in these roots might have been.

point in the Quechuan lineage. As discussed in the introduction, this might have taken place between 3500 and 5500 years BP, when agropastoralism first developed in the Andean highlands. On the other hand, it is not necessarily the case that the speakers of Pre-Proto-Quechua adopted agropastoralism upon its first emergence in the Andean highlands (for instance, if Pre-Proto-Quechua made its way to the Andes from another part of South America where agropastoralism was not practiced). However, other Quechuan agropastoral terms do not appear to have been formed this way, and the nature of this process itself is still poorly understood (for more on this topic, see Emlen to appear).⁵

4. Conclusions

A few conclusions can be drawn from the foregoing presentation of agricultural and pastoral terminology in Proto-Quechua and Proto-Aymara. To begin with, some comments are in order regarding the relevance of this case to the Farming/Language Dispersal Hypothesis that is the topic of this volume.

Despite the fact that the Quechuan and Aymaran languages are widely distributed across a landscape with a long history of agriculture and pastoralism, they do not constitute a straightforward test of the Farming/Language Dispersal Hypothesis. That hypothesis proposes that the languages of agriculturalists replace the languages of neighboring hunter-gatherers. However, as discussed in the introduction to this chapter, the initial dispersal of the Quechuan and Aymaran families (perhaps one or two millennia BP) took place long after an agropastoral economy had already developed across the Central Andes (between 3500 and 5500 years BP). Thus, the Quechuan and Aymaran families spread across a landscape that had already been populated by farmers and herders, rather than hunter-gatherers, as the Hypothesis asserts. Furthermore, many of the languages with which the Quechuan and Aymaran families came into contact during their dispersals already had their own agricultural lexicons. This is not consistent with a scenario in which the Quechuan and Aymaran families were propelled across the landscape because their speakers possessed a subsistence advantage over their hunter-gatherer neighbors.

This leaves open the question of what economic and social forces propelled the families across the region. On this question, Heggarty and Beresford-Jones (2010) refine the Farming/Language Dispersal Hypothesis for the Central Andean context

5. It is interesting to note that some of these monosyllabic elements are the basis of ideophones in Quechuan languages (as well as others across Western South America). To give just one example, Nuckolls (1999: 242) reports that in Pastaza Quechua, *tak* (related to Pre-Proto-Quechua *ta discussed above) refers to “the sound of contact between two firm surfaces.”

by arguing that it was not the advent, but rather the later intensification of maize cultivation, long constrained by the diversity of Andean micro-environments, that led to the more recent dispersal of the Aymaran family. Our findings do not point to an alternative scenario for the initial Quechuan and Aymaran dispersals, but rather simply suggest that any link between the adoption of agropastoralism (or particular domesticates) and the expansions of those families is indirect at best.

It should be noted in passing that the history of Quechuan and Aymaran agropastoral terms can be correlated with the dates offered by the archaeological record, in a manner similar to the analysis put forth by proponents of the Steppe Hypothesis of Indo-European origin. According to that hypothesis, the presence of terminology referring to wheeled vehicles in the earliest periods of Proto-Indo-European suggests that the speakers of that language cannot have lived earlier than 6000 years BP, when wheeled vehicles first appear in the archaeological record (Anthony & Ringe 2015; Mallory & Adams 2006; see also Chang et al. 2015). Similarly, the presence of agricultural and herding terminology in both Pre-Proto-Quechua and Pre-Proto-Aymara suggests that the speakers of those languages cannot have lived before the advent of agriculture and herding in the Andes, which developed between 3500 and 5500 years BP. However, since Pre-Proto-Quechua and Pre-Proto-Aymara were likely spoken much later than these dates (see Figure 1), this merely confirms what was already evident.

But while the relationship between the adoption of agropastoralism and the Quechuan and Aymaran dispersals remains murky, our reconstructions do yield a number of other novel insights regarding cultivation and herding among speakers of Pre-Proto-Quechua and Pre-Proto-Aymara. Indeed, when we begin to disentangle the layers of lexical borrowing between the two lineages – a methodological prerequisite for any consideration of Quechuan and Aymaran prehistory – two notable facts become clear.

First, the parts of the Proto-Quechua and Proto-Aymara lexicons that refer to agropastoralism, including the names of domesticates, tools, techniques, products, etc., are mostly separate. This indicates that speakers of both Pre-Proto-Quechua and Pre-Proto-Aymara were likely both engaged in mixed agricultural-pastoral economies before the initial convergence some 1500–2500 years BP. If they were not, we would expect some degree of borrowing in these lexical domains, particularly during the initial convergence when the Aymaran lineage took on around a third of its lexicon from Pre-Proto-Quechua.

Second, both pre-*proto*-languages exhibit terms for cultivation and herding at a wide range of ecological and elevational zones, including camelid pastoralism above 4000 meters; the cultivation of tubers above 3500 meters; maize agriculture from 2300 to 3500 meters – and in some places as high as 4100 meters (Staller 2016); and in the case of the Quechuan lineage, tropical crops like guava, grown

below 1600 meters. The speakers of both pre-proto-languages, in other words, appear to have moved or sustained contact across elevations and engaged in various subsistence practices – perhaps, in the case of Quechua, into the tropical lowlands. This would be consistent with the typically Andean model of ecological complementarity, as well as with an integrated vision of highland-lowland socio-economic and linguistic continuities in Western South America (Emlen 2016). This discontinuous settlement pattern may have created a sort of Jackson Pollock-esque array of overlapping social contacts, generating what Mannheim calls, referring to the Southern Peruvian Andes some time later, a “mosaic of territorially interspersed languages” (Mannheim 1991: 60). Such a scenario, in which a variety of related and unrelated languages were likely spoken side by side in a multilingual environment spanning ecological zones, may help explain the pervasive and continuous language contact effects found in the Central Andean region. Furthermore, since ecological complementarity was the foundation of robust Andean economies, the inter-elevational nature of Quechuan and Aymaran-speaking social networks may itself have contributed to the dispersals of both language families.

Acknowledgements

Thanks to Martine Robbeets, Alexander Savelyev, Anna Berge, and Bruce Mannheim, who all provided helpful comments on earlier versions of this chapter. Thanks also to Sophia Nicolay. The research leading to these results received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement number 295918.

References

- Adelaar, Willem F. H. 1986. La relación Quechua-Aru: Perspectivas para la separación del léxico. *Revista Andina* 4(2): 379–426.
- Adelaar, Willem F. H. 2008. The vicissitudes of directional affixes in Tarma (Northern Junin) Quechua. In *What’s in a Verb? Studies in the Verbal Morphology of the Languages of the Americas*, Grażyna J. Rowicka & Eithne B. Carlin (eds), 121–142. Utrecht: LOT.
- Adelaar, Willem F. H. 2012a. Languages of the Middle Andes in areal-typological perspective: Emphasis on Quechuan and Aymaran. In *The Indigenous Languages of South America: A Comprehensive Guide*, Lyle Campbell & Verónica Grondona (eds), 575–624. Berlin: Walter de Gruyter. doi:10.1515/9783110258035.575
- Adelaar, Willem F. H. 2012b. Modeling convergence: Towards a reconstruction of the history of Quechuan-Aymaran interaction. *Lingua* 122 (5): 461–469. doi:10.1016/j.lingua.2011.10.001
- Adelaar, Willem F. H. & Muysken, Pieter C. 2004. *The Languages of the Andes*. Cambridge: CUP. doi:10.1017/CBO9780511486852

- Anthony, David W. & Ringe, Don. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1(1): 199–219.
doi:10.1146/annurev-linguist-030514-124812
- Campbell, Lyle. 1995. The Quechumaran hypothesis and lessons for distant genetic comparison. *Diachronica* 12(2): 157–200. doi:10.1075/dia.12.2.o2cam
- Cerrón-Palomino, Rodolfo. 1987. *Lingüística Quechua*. Cuzco: Centro de Estudios Rurales Andinos Bartolome de las Casas.
- Cerrón-Palomino, Rodolfo. 2000. *Lingüística Aimara*. Cuzco: Centro de Estudios Regionales Andinos Bartolomé de Las Casas.
- Cerrón-Palomino, Rodolfo. 2008a. *Quechumara: Estructuras paralelas del quechua y del aimara*. La Paz: Plural Editores.
- Cerrón-Palomino, Rodolfo. 2008b. *Voces del Ande: Ensayos sobre onomástica andina*, Vol. 3: Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Chang, Will *et al.* 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1): 194–244. doi:10.1353/lan.2015.0005
- Diamond, Jared & Bellwood, Peter. 2003. Farmers and their languages: The first expansions. *Science* 300 (5619): 597–603. doi:10.1126/science.1078208
- Emlen, Nicholas Q. 2016. Multilingualism in the Andes and Amazonia: A view from in-between. *Journal of Latin American and Caribbean Anthropology*. doi:10.1111/jlca.12250.
- Emlen, Nicholas Q. 2017. Perspectives on the Quechua-Aymara contact relationship and the lexicon and phonology of Pre-Proto-Aymara. *International Journal of American Linguistics* 83(2): 307–340. doi:10.1086/689911
- Emlen, Nicholas Q. To appear. On the polymorphemic genesis of some Proto-Quechua roots.
- Gade, Daniel W. 1975. *Plants, Man and the Land in the Vilcanota Valley of Peru*. The Hague: W. Junk. doi:10.1007/978-94-010-1961-3
- Hardman, Martha J. 1966. *Jaqaru: Outline of Phonological and Morphological Structure*. The Hague: Mouton.
- Heggarty, Paul & Beresford-Jones, David. 2010. Agriculture and language dispersals: Limitations, refinements, and an Andean exception? *Current Anthropology* 51(2): 163–191.
doi:10.1086/650533
- Mallory, James P. & Adams, Douglas Q. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: OUP.
- Mannheim, Bruce. 1991. *The Language of the Inka since the European Invasion*. Austin TX: University of Texas Press.
- Murra, John V. 1972. El control vertical de un máximo de pisos ecológicos en la economía de las sociedades andinas. In *Visita de la Provincia de León de Huánuco en 1562*, Tomo II, John V. Murra (ed.), 429–462. Huánuco, Perú: Universidad Nacional Hermilio Valdizán.
- Muysken, Pieter C. 2011. Modelling the Quechua-Aymara relationship: structural features, sociolinguistic scenarios, and possible archeological evidence. In *Archaeology and Language in the Andes: A Cross-Disciplinary Exploration of Prehistory*, Paul Heggarty & David Beresford-Jones (eds), 85–110. Oxford: OUP.
- Nuckolls, Janis B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28: 225–252. doi:10.1146/annurev.anthro.28.1.225
- Orr, Carolyn & Longacre, Robert E. 1968. Proto-Quechumaran. *Language* 44(3): 528–555.
doi:10.2307/411720

- Pearsall, Deborah M. 2008. Plant domestication and the shift to agriculture in the Andes. In *The Handbook of South American Archaeology*, Helaine Silverman & William H. Isbell (eds), 105–120. New York NY: Springer. doi:10.1007/978-0-387-74907-5_7
- Perry, Linda *et al.* 2006. Early maize agriculture and interzonal interaction in southern Peru. *Nature* 440(2): 76–79. doi:10.1038/nature04294
- Pulgar Vidal, Javier. 1987. *Geografía del Perú. Las Ocho Regiones Naturales*, 9th edn. Lima: PEISA.
- Sandweiss, Daniel H. & Richardson, James B. 2008. Central Andean environments. In *The Handbook of South American Archaeology*, Helaine Silverman & William H. Isbell (eds), 93–104. New York NY: Springer. doi:10.1007/978-0-387-74907-5_6
- Staller, John E. 2016. High altitude maize (*Zea Mays* L.) Cultivation and endemism in the Lake Titicaca Basin. *Scientific Pages of Botany* 1(1): 8–21.
- Torero, Alfredo. 1964. Los dialectos quechuas. *Anales Científicos de la Universidad Agraria* 2(4): 446–478.
- Tykot, Robert H. *et al.* 2006. The importance of maize in the Initial period and Early Horizon Peru. In *Histories of Maize: Multidisciplinary Approaches to the Prehistory, Linguistics, Biogeography, Domestication, and Evolution of Maize*. John Staller, Robert Tykot & Bruce Benz (eds), 187–197. London: Academic Press.
- Urton, Gary. 2012. The herder-cultivator relationship as a paradigm for archaeological origins, linguistic dispersals, and the evolution of record-keeping in the Andes. In *Archaeology and Language in the Andes*, Paul Heggarty & David Beresford-Jones (eds), 321–343. Oxford: OUP. doi:10.5871/bacad/9780197265031.003.0013
- Wheeler, Jane C. 1995. Evolution and present situation of the South American Camelidae. *Biological Journal of the Linnean Society* 54 (3): 271–295. doi:10.1016/0024-4066(95)90021-7
- Zimmerer, Karl S. 1995. The origins of Andean irrigation. *Nature* 378 (6556): 481–483. doi:10.1038/378481a0

Subsistence terms in Unangam Tunuu (Aleut)

Anna Berge

Alaska Native Language Center

The Eskimo-Aleut are arctic and subarctic hunter-gatherers known for their geographic spread and successful adaptation to a harsh climate; they are one of the canonical examples of a people that spread without agriculture. One of the most prehistoric recent spreads in this language family occurred about 1000 years ago, with effects felt throughout coastal Alaska. One area of language contact and possible spread was in Southeast Alaska, between the Pacific Coast Yupik language Alutiiq and the Aleutian language Unangam Tunuu. In this paper, I look at the distribution of cognates and borrowings of subsistence terminology in Unangam Tunuu, and I show that Alutiiq must have spread into a previously Unangaḵ area as a result of warfare rather than subsistence activities.

Keywords: Eskimo-Aleut, hunter-gatherers, prehistoric language contact, distribution of cognates, borrowed subsistence terminology, warfare

1. Introduction

The Eskimo-Aleut are an arctic and subarctic people known for their geographic spread, successful adaptation to a harsh climate, and hunter-gathering lifestyle; they are one of the canonical examples of a people that spread without agriculture. They spread from Siberia to Greenland in several migrations, splitting into the respective Eskimo and Unangan groups perhaps around 4000 years ago.¹ Those who settled on the Pacific Coast (i.e. Southeast Alaska and the Aleutians) developed large,

1. The Unangan are better known in the literature as the Aleut, but the currently preferred ethnonym is Unangaḵ in the singular, Unangan in the plural; the language is known as Unangam Tunuu. In this paper, I use the linguistically preferred term “Eskimo-Aleut” to refer to the language family, and “Unangam Tunuu” to refer to the language. The term “Eskimo” is dispreferred in some parts of the Arctic, but no term has replaced it to refer to both Yupik and Inuit languages. “Yupik” and “Inuit” are used both in reference to the languages and the people; however, in designating the latter, terms are pluralized with a *-t*: Yupik, PL. Yupit, or Yup’it specifically in reference to the Central Alaskan Yup’it; and Alutiiq, PL. Alutiit.

sedentary populations, stratified societies, food storage practices, and other characteristics often associated with agricultural societies. In the context of the farming/language dispersal hypothesis (cf. Renfrew 1987; Bellwood 2011), one could ask whether hunter-gathering subsistence activities in a resource-rich area have had effects equivalent to those brought by agriculture. The research that led to this paper was initially undertaken to address this question, and indeed, Fitzhugh (2003) essentially argues for this in a study of hunter-gatherers from Kodiak Island, Alaska, although from an anthropological rather than linguistic perspective. However, whether or not resource surpluses led to population replacement and non-Eskimo-Aleut language spread in the Alaskan Pacific Coast, they were certainly not the motivation for the different waves of Eskimo-Aleut spread. Quite the opposite, Eskimo languages in particular spread into the resource-rich and already settled Pacific Coast area. In this paper, I look at one instance of language spread in recent prehistory, involving the interaction of Unangam Tunuu and Alutiiq.

Around 1000 BP, there appears to have been a cultural shift in the Alaskan Peninsula, Kodiak Island, and Aleutian Islands. The nature of this cultural shift is hotly debated in archaeological circles: some view the area as an Alutiiq (Pacific Coast Yup'ik Eskimo) cultural homeland for thousands of years, with substantial recent external influence from northern Alaska; others see it as erstwhile Unanga territory, with the Alutiit arriving and replacing the Unangan about 800 BP. This period clearly resulted in linguistic contact between the Unangan and the Alutiit. Among linguists, this period is generally thought to have signaled the genesis of the modern Alutiiq language and likewise the beginning of a westward expansion of the Eastern dialect of Unangam Tunuu.

Neither the motivation for this contact nor the question of whether Alutiiq spread into a formerly Unanga-speaking area and pushed one dialect of Unangam Tunuu westward have been systematically investigated, however. In this paper, I investigate the subsistence terminology in Unangam Tunuu for clues as to the nature of this language contact. I first provide a background to the Eskimo-Aleut language family and to the Unangan subsistence activities (Section 2). I then establish that there is no evidence for a tradition of prehistoric agriculture, and I review the subsistence terminology in Unangam Tunuu, specifically with respect to the distribution of Unangan and Eskimo cognates and borrowings (Section 3). Distribution studies reveal several patterns: an unequal distribution of cognates in different semantic domains; a higher number of cognates in domains relating to most men's activities; and a correlation between domains with high numbers of cognates and high numbers of borrowings (Section 4). I then discuss possible motivation(s) for this language contact and possible spread: the patterns seen here suggest an influx of Alutiiq men into a previously Unanga area, and the resulting language spread must have occurred as a result of warfare rather than subsistence activities.

This study is part of a larger series of studies (Berge forthcoming; Berge 2016) on the Unangaᵗ lexicon, which together are indicative of prehistoric language contact, if not outright language mixing (cf. Ross 2003).² These studies do not indicate the source(s) of this presumed language contact, beyond speculations already common in the literature which focus on non-Eskimo groups (e.g. Leer 1991; Fortescue 1998). Bergsland (1989, 1994) pointed out that the numerous borrowings between Yupik languages and Unangam Tunuu suggested some post-Eskimo-Aleut split contacts. In this study, I show that the most important source of this contact is likely to have been Alutiiq, and that many presumed Eskimo-Aleut cognates are probably best understood as late borrowings between Unangam Tunuu and Pacific Coast Yup'ik, especially Alutiiq.

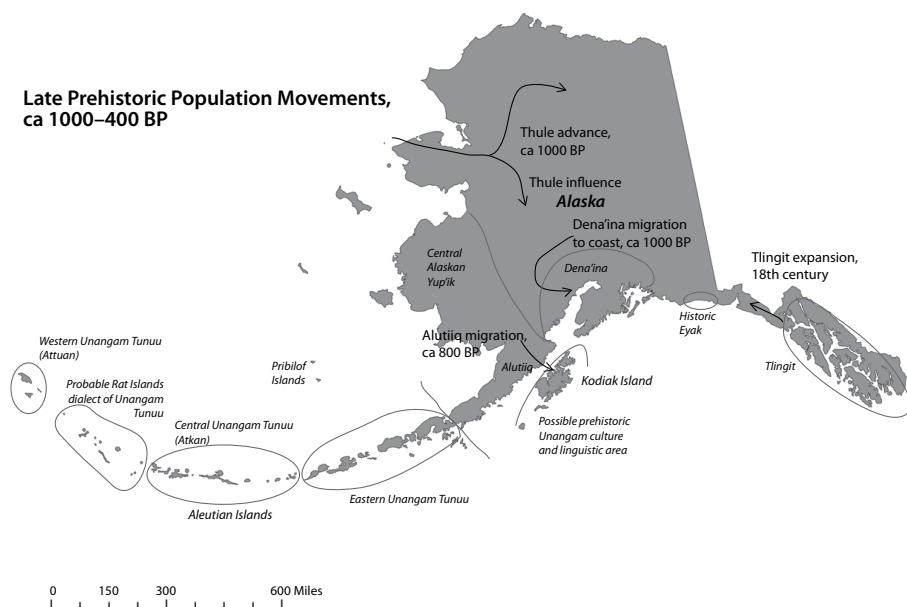


Figure 1. Late Prehistoric population movements, ca. 1000–400 BP

2. Evidence of such contact includes a split in the lexicon, with more cognates among grammatical terms (e.g. inflectional morphology, particles, pronouns, deictic terms, etc.) than lexical terms; more cognates among verbs than nouns; more cognates that are derived, as opposed to cognate roots that are morphologically simple and semantically general; more cognates in semantic domains relating to men's activities than to women's; and cognates more central to the domain in men's rather than in women's domains (Berge forthcoming; Berge 2016).

2. Background

Eskimo-Aleut is the last major language family to have arrived in North America; it is thought to have split into Eskimo and Unangam Tunuu by about 4000 BP. Unangam Tunuu, traditionally spoken along the Aleutian Chain, is the only language in its branch of Eskimo-Aleut, consists of three documented dialects, Attuan, Atkan, and Eastern, and is known for its substantial divergence from Eskimo languages; the internal differences between the dialects, however, are shallow enough to suppose a late dialect spread (Woodbury 1984: 62). The Eskimo branch is larger and more diverse, with two major branches, Yupik, with 4 or 5 languages, and Inuit, with a number of major dialects. Yupik and Inuit are thought to have diverged about 2000 years ago. The language groups have had several periods of contact postdating the split of Eskimo-Aleut.

The Aleutian Islands lie between Alaska and Eastern Russia, separating the Pacific Ocean to the south from the Bering Sea to the north. Several marine current systems meet there, creating conditions for an extremely rich and diverse marine ecosystem. The Aleutians are not an ecologically marginal area; and although not conducive to agriculture, they supported dense populations, extensive food storage systems, and some of the most complex hunter-gatherers in the world (Erlandson 2001: 289; Heggarty 2015: 620), over a 9000 year history of habitation and a number of distinct cultural periods. The earliest cultural period was short-lived, appears to have been a non-marine adapted group from the Alaskan mainland (Potter 2010), and involved a small and localized population in the Eastern part of the Aleutian Chain (Maschner 2016: 326). The following three periods are generally more crucial to the interpretation of Eskimo-Aleut presence, and specifically to the question of the identity of the original inhabitants of Kodiak Island.

The period from 7000–4500 BP is variously seen as a continuation of the preceding (Davis and Knecht 2010) or as a new culture with ties to or origins around Kodiak Island (Maschner 2016). It is characterized by near-shore marine adaptations such as boating technology, harpoons for hunting large sea mammals from the shoreline and near-shore, and extensive harvesting of fish and shellfish, and it shares numerous cultural traits with the Pacific Coast. The period from 4400–300 BP shows cultural continuity with the preceding one, with several significant developments. Around 3500 BP, following a period of volcanic eruptions, Kodiak was effectively split off from the Eastern Aleutians, each area thenceforth developing separate cultural traditions, which Maschner (2016) views as separate Unangan traditions, while Fitzhugh (2003) views Kodiak as already Alutiiq. Around 3000 BP in the Alaskan Peninsula and Eastern Aleutians, there is evidence of a small infiltration of Arctic Small Tool tradition materials, typically associated with the Eskimo. These are important details: the interpretation of the Arctic Small Tool

tradition infiltration is crucial to understanding the relationship of the Unangan to the Eskimo and to dating the Eskimo-Aleut split; the interpretation of the culture on Kodiak, as originally Unangaꝯ or the ancestors of the Alutiit, is crucial to understanding the nature of the language contact in the Kodiak area in the next period.

About 1500–1000 BP, there was a broad Alaska-wide upheaval resulting from climate changes, volcanic activity, rapid population movements, including, among others, the southward movement of the Thule Eskimo from Siberia and Northern Alaska. The latter led to an influx of Eskimo-related culture and possibly the Alutiiq people to the Pacific coast and Kodiak (Dumond 2001; Maschner et al. 2009, 2016; contra Fitzhugh 2003; Steffian et al. 2016). It also points to what is generally considered the beginning of the modern Alutiiq language. The extent of Eskimo influence in the Unangaꝯ area to the east is not clear; but the first real evidence for the use of kayaks in this area dates from the early part of this period, and hunting of large sea mammals is much more common, judging from the types and size of spear and harpoon points (Maschner 2016), faunal remains, and isotope studies of bones (Byers et al. 2011); both are commonly associated with Thule culture. The Thule also reintroduced the bow and arrow for warfare and brought slatted wood and hide armor, and these made their way to the Aleutians (Dumond 1987; Maschner & Reedy-Maschner 1998). At the same time, there was significant cultural diffusion throughout the Pacific Coast, probably originating in the south and encompassing the Unangan, Alutiit, Dena'ina, Eyak, Tlingit, and others. The Pacific Coast culture area is characterized by a highly stratified society, slavery, increased trade and warfare, the development of longhouses (suggestive of a patrilineal society), etc. (Byers et al. 2011). There is an influx of new people into the Eastern Aleutians, with genetic evidence for replacement of the female line by the end of this period (Smith et al. 2009).

The final period, after a severe climate-driven resource and population crash around 900–700 BP, is similar, but also shows a conspicuous increase in fortifications and refuges (Maschner et al. 2009), indicating more emphasis on wars and defensive activity, and a new westward Unangaꝯ expansion along the Chain. Misarti and Maschner (2015) speculate that this expansion is a result of the acquisition of Kodiak wives (long-separated Unangan according to them), although it may also be seen as a push-effect from the influx of Alutiit into Kodiak and Thule influence in general on Yupik areas destabilized by their own wars and population movements (cf. Funk 2010 for a description of the Yupik wars). Linguistic traces of this westward expansion of Unangan are apparent in the dialects (cf. Bergsland 1994: XVff).

The Unangan were complex hunter-gatherers. They had a long history of permanent settlements, smaller or larger depending on the period, with seasonal hunting and fishing camps. Subsistence activities encompassed marine, littoral, and terrestrial areas, and they were strongly differentiated by gender: marine and bird

hunting were exclusively conducted by men, whereas women were responsible for most littoral and terrestrial activities, such as river/stream fishing, berry, grass, and wood gathering. Marine subsistence activities involved the use of the kayak and included deep-sea fishing, using deep-sea fishing lines; surface fishing using regular fishing lines, and sea mammal hunting, using harpoons, spears, and atlatls (spear throwers). Whales were hunted using poison darts and harvested if and when they washed up onshore. Littoral subsistence activities consisted especially of intertidal gathering of shellfish, seaweed, etc. Terrestrial subsistence activities included stream fishing using seines, dipnets and fishing lines; berry, root, and grass gathering, the latter for basket and mat production; and bird hunting, using bolas, spears, darts, and in some places also nets. Subterranean storage pits or shelving areas in houses, drying racks and huts, use of sea mammal intestines or grass baskets for storage, etc. were used for processing and storage. Traditional materials included ivory, bone, driftwood, stone, and in eastern areas slate, for tool production; grass for woven baskets and mats; sinew, skin, feathers, and intestines for clothes, boat coverings, and storage; and shells for decoration and currency. The Unangan were well integrated into a larger culture area; ethnographic sources mention the esteem or fear engendered by the Unangan as fierce fighters and notable traders. Their primary enemies were the Alutiit, but they were known to non-Eskimo neighbors along the coast.

3. Subsistence terminology and language spread

3.1 Evidence for early agriculture?

Eskimo-Aleut is famously spoken in a region that appears never to have known agriculture; however it is also one of the last prehistoric language families to have arrived on the North American continent from Asia. We might therefore wonder whether there is evidence for early agriculture that may have subsequently been lost. There is no such evidence within Eskimo-Aleut, for several reasons. For one, agricultural terms in the daughter languages are clearly derived from hunter-gatherer terminology. The primary and earlier meaning of UT *itxi-lix*³ ‘to plant,’ for example, is ‘to drop, let down a net, fishline, etc.’ (Bergsland 1994: 214). The root from which this word is derived, *it-*, *isix* means ‘to fall, to drop down,’ and other derivatives of the root include, e.g., *itxuli-x̄* ‘a kind of fox snare.’ Other Unangan examples of agricultural terms created from original hunter-gatherer terminology include:

3. UT lexical items are given in citation form as found in Bergsland (1994) (usually but not exclusively -*x̄* on nouns and -*lix* on verbs).

- (1) *chisi-lix* ‘to sow,’ primarily ‘to divide, distribute food,’ from the practice of distributing sea-catch
- (2) *tanasx-a* ‘cultivated field, kitchen garden with vegetables,’ from ‘field, hunting and fishing area away from village, camping area’
- (3) *angusu-x̂* ‘mortar,’ from ‘soapstone (used for making seal oil lamps)’

Further, some Unangan agriculture terms are clearly post-contact coinages (often specially created for Bible translations):

- (4) *chiqix̂ mâxiisâx̂* ‘farmer,’ lit. ‘worker with mud’
- (5) *hitnisâx̂* ‘cultivated plant,’ lit. ‘that one waits to grow’
- (6) *chiqimaan aguusix* (du), lit. ‘(bipartite) instrument for soil’

Interestingly, pre-contact tools that are functionally equivalent to agricultural terms were not extended to include agricultural activities; thus *hinkuulûx̂* ‘anvil stone on which paint was ground’ and *hinkuulûgim chaa* ‘stone pestle for grinding paints,’ lit. ‘hand of the anvil stone ...,’ were not extended to mean mortar and pestle; the modern term *tulkiisi-x̂* ‘masher, pestle,’ from *tulki-lix* ‘to mash,’ is from Russian *tolki* ‘push, strike, pound.’ Likewise, none of native terms for harvested roots, such as *qunglux* ‘root of any plant,’ are used for modern agriculture. Biblical coinages (4–6), Russian borrowings such as *siniicha-x̂* ‘wheat’ from Russian *pshenitsa* ‘wheat,’ or descriptive phrases such as *inisam uluudaa* ‘carrot,’ lit. ‘red cultivated plant,’ are used instead.

Finally, there is no reconstructed shared agricultural terminology in Eskimo-Aleut:⁴

- (7) ‘to plant’ UT *tini-lix* lit. ‘to set up (tents, nets)’ or *itxi-lix* lit. ‘to let down (nets, fishline),’ cf. Inuit (WG) *ikut(i)-* ‘to plant, fix, place over fire’ (PE **əkə-* ‘to let or put in’)
- (8) ‘to sow’ UT *chisi-lix* lit. ‘to scatter, distribute,’ cf. Yupik (CAY) *naucecii-* ‘to plant, sow, make grow’ (PE **nayu-* ‘to grow’), Inuit (WG) *siaruarter-* ‘to strew something’ (transitive form) (PE **ciḍay-* ‘to spread’)

3.2 Traditional subsistence terminology

There are about 373 subsistence terms in Unangam Tunuu, of which about 60 are cognate with Eskimo terms. Subsistence terms in Unangam Tunuu come from among the following domains (numbers in parentheses refer to the number of

4. Proto-Eskimo is a relatively young family with a time depth not much greater than 2000 years. Proto-Eskimo-Aleut has been assumed to be twice as old based on the divergence of its lexicon.

terms found in Bergsland 1986, 1994 and Fortescue et al. 2010, the main sources of data for this study):

Table 1. Unangan subsistence terms by gendered activity and semantic domain

Women's domain (total: 144)	Men's domain (total: 215)
Skin preparation (20)	Sea Mammal Hunting (53, overlap with war terms)
Gathering (berries, grass) (29)	Bird hunting (11)
Weaving (mats: 14, basketry: 22, general terms: 8) (44)	Deep Sea Fishing (13)
Sewing (24)	Boats (64)
Fishing from land (rivers, shore) (27)	Trapping (10)
	Knives (39)
	General tools (drills, adzes, etc.) (10)
	Fire-making (15)

In the years since the Eskimo-Aleut split, there have been numerous technological innovations and points of differentiation. Important differences between the groups include the use of pottery, dog-sleds, ice fishing techniques, and toggling harpoons for hunting whales among the Eskimos, as opposed to the exclusive use of grass baskets, the lack of dog-sleds, the practice of deep sea fishing and salmon harvesting on major rivers and streams, and whale hunting using poison darts (the latter two also found among the Pacific Eskimos). At the time of European contact, both groups made use of the kayak, and the open-skin boat for group travel (UT *niġ*, PY *anyaq*, PI *umiaq*). The dog-sled and the whale hunt with toggling harpoons are associated with the Thule advance.

Unsurprisingly, we find both a common core of shared vocabulary as well as independent development in the respective vocabularies reflecting the different experiences and cultural innovations of the groups. There are cognates in almost all domains listed in Table 1, although the domains are elaborated differently in the different branches of Eskimo-Aleut. Thus, the Unangan have about 30 terms for grass baskets, whereas only about 5 terms for bags or pouches reconstruct to Proto-Eskimo; conversely, there are no terms relating to dog-sleds or their parts in Unangam Tunuu, as against about 10 terms in Eskimo found in Fortescue et al. (2010). There are, however, some observations to make regarding the distribution of cognates within these subsistence domains. Some domains have a relatively low proportion of cognates, while others have a far greater proportion thereof. In Section 3, I discuss first the methodology used to examine the distribution of cognates (Section 3.2.1) and the relative proportions of cognates across different domains (Section 3.2.2). The patterns observed are described in Section 4. In brief, these include differences in the proportion of cognates found in the different

semantic domains relating to subsistence (Section 4.1), as well as between semantic domains relating to gender-specific activities between Unangam Tunuu and Eskimo (Section 4.2); the high correlation between number of cognates and number of borrowings in a particular semantic domain (Section 4.3); and the complication of high levels of cognates referring to technologies that post-date the Eskimo-Aleut split (Section 4.4).

3.2.1 *Methodology*

Eskimo-Aleut languages are known for their extreme polysynthesis, which complicates the matter of determining what is a cognate. Most words, and especially verbs, involve some degree of spontaneous derivation. However, some morphologically complex words are so common as to be lexicalized, as in UT *kigusi-ġ* ‘tooth,’ literally ‘thing used to bite with,’ from *kix-six* ‘to bite,’ or as in Inuit *kiguti* ‘tooth,’ from *kii-* ‘to bite something’ (transitive form). It can be difficult to determine if a pair of words are true cognates or independent derivations arising after the Eskimo-Aleut split (Bergsland 1986: 102). For the purposes of this study, words or roots are counted as cognates if they are so identified in the main sources. One in which all parts of the words being compared are reconstructible is referred to as a *cognate*. Thus, UT *iqya-ġ* is assumed to be cognate with Yupik and Inuit *qayaq* ‘kayak,’ and more specifically with the reconstructed PE **qayar*. *Cognate roots* are the reconstructible roots in pairs of words with additional morphology, as in the bound root *ahya-* – in UT *ahyaaku-* ‘play dart,’ cognate with PE **ayay-* ‘thrust or push with a pole.’ *Questioned cognates* are those whose reconstruction or semantic relation is irregular, as in UT *ġaasi-ġ*, *haasi-ġ*, PE **paŋarun* ‘kayak paddle,’ where UT /ġ/ is in an unusual correspondence with Eskimo /p/, and a medial syllable is lost in Unangam Tunuu; another illustrative example is UT *qigda-ġ* ‘single hook of fish spear,’ which has an unclear semantic connection with PE **qamir* ‘ridge,’ cf. AI *qimiq* ‘hill, mound, lead line or float line of net.’

The choice of one morphological analysis over another also affects the identification of a cognate. For example, the word *snuuġi-ġ* is listed as having two senses (Eastern) ‘alien, stranger’ and (Atkan) ‘low-class person, subordinate.’ Bergsland (1994: 370) relates it to *sna-* ‘side,’ which is assumed to be cognate with PE **cinā-* ‘shore or edge,’ both because of an assumed semantic link between ‘side’ and ‘being on one side, alien,’ and because of properties of the suffix – *Vġi-*. However, the vowel change is unexplained, and one would have expected *snaaġi-* instead. There is, on the other hand, an entry *snu-* ‘to send on an errand, to order; to hire,’ under which we find *snu-ġ* ‘person to do something.’ The word *snuuġi-ġ*, therefore, may actually be derived from *snu-ġ* with a different suffix than the one suggested (cf. Bergsland 1994: 477). If so, it does not have an Eskimo cognate. Likewise, derivations may sometimes mask the ultimate relatedness of their roots; for example,

Bergsland (1994: 411) lists UT *tutax* (du) ‘earrings’ as a separate entry, although he tentatively links it to *tut-*, *tusix* ‘to hear,’ which is cognate with PE **tucar-* ‘to hear or understand.’ For the purposes of the discussion in Sections 3.2.2 and 3.2.3, entries in Bergsland (1994) and Fortescue et al. (2010) are analyzed as is, and they total about 450 proposed cognates. However, the issue of what counts as cognate is far from resolved, as I will show in Section 3.2.4.

Proposed cognates differ not only in how regular the sound correspondences are or in the degree to which they are simple vs. derived words, but also in the degree of semantic extension they show. For example, UT *ila-lix* ‘to be part (of something)’ has no phonological variation between Unangan dialects and is cognate with PE **ila-* ‘part;’ it also has a certain amount of semantic variation, both within Unangam Tunuu and within Eskimo: in addition to the main sense, it also means ‘to pass by, to do too much’ and has the nominal forms ‘part of, piece of, some, relative of.’ This is in contrast to words like UT *ayaġu-ġ* ‘walking stick,’ a word that was presumably morphologically complex at one point, with no phonological or semantic variation within Unangam Tunuu and said to be cognate with PE **ayaRUR* ‘walking stick.’

Proposed cognates were sorted into exact cognates, cognates with regular sound correspondences, cognate roots, cognates with irregular sound correspondences, and cognates with dubious semantic correlations. They were also coded for level of phonological variation within a branch of the language family (e.g. within Unangam Tunuu) and level of semantic variation. The results are summarized in Berge (forthcoming). The focus of the discussion here is limited to examining the distribution of the cognates and borrowings identified in the major sources in semantic categories, or domains.⁵ The findings are presented from the perspective of Unangam Tunuu.

3.2.2 *Eskimo-Aleut cognates*

Cognates and cognate roots appear in most semantic domains related to subsistence; however, we can divide the semantic domains into those with low and those with high numbers of cognates. Low-cognate domains have between 0–20% cognates or cognate roots, whereas high-cognate domains between roughly 40–53% cognates or cognate roots. This division is admittedly arbitrary, and reflects a noticeable gap of almost 20% between the two groups. Only one category (hunting) reviewed appeared to fall in between these numbers, but subdividing the category

5. While the traditional criteria for determining cognate status include especially regular sound correspondences, I avoid categorically assuming that proposed cognates with regular correspondences are in fact cognates. My goal is first to find correlations and patterns.

revealed important differences in the distribution of cognates.⁶ In the discussion that follows, I discuss the characteristics of the subsistence terms that are represented in Table 2 first before providing interpretations of the percentages.

Table 2. Percentages of Eskimo-Aleut cognates in semantic domains pertaining to subsistence terminology

Low cognate domains		High cognate domains	
domain	% cognates / cognate roots	domain	% cognates / cognate roots
Sewing	16	Fire-making	53
Gathering	14	General Tools	44
Fishing from land	13	Boats	40
Skin Preparation	11	Deep Sea fishing	38
Knives	10		
Weaving	5		
<i>Mats</i>	0		
<i>Baskets</i>	9		
Bird hunting	0		
Trapping	0		
		Hunting / War	30
<i>Sea mammal hunting with harpoons (16 terms)</i>	6	<i>Sea mammal hunting weapons not including harpoons (27 terms)</i>	36
		<i>Hunting weapons also used for war (17 terms)</i>	>50

The percentages in Table 2 refer to both cognates and cognate roots; but several semantic domains have almost no cognates, if indeed they have any. Terms related to weaving (in Unangam Tunuu, this almost exclusively concerns weaving mats or baskets), for example, include only one proposed cognate, *kuygi-s* ‘grass ribs in basket,’ from *kuygi-â* ‘vertebra of the loin, rumpbone of bird’ ~ PE **kuyay* ‘lumbar vertebra or keel of boat,’ and one questioned cognate root, *qitxu-â* ‘whirl of hair, grass mat,’ possibly derived from *qit-* ‘to contract, be twisted (of rope),’ cf. PE **qit-* ‘be convulsed.’ Likewise, there are no cognates and only three cognate roots relating to fishing from land, and no cognates or cognate roots for bird hunting or trapping.

6. Hunting terminology includes spears, darts, and harpoons, as well as their parts; as mentioned in Section 2, hunting specifically refers to the hunting of sea mammals. Some hunting tools were also used for war against humans; bows and arrows were specifically used for war on the Aleutian Islands, although they were also used against big game on the mainland. Bird hunting terminology include bolas and bird darts and spears and their parts.

Several domains, including sewing, cleaning/processing, gathering, and words relating to knives are low-cognate categories, with no more than 20% cognate roots. These roots (illustrated in (9)–(12)) are characterized by semantic generality and by high level of semantic variation within a language group and/or divergence between branches of the language family (as in *kala-*):

- (9) UT *da-* ‘eye’ ~ PE **əðə* ‘eye’ → UT *damiku-x̂* ‘hole around upper edge of basket’, *dagalukix* ‘notches on edge of ancient bone needle’
- (10) UT *qa-* ‘food, fish, eat’ ~ PE **nəqə* ‘food, meat’ → UT *qalima-x̂-six*, *qaligda-l* ‘to clean fish’, *qasi-lix* ‘to fish for supply’, *qanaa-x̂-six* ‘to fish’, etc.
- (11) UT *igu-* ‘to take out’ ~ PE **niyu-* ‘to disembark, take off, take out’ → UT *igula-x̂* ‘bone root digger for lupine roots’
- (12) UT *kala-lix* ‘to string fish (by the mouth on a rope)’ ~ PE **kala-* ‘to tow’.

There are two or three apparent cognates or cognate roots with a more specific relationship with Proto-Eskimo:

- (13) UT *yu-n* ‘crimp, seam in leather work’, cf. CAY *yuurte-* ‘to curve, bend’ (not listed as a cognate in Fortescue et al. 2010)
- (14) UT *hinguqa-x̂* ‘needle’, from *hingu-lix* ‘to push’ ~ PI **pingu-* ‘to push’
- (15) UT *tani-x̂taasi-x̂* ‘burning lamp’ ~ PE *nanir* ‘lamp’
- (16) UT *saami-x̂* ‘stone knife’ ~ PE **caviy* ‘knife’

The results from these low cognate domains generally suggest a long period of separation and independent development, as many have already suggested (e.g. Bergsland 1986); I show elsewhere (Berge, forthcoming) that these are more likely recent borrowings (cf. also Section 4.3) than true cognates.

Semantic domains with high numbers of proposed cognates and cognate roots include boats and boating, fire-making, and general tools. For example, about 40% of boating terms appear to be cognates or cognate roots. The majority are verbs (17)–(21), with a few cognate nouns (22)–(24) and questioned cognate nouns (25)–(26):

- (17) UT *chala-lix* ‘to slide, come ashore’ ~ PE **tulay-* ‘to land (come ashore)’
- (18) UT *sayu-lix* ‘to pull’, Eastern ‘to row in a baidarka’ ~ PE **cayuy-* ‘to pull or twitch’
- (19) UT *hayamda-lix* ‘to be unsteady (of a box in a boat)’ ~ PE **pay(y)ay-* ‘to be unsteady or weak’

- (20) UT *knachi-lix*, *knaxchi-lix* ‘to soak the seams of the skin covering of a baidarka, put cover on baidarka; to dry skin in the sun, oil, and put on baidarka’ ~ PE **kənit*- ‘to soak’ (Bergsland 1994) or PE **kinəR*- ‘to drip dry’ (Fortescue et al. 2010)
- (21) UT *hum-six* ‘to inflate (bladder, floats), swell, swell up’ ~ PE **puvə*- ‘to swell; cf. also the derived UT forms *humaġi-x̂* ‘inflated decoy seal,’ *umalu-x̂* ‘small bone pipe for bladders to carry fresh water or for inflating sealskin floats for harpoons’
- (22) UT *iqya-x̂* ‘kayak’ ~ PE **qayar* ‘kayak’
- (23) UT *taamġax̂*, *taamx̂ax̂* ‘cross-strap on kayak deck’ ~ PE **taprar* ‘skin rope or strap,’ in Inuit dialects ‘cross-strap on kayak deck’
- (24) UT *qisa-x̂* ‘strap for tying up baidarka’ ~ PE **qitəR*- ‘to tie’
- (25) UT *x̂aasi-x* ‘double-bladed paddle for baidarka’ ~ PE **paŋərun* ‘paddle’
- (26) UT *suka-x̂* ‘baidarka spray skirt’ ~ PI *cukak*- ‘to be tight or tighten’

Sea mammal hunting appears to fall between these two groups. However, within this domain, cognates are unevenly distributed: 36% of cognate terms relate to spears, darts, and arrows (e.g. (27)–(28)), while those relating to harpoons are almost entirely non-cognate (e.g. (29)).⁷ Most cognate roots are disputed: Bergsland (1994), for example, lists (30)–(31) as cognate (roots) while Fortescue et al. (2010) does not:

- (27) UT *qaḵuu-* ‘butt of shaft, back part of shaft covered by spear thrower’ ~ PE **qaqu*(*R*) ‘shaft of harpoon thrower’
- (28) UT *ayaqudaax* ‘sea otter spear,’ *ahyaaku-* ‘play dart’ ~ PE **ayay-* ‘thrust or push with a pole’
- (29) UT *tunumulgu-x̂* ‘simple harpoon for hunting sea otters and fur seals’ (no Eskimo cognate)
- (30) UT *agalġi-x̂* ‘dart’ ~ PE **ay*(*γ*)*a*(*R*)- ‘to hang’
- (31) UT *hasxu-x̂*, *haasxu-x̂*, *haaxsu-x̂* ‘spear thrower, throwing stick’ ~ PE **patəy-* ‘to slap’

Although war is not, per se, a subsistence activity, some sea mammal hunting weapons terms are also used for war; and if we look at this subset of subsistence terms, more than 50% are cognates or cognate roots. For example, *anax* ‘club, stick

7. Menovshchikov, cited in Liapunova (1996: 14), observed that there were fewer cognates among men’s sea-hunting tools than expected; however, as we see, it depends on the kinds of hunting tools.

for clubbing animals and big fish' (~ PE *anayu-* 'hit (with club)') has a derivative *anagasi-x̂* 'native axe,' used in a compound name *Alitxum Anagacha* 'War Axe' (Bergsland 1994: 70; cf. also (31)). In Unangam Tunuu, bow and arrow terminology is associated with war and is predominantly cognate or cognate root (33)–(34):

- (32) UT *ingadusi-x̂* 'spear, dart (for war)' (Bergsland → AAY *ingaquq* 'spear with blade')
- (33) Eastern UT *kaluuda-x̂* 'crossbow,' from *kalu-* 'to shoot with bow or gun,' ~ PE **katluy* 'thunder'
- (34) Atkan UT, *saygiida-s* 'crossbow,' Eastern UT *saygiida-x̂* 'toy bow,' from the root *sayu-lix* 'to pull,' cognate with PE **cayuy-* 'to pull or twitch'

4. Major patterns

Three main observations stand out: the cognates are not evenly distributed between semantic domains (Section 4.1), domains with higher levels of cognates are those related to men's activities (Section 4.2), and there appears to be a correlation between higher levels of cognates and higher levels of borrowings (Section 4.3). A closer examination of some of the proposed cognates in light of the dating of technological innovations also suggests that they may be better thought of as late borrowings (Section 4.4).

4.1 Unequal distribution of cognates between semantic domains

All things being equal, one might expect an even distribution of cognates except where subsistence activities required environmentally specific innovations, e.g. Unanga grass weaving vs. Inuit dog sledding, as mentioned in Section 3. The data reveal notable differences precisely between domains involving such specific cultural innovations. Low-cognate subsistence domains include river or near-shore fishing, bird hunting, gathering, weaving, skin processing, trapping, etc. Both Eskimo and Unangan groups engaged in these activities, but they share mostly cognate roots (as opposed to cognates) of a very general nature (e.g. 10), and with different semantic extensions (e.g. 11), suggestive of long-term independent development. High-cognate subsistence domains, on the other hand, include sea mammal hunting, deep-sea boating, fire-making, general tools, and subsistence terms also used for war. Some of these domains, particularly sea mammal hunting and boating, are also associated with technological advances contributed by the Thule Eskimo. Likewise, some tools associated with war are known to have been reintroduced into the Aleutians with Thule-era influence (Maschner & Mason 2013; Mason 2009; see

discussion in Section 4.3). Together, these observations suggest that independent development cannot alone explain the discrepancy in levels of cognates between semantic domains. I suggest a possible explanation in Section 4.4 below.

4.2 Gender differences in proportions of cognates

Interestingly, the high cognate domains are exclusively domains relating to men's activities, while low cognate domains include those relating to women's activities as well as some men's activities or tools. Further indicative of a gender difference in distribution of cognates, of 60 cognates among the subsistence terms, a handful could refer to men's or women's activities, and another 10 refer to typically women's activities, such as *la-lix* 'to gather' or *cham-six* 'to clean skins for scraping.' However, around 45 refer to men's domains, including names of tools, e.g. *angaaġu-x̣* 'single-bladed paddle for skin boat,' or verbs denoting men's activities, e.g. *sat-*, *sasix* 'to pierce or wound game.' This indicates a difference in language development and change between the gender-based activities in Unangam Tunuu, and thus requires an explanation.

A comparison of cognates between the Yupik and Inuit branches of Eskimo shows no such difference in proportions of cognates between men's and women's subsistence terminology. Not surprisingly given the shallow time-depth of these branches of the Eskimo-Aleut language family, they have a far higher proportion of cognates to non-cognates ranging from 75–83%, as opposed to 15–25% cognates between Unangam Tunuu and Eskimo. However, there is no significant difference in number of cognates in male and female domains, with all subsistence domains in the 75–83% cognate range (Berge 2016). This suggests steady language development and change between Yupik and Inuit. Further, from an Eskimo perspective, different sets of roots are cognate with Unangam Tunuu, suggesting that the respective terminologies developed after the Eskimo and Unangaġ split. For example, Eskimo and Unangam Tunuu have cognate roots for sewing terminology (e.g. 'needle') but they are based on different cognates in the respective languages:

Eskimo sewing terms with Unangan cognates

- (35) PE **qu(C)ayðuləy* 'three edged needle', from PE **quyðar-* 'split with wedge' ~ UT *quxsu-x̣* 'wedge'
- (36) PE **qipðar* 'thread', from PE **qipə-* 'twist' ~ UT *qihmay-* 'be twisted'

Unangan sewing terms with Eskimo cognates

- (37) UT *hinguqa-* 'needle', from *hingu-* 'to push' ~ PE **piŋu-* 'to push'

- (38) UT *yu-n* ‘crimp, seam’, cf. CAY *yuurte-* ‘to curve, bend’ (Fortescue et al. 2010 do not include this as a cognate)

This gender-based difference in Unangam Tunuu subsistence terminology extends to borrowings, as we will see in 4.3. I have argued elsewhere (Berge, submitted), that together with other characteristics of Unangaꝥ divergence from Eskimo, this split in gender-based terminology is suggestive of language mixing. This appears to be supported by genetic studies indicating the replacement of the female line around 1000 BP (Smith et al. 2009; Misarti & Maschner 2015), although my interpretation of the data has Alutiiq men rather than women coming in to the Unangaꝥ-speaking community. Further discussion of this is given in 4.4 and 4.5.

4.3 Correlations in proportions of cognates and borrowings between Eskimo and Unangam Tunuu

About 4%, or 15 of the 373 subsistence terms in Unangam Tunuu, are borrowed from Alutiiq or Central Alaskan Yupik (CAY) according to the sources. Interestingly, if we examine which terms are borrowed and in which semantic domains, we find that those domains with higher percentages of cognates have larger percentages of borrowings than those with fewer numbers of cognates. Likewise, there also seems to be a correlation between gendered activity and the direction of borrowing.

Bergsland (1994: 654ff.) provides the most complete list to date of borrowings so far identified either from or into Unangam Tunuu. He identifies about 110 borrowings between CAY or Alutiiq and (mostly Eastern) Unangam Tunuu; almost 80 have been borrowed into the Yupik languages rather than vice versa, mostly terms for flora and fauna, with some from material culture. Most are not clearly datable; however some obviously predate Russian contact (e.g. UT *kud-machi-x̄* ‘seine,’ borrowed into or from Proto-Yupik before certain sound changes in either language (Bergsland 1986: 123), while others must date from the Russian period, e.g. UT *anach̄xuuda-x̄* ‘sight of gun (rear and front)’ → *anacruq* ‘gunsight’ (Bergsland 1986: 45).

Table 3. Percentages of Eskimo-Aleut borrowings in semantic domains pertaining to subsistence terminology

	CAY / Alutiiq → UT	UT → Alutiiq / CAY	Uncertain direction
Total number of borrowings	20	80	10
Borrowed subsistence terms, % of 373 total UT subsistence terms	6 terms, 1.6%	15 terms, 4%	3 terms, 0.8%
women’s domain	1 term, 0.3%	6 terms, 1.6%	3 terms, 0.8%
men’s domain	5 terms, 1.3%	9 terms, 2.4%	

Among these borrowed subsistence terms, 6 relate to women's domains are borrowed from Unangam Tunuu into Alutiiq (e.g. (39)–(41)), and one is borrowed from Alutiiq or CAY into Unangam Tunuu (42):

- (39) UT *halaya-x̂* 'board on which one cuts auk skins for parkas' → AAY *all'aq, allaaq* 'cutting board for sewing skins'
- (40) UT *kaasxi-x̂* 'post for the squeezers of sewing stand, sewing or weaving stand' → AAY *kas'iq* 'sewing stand'
- (41) UT *asu-x̂* 'cooking pot' → AAY *asuq* 'pot'
- (42) AAY *camru(q)* 'small pieces of driftwood,' CAY *ciamruq* 'small stick used for kindling' → UT *chaamġu-x̂* 'driftwood, piece of wood drifted ashore, picked up for firewood,' *chaamġu-lix* 'to pick up pieces of driftwood'

Borrowed terminology relating to men's activities is more abundant, with 14 terms identified in Bergsland (1994); but not all domains have borrowed equally. From low-cognate domains such as fishing and bird hunting, there is only one identified borrowing from Unangam Tunuu into Alutiiq (43), and there are only two from Alutiiq into Unangam Tunuu (44)–(45):

- (43) UT *dux̂ta-x̂* 'compound bone fishhook, gaffhook' → AAY *uqtaq* 'hookless lure used to attract fish when dipnetting or spearing'
- (44) CAY *nuik* 'dart for hunting birds' → UT *nugi-n* '3-pronged bird dart'
- (45) AAY *kugyuasiq* 'large fish net' → UT *kudmachi-x̂* 'seine net' (the directionality is suggested in Bergsland 1994:654)

On the other hand, domains that have high proportions of proposed cognates, such as boating, general tools, and hunting/war also have higher levels of borrowings. There are 8 such terms borrowed from Unangam Tunuu to Alutiiq or CAY (e.g. (46)–(50)) and 3 from CAY or Alutiiq into Unangam Tunuu (51)–(53):

- (46) UT *kagalu-x̂* 'heel, stern of baidarka' → AAY *kagaluq* 'stern of kayak'
- (47) UT *unaġda-n* 'bottom stringers, smaller longitudinal ribs in baidarka' (probably from *una* demonstrative root 'down there') → AAY *unarat* (pl) 'bottom stringers, smaller longitudinal ribs in baidarka'
- (48) UT *uġalu-x̂* 'spear, lance (for sea lion, whale, war)' → AAY Ch *waloo* 'point of spear' (from an 18th century source, not further attested, orthography non-standard), cf. also CAY *urluveq* 'bow'
- (49) UT *aniix-six* 'to chop with hatchet, axe' → AAY *aniig-* 'to chop, hack, carve with hatchet,' *annil'an* 'hatchet, axe'
- (50) UT *umna-x̂* 'rope, string' → AAY *umnaq* 'rope'

- (51) CAY *aklegaq* ‘spear with float and line attached, bird arrow’ (AAY *akleguaq* ‘barbed spear’) → UT *akliga-ġ* ‘harpoon with bladder float’
- (52) CAY *caniurtaq*, CSY *saninghughtaq* ‘quiver’ → UT *chngayuxta-ġ* ‘quiver’
- (53) AAY *kuukusaaq* ‘unraveled rope’ → UT *kuukusa-ġ* ‘Manila rope’

There is reason to believe that there are far more borrowings than originally assumed, as I discuss in Section 4.4.

4.4 Cognates and Post-Eskimo-Aleut split technology

The high-cognate domains are complicated: what at first glance appears to show strong and ancient cultural connections between the branches of the Eskimo-Aleut language family is actually suspect, and many proposed cognates look more like late borrowings, having either irregular constructions, being too phonologically and semantically similar despite the time depth since the language family split up, and reflecting technologies that were developed after this split. For example, the nominal cognates relating to boats all involve irregular reconstructions; and they show remarkably specific and stable meanings for having supposedly evolved independently over some 4000 years. While commonly and frequently used words such as *iqya-ġ* ‘kayak’ may be assumed to be relatively stable, a word like *angaaġu-ġ* ‘single-bladed paddle for skin boat’ is more suspect. In fact, the latter has an unclear relationship with its assumed cognate PE **aŋuðarun* ‘paddle.’ Like many cognates in this domain, it is morphologically complex and could have resulted either from independent development or inheritance from the proto-language. In this case, *angaaġu-ġ* is assumed to be morphologically derived from *anga-* ‘longitudinal half, match’ and the postbase *-aaġu-*, whose derivation meaning is not immediately clear (Bergsland 1994: 81); no other words derived from *anga-* relate to boating. PE **aŋuðarun*, however, is morphologically transparently from **aŋuðar-* ‘to paddle, row’ (thought to be somehow related to **angu-* ‘to catch’) and the applicative *-un* ‘means with which to ...’ (Fortescue et al. 2010: 37). In fact, there is an Alutiiq word *anguarun* ‘single bladed paddle’ (as well as the verb from which it is derived, *anguar-* ‘to row’), from which the UT *angaaġu-ġ* is phonologically explainable (leveling of a diphthong and loss or reinterpretation of the final /n/). It makes more sense to view this as a loan into Unangan Tunuu from Alutiiq.

In support of this are recent suggestions from other fields that early boat technology in the Aleutians was likely primitive (Maschner 2016: 331), that the kayak may have been a recent technological advance in this area, i.e. within the past 1500 years (Anichtchenko 2012; Maschner 2016: 336), and thus after the split of the Eskimo and Unangan branches of the language family, and that kayaks may have

been associated more with war than with deep-see hunting (Anichtchenko 2012).⁸ Whether or not we accept these suggestions, common boating terms related to the kayak often look like borrowings from neighboring Yupik languages to Unangam Tunuu; these include the very term *iqya-x̂* ‘kayak’ itself, as well as many others that have either irregular correspondences or unusual semantic or phonological identity despite the time depth since the split (these are discussed more fully in Berge, forthcoming):

- (54) PE **qayar* (or **qan'ar*) ‘kayak’ → UT *iqya-x̂* ‘kayak’
- (55) AAY *anguarun* ‘single-bladed paddle’ from PE **aṅuḍarun* ‘paddle’ → UT *angaag̃u x̂* ‘single-bladed paddle’
- (56) PE **paṅər̃un* ‘kayak paddle’ (note AAY *paṅəXsuun* ‘paddle’) → UT *x̂aasi-x* ‘double-bladed paddle for baidarka’
- (57) PE **taprar* ‘skin rope or strap’ (in Inuit dialects ‘cross-strap on kayak deck’) → UT *taam̃g̃aa x̂*, *taam̃xaa x̂* ‘cross-strap on kayak deck’
- (58) PE **kənit-* ‘to soak’ (in Bergsland 1994) or PE **kinər-* ‘to drip dry’ (in Fortescue et al. 2010) → UT *knachi-lix*, *knaxchi-lix* ‘to soak the seams of the skin covering of a baidarka, put cover on baidarka; to dry skin in the sun, oil, and put on baidarka’

The same types of considerations hold for hunting terminology, where many of the proposed cognates or cognate roots may also have been borrowed. For example, there are a very limited group of words assumed to be derivatives of a reconstructed root **ahya-*, cognate with PE **ayay-* ‘thrust or push with a pole’. The derivatives involve obsolete or reconstructed suffixes, as in *ayaqsa-x̂* ‘boy’s play spear’, with the reconstructed suffix **q(u)sa-* being the only instance of this proposed suffix in Bergsland (1994), or as in *ayaqu-x̂* ‘fish spear’ with the reconstructed suffix **-qu-*, found in only one other word in the dictionary. *Ayaqu-x̂* ‘fish spear’ could be an old Eskimo loan or a cognate (Bergsland 1986: 60), although it is listed as a cognate in the later sources. It seems more likely that UT *ayaqu-x̂* is a loan from Alutiiq *ayaqu(q)* ‘harpoon’ rather than the result of an independent 3000–4000 year old history.

8. The belief in an earlier use of kayaks hinges on two things: the discovery of a kayak keel dated to 4000 BP in Greenland (Anichtchenko 2012: 159) and the assumption that the kayak was the boat of choice for travel by sea in the Arctic. However, the earlier use of the kayak in the Aleutians and the Kodiak area is unsubstantiated, as the few boat remains in the archaeological record of the Aleutians date to no earlier than about 1100 BP (Anichtchenko 2012; Maschner 2016); and the open skin boat used for transporting more than 2–3 people is technologically simpler, and likely predates the kayak. The peopling of the Aleutians did not happen with the kayak, but with the open skin boat.

The timing of technological advances and actual usage of hunting tools also suggest convincing reasons for viewing these sets of terms as the result of late borrowing. For example, bow and arrow technology, although known in early Aleutian tradition, fell out of use for more than 2000 years between about 3500 BP and 1300 BP (Maschner & Mason 2013), with crossbows being introduced slightly later, about 1000 BP. The likelihood of high numbers of cognates in this semantic domain is small when compared with the low numbers in domains such as fishing or weaving; yet, as we have seen, bow and arrow terms are 50% or more “cognate”. In fact, close inspection suggests the proposed cognates should be reconsidered. For example, the semantic link between *kaluuda-x̂* ‘crossbow’ and PE **katluy* ‘thunder’ is unlikely, given that crossbows are intended to be silent. UT *kalu-lix* refers exclusively to shooting, not to the sound of thunder, and PE **katluy* ‘thunder’ never develops into Eskimo words relating to bows or arrows. A gun, however, does make a loud sound; and guns were introduced by the Russians. The Russian word *strela* ‘arrow,’ borrowed into Eastern UT as *strila-x̂* ‘aurora,’ is also used in the Russian compound *gromovaya strela* ‘thunderbolt’ (lit. ‘thunderous arrow’), which hints at the probable source of the semantic extension from thunder to bow and arrow terminology. The archaic Alutiiq word *katluk* ‘thunder,’ could, conceivably, have been introduced into Unangam Tunuu with the two Russian senses. If so, it belongs to a much more recent period of language contact, but it nevertheless supports the pattern of directionality of borrowing, from Alutiiq to Unangam Tunuu, of terms from the men’s sphere of activities.

In light of the discussion above, I would reinterpret the proposed hunting cognates in Section 3.2.2 as follows:

- (59) PE **qaqu(r)* ‘shaft of harpoon thrower’ → UT *qaḵuu-* ‘butt of shaft, back part of shaft covered by spear thrower’⁹
- (60) AAY *ayaquq* ‘harpoon’ → UT *ayaqudaax* ‘sea otter spear,’ *ahyaaku-* ‘play dart’
- (61) AAY *ayaruq* ‘walking stick’ → UT *ayaḵu-x̂* ‘walking stick’
- (62) AAY *cayuy-* ‘to pull, tug toward self’ → UT (Atkan) *saygi-x̂* ‘bow’
- (63) AAY *katluy* ‘thunder’ → UT *kaluuda-x̂* ‘crossbow’

If we now examine the list of probable borrowings of uncertain direction in Bergsland (1994: 655), it is most likely that the following, at least, were borrowed into Alutiiq or Yupik from Unangam Tunuu, given the tendencies noted here for women’s terms to come from Unangam Tunuu:

- (64) UT *kaluka-x̂* ‘plate’ → AAY *kaluqaq* ‘basket, casket’

9. No Alutiiq form is attested, and the senses in various Eskimo languages vary greatly.

(65) UT *is̄x̄ati-x̄* ‘grass basket’ → AAY *ishxat*, CAY *issran* ‘carrying bag’

These results are summarized in Table 4.

Table 4. Relative levels of borrowings of subsistence terminology between neighboring Yupik languages and Unangam Tunuu

Borrowings	Alutiiq/CAY → UT	UT → Alutiiq/CAY
Totals	16	17
Women’s domain	1	8
Men’s domain	15	9
<i>Boating</i>	5	3
<i>Hunting/war</i>	7	1 (+ 1 <i>postcontact</i>)
<i>Other (low-cognate)</i>	3	4

In other words, there are clear gender differences in directionality of borrowing, and there is some indication that levels of cognates and borrowings are correlated: low-cognate domains tend to prefer borrowings from Unangam Tunuu into Alutiiq, whereas high-cognate domains tend to have more borrowings from Alutiiq to Unangam Tunuu. Further, the high-cognate domains may in fact not be high in proportion of cognates, but rather in proportion of borrowings.

The pattern is suggestive of Alutiiq men moving into previously Unangā territory, intermarrying with Unangan women, and introducing certain kinds of boat, hunting, and war technology. The non-cognate terms are illustrative of what was not replaced: the larger boats, the fishing equipment, and the tools for hunting from shore (with some exceptions, e.g. a bird dart).

4.5 The motivations for language spread

In this study, I have shown that there is an unequal exchange of subsistence terms between the Unangan and the Alutiit, with boating, hunting, and war terms predominantly flowing from Alutiiq to Unangam Tunuu and rather fewer fishing, weaving, sewing, food and hide processing terms, etc. flowing from Unangam Tunuu to Alutiiq. The archaeology suggests population influx although not population replacement, around 800 BP (although Maschner et al. 2009 argue for the latter), associated with the modern Alutiiq language; the subsistence borrowings suggest specifically an influx of Alutiiq men into an Unangā-speaking area.¹⁰ On

10. Interestingly, one of the few exact cognates among kinship terms between Aleut and Eskimo is UT *ugi-x̄* ‘husband’ ~ PE **uyi* ‘husband’.

Kodiak, Alutiiq may have replaced Unangam Tunuu; in the historical Unanga territory, we see Alutiiq influence from this influx.

The question is then what the motivation for this influx was. There is no evidence of an expansion resulting from resource surpluses out of Kodiak and the Aleutians during this time (although cf. Fitzhugh 2003, who argues for population increase and subsequent increase in social complexity on Kodiak because of such surpluses). Nor do better subsistence strategies as a result of Thule influence alone explain the constellation of borrowings. The Unangan did not adopt Thule whale hunting techniques, for example, although they did hunt more widely in the open ocean, and the cultural practices are primarily shared with the Pacific Coast, and not with the Eskimos. In historic times, the Unangan and the Kodiak Islanders shared unique whaling customs that do not obviously have their sources in the Eskimo north (Lantis 1938:456).

There is, however, evidence of an increase in warfare bringing Eskimo groups from the more sparsely populated north into this region, at the same time as there was a westward expansion of Eastern Unangan along the Aleutian Chain. The entire region was facing upheavals from climate changes (warmer temperatures, increased storminess, environmental stress), volcanic eruptions, and migrations from the north (the Thule) and the east (e.g. the Dena'ina). Mason (2009) argues that the Thule were a military people, responsible for reintroducing the bow and arrow into Alaska, and most especially the crossbow, and that violent confrontations increased around 700 BP. Funk (2010) also highlights a little-known aspect of Alaskan Yupik history known as the Bow and Arrow wars; although it is unclear when they started, they lasted until the 19th century and their effects were felt throughout the Yupik area, down to the Alaskan Peninsula. One hypothesis regarding the origins of these wars involves the displacement of a Yupik tribe from the Norton Sound area, some 500 years BP (Funk 2010:534), at about the same time that Thule slat armor starts to appear in that area (Mason 2009:112). Kari (1989:553) mentions a series of wars between the Central Alaskan Yup'it and Alutiit and the Dena'ina, in which the Dena'ina were driving a wedge between Yupik tribes as they moved toward the coast. The Alutiit were pushed south at a time when Kodiak was environmentally stressed and suffering from a population crash (Maschner et al. 2007; for Fitzhugh 2003, the Alutiit were indigenous to Kodiak). Anichtchenko (2012:159), citing Turner (2008), mentions old Unangan oral traditions suggesting that the kayak had been developed for warfare. The high incidence of kayak terms and terms for weapons used for war (bow, arrow, and certain harpoons with dual functions of hunting and war) are explainable if war was a significant factor in the spread of Alutiiq. Other features of Unangam Tunuu begin to make sense with this scenario. Bergsland (1986, 1994) remarked on the westward expansion of the Eastern dialect, apparently still going on at the time of Russian contact. In Berge (submitted), I

point out the unusual number of synonyms and phonological variability of lexical items, especially in the Eastern dialect. If indeed Maschner 2016 is correct that the Alutiit met a distinct Unangǎ-speaking group on Kodiak Island and pushed them westward, this would explain both the motivation for the westward expansion and the existence of variant forms in Eastern Unangam Tunuu.

Any discussion of warfare in the Kodiak and Aleutian area is incomplete without reference to already existing patterns of warfare along the Pacific Coast. The Thule did not introduce war around 1000 BP: signs of violence are present in the archaeological record for at least 5000 years along the Pacific Coast, and there is evidence of an increase in warfare by about 1300 BP in the coastal areas immediately to the south of Kodiak. The source of violence may have its roots in the generally stressed environment (Maschner et al. 2009); or, it may have been related to relative status and the acquisition of prestige goods, rather than to food production or shortage (Lambert 2002: 215; Maschner & Reedy 2007). The Alutiit arrived in an area with a different tradition of warfare during a time of great upheavals in Alaska and the Pacific Coast.

The proximate cause for language spread at this time was war; ultimately, it resulted from a combination of climate changes and natural disasters leading to population migrations; technological changes brought by new peoples; and existing cultural practices involving status, trade, slavery, and the exchange of surplus goods.

5. Conclusions

A study of the subsistence terminology in Unangam Tunuu and Alutiit indicates that there was no prior tradition of agriculture, and that agriculture was responsible neither for the original spread of Eskimo-Aleut, nor for the most recent instance of language spread in the Unangǎ area, namely the advance of Alutiit and the retreat of Unangam Tunuu. In fact, subsistence activities likely did not cause the latter either. However, although warfare was a major factor in the spread of Alutiit, and in the population movements throughout Alaska at this time, it is by no means the only type of interaction that the Unangan and Alutiit had. Cultural exchanges resulted in the Alutiit adopting some fishing, fish and animal processing, and weaving or sewing terminology, and a very few boating and hunting terms, while giving harpooning and boating technology to the Unangan. While this technology was used in wars, it also facilitated subsistence activities and as such had a huge effect on Unangǎ culture.

The results of this research seem to support Dumond's (1987, 2001) and Maschner's (2016) positions that Alutiit culture in the Kodiak area replaced a prior Unangǎ culture around 800 BP. This differs from long held beliefs that the

prehistoric boundary between Pacific Yupik and Unangan groups is further west on the Alaskan Peninsula, and that Kodiak was always Alutiiq territory. It also provides a motivation for the recent westward expansion of Eastern Unangan Tunuu (Woodbury 1984; Bergsland 1986; Berge 2010) and explains the Yupik influence we see in Unangaʔ vocabulary.

The results do not support genetic research (Smith et al. 2009) or archaeological research (Misarti & Maschner 2015) suggesting the replacement of the female line in the Unangaʔ area, although there is a gender-based split in subsistence terminology. As argued above, the results suggest an influx of Yupik men. This does not disprove Smith et al. 2009, however. There are ethnographic records of frequent raids involving the capture of women and children (Maschner & Reedy 2007), which in principle could lead to the replacement of female terminology. If the female line was replaced, it is manifested differently in the language, as there is as yet no compelling evidence for Denaʔina, Eyak, or other borrowings into Unangan Tunuu to explain this state of affairs. There is evidence of a region-wide system of lexical replacement involving language internal word constructions (cf. Kari 2013 for elite replacements in Denaʔina, and Berge & Holton 2015); however, this requires further research.

Finally, some wars involved a struggle for insufficient resources and some involved instead the establishment of status. They may have had different effects on language use, ranging from language replacement to various degrees of a linguistic area, in some cases involving both at the same time. One might imagine looking at different lexical domains for evidence of this, e.g. terms relating to social status.

Acknowledgements

This research was conducted with the help of student assistant Josh Pharris under a grant from the Linguistics Program of the University of Alaska Fairbanks.

Abbreviations

AAʔ	Alutiiq	PI	Proto-Inuit
AI	Alaskan Inuit	pl	plural
CAY	Central Alaskan Yupʔik	PY	Proto-Yupik
du	dual	UT	Unangan Tunuu
EA	Eskimo-Aleut	WG	West Greenlandic
PE	Proto-Eskimo		

References

- Anichtchenko, Evguenia. 2012. Open skin boats of the Aleutians, Kodiak Island, and Prince William Sound. *Études/Inuit/Studies* 36(1): 157–181. doi:10.7202/1015958ar
- Bellwood, Peter. 2011. *First Migrants: Ancient Migration in Global Perspective*. Malden MA: Blackwell.
- Berge, Anna. forthcoming. Reevaluating the reconstruction of Proto-Eskimo-Aleut. *Journal of Historical Linguistics*.
- Berge, Anna. submitted. Reexamining the linguistic prehistory of Aleut (Unangam Tunuu). In *Archaeology and Language – And the Future of Archaeo-linguistic Studies. Selected papers from the Proceedings of the XV Nordic TAG, Copenhagen, Denmark, April 16–18, 2015*, Rune Iversen & Guus Kroonen (eds). Oxford: British Archaeological Reports (BAR).
- Berge, Anna. Forthcoming. Language contact in Aleut. In *Handbook of Language Contact*, Anthony Grant (ed.). Oxford: OUP.
- Berge, Anna. 2016. *Lexical Differentiation in Aleut (Unangam Tunuu)*. *Frontiers of Early Human Expansion in Asia: Linguistic Perspectives on Ainu, Japan, and the North Pacific Rim, Zurich, March 13–14*. <https://www.academia.edu/30870719/Lexical_Differentiation_in_Aleut_Unangam_Tunuu_>.
- Berge, Anna & Holton, Gary. 2015. Toward a history of the Aleut language (Unangam Tunuu). Paper presented at the conference of the Societas Linguistica Europæa (SLE), Leiden, the Netherlands, September 2–5, 2015. Hdt posted on <https://www.academia.edu/15391546/Toward_a_history_of_the_Aleut_language_Unangam_Tunuu_>.
- Berge, Anna. 2010. Origins of linguistic diversity in the Aleutian Islands. *Human Biology* 84(4): 557–582. (October–December 2010 special issue).
- Bergsland, Knut. 1986. Comparative Eskimo-Aleut phonology and lexicon. *Journal de la Société Finno-ougrienne* 80: 63–137.
- Bergsland, Knut. 1989. Comparative aspects of Aleut syntax. *Journal de la Société Finno-ougrienne* 82: 7–80.
- Bergsland, Knut, compiler. 1994. *Aleut Dictionary: Unangam Tunudgusii*. Fairbanks AK: ANLC.
- Byers, David A., Yesner, David R., Broughton, Jack M. & Coltrain, Joan Brenner. 2011. Stable isotope chemistry, population histories, and late prehistoric subsistence change in the Aleutian Islands. *Journal of Archaeological Science* 38: 183–196. doi:10.1016/j.jas.2010.09.005
- Davis, Richard S. & Knecht, Richard A. 2010. Continuity and change in the Eastern Aleutian archaeological sequence. In *Special Issue on the Origins of the Populations of the Aleutian Islands*, Michael H. Crawford, Dixie L. West, & Dennis H. O'Rourke (eds), 507–524. *Human Biology* 82(5–6).
- Dumond, Don. 1987. A reexamination of Aleut prehistory. *American Anthropologist* 89(1): 32–56. doi:10.1525/aa.1987.89.1.02a00020
- Dumond, Don. 2001. Toward a (yet) newer view of the (pre)history of the Aleutians. In *Archaeology in the Aleut Zone of Alaska, Some Recent Research* [University of Oregon Anthropological Papers 58], Don E. Dumond (ed.), 289–309. Eugene OR: University of Oregon Press.
- Erlandson, Jon M. 2001. The archaeology of aquatic adaptations: Paradigms for a new millenium. *Journal of Archaeological Research* 9(4): 287–350. doi:10.1023/A:1013062712695
- Fitzhugh, Ben. 2003. *The Evolution of Complex Hunter-Gatherers: Archaeological Evidence from the North Pacific*. Dordrecht: Kluwer. doi:10.1007/978-1-4615-0137-4

- Fortescue, Michael. 1998. *Language Relations Across Bering Strait: Reappraising the Archaeological and Linguistic Evidence*. London: Cassell.
- Fortescue, Michael, Jacobson, Steven & Kaplan, Lawrence. 2010. *Comparative Eskimo Dictionary with Aleut Cognates*, 2nd edn. Fairbanks AK: ANLC.
- Funk, Caroline. 2010. The bow and arrow war days on the Yukon-Kuskokwim Delta. *Ethnohistory* 57(4): 523–569. doi:10.1215/00141801-2010-036
- Heggarty, Paul. 2015. Prehistory through language and archaeology. In *The Routledge Handbook of Historical Linguistics*, Claire Bowern & Bethwyn Evans (eds), 595–626. London: Routledge.
- Kari, James. 1989. Some linguistic insights into Dena'ina prehistory. In *Athabaskan Linguistics: Current Perspectives on a Language Family*, Eung-Do Cook & Keren Rice (eds), 533–574. Berlin: Mouton de Gruyter.
- Kari, James. 2013. *Dena'ina Topical Dictionary*. Fairbanks AK: ANLC.
- Lambert, Patricia M. 2002. The archaeology of war. *Journal of Archaeological Research* 10(3): 207–241. doi:10.1023/A:1016063710831
- Lantis, Margaret. 1938. The Alaskan whale hunt and its affinities. *American Anthropologist* 40(3): 438–464. doi:10.1525/aa.1938.40.3.02a00070
- Leer, Jeff. 1991. Evidence for a Northwest Coast language area: Promiscuous number marking and periphrastic possessive constructions in Haida, Eyak, and Aleut. *IJAL* 57(2): 158–193.
- Liapunova, Roza G. 1996. *Essays on the Ethnography of the Aleuts: At the End of the Eighteenth and the First Half of the Nineteenth Century*, Jerry Shelest (transl.), William B. Workman & Lydia T. Black (eds). Fairbanks AK: University of Alaska.
- Maschner, Herbert D. G. 2016. Archaeology of the Eastern Aleut region. In *The Oxford Handbook of the Prehistoric Arctic*, T. Max Friesen & Owen K. Mason (eds), 323–348. Oxford: OUP.
- Maschner, Herbert & Mason, Owen K. 2013. The bow and arrow in northern North America. *Evolutionary Anthropology* 22: 133–138. doi:10.1002/evan.21357
- Maschner, Herbert D. G. & Reedy-Maschner, Katherine L. 1998. Raid, retreat, defend (repeat): The archaeology and ethnohistory of warfare on the North Pacific Rim. *Journal of Anthropological Archaeology* 17: 19–51. doi:10.1006/jaar.1997.0315
- Maschner, Herbert D. G. & Reedy, Katherine L. 2007. Heads, women, and the baubles of prestige: Trophies of war in the Arctic and Subarctic. In *The Taking and Displaying of Human Body Parts as Trophies by Amerindians*, Richard J. Chacon & David H. Dye (eds), 32–44. New York NY: Springer. doi:10.1007/978-0-387-48303-0_3
- Maschner, Herbert D. G. et al. 2009. Did the North Pacific ecosystem collapse in AD 1250? In *The Northern World, AD 900–1400*, Herbert Maschner, Owen Mason & Robert McGhee (eds), 33–57. Salt Lake UT: University of Utah.
- Mason, Owen K. 2009. Flight from the Bering Strait: Did Siberian Punuk/Thule military cadres conquer Northwest Alaska? In *The Northern World, AD 900–1400*, Herbert Maschner, Owen Mason & Robert McGhee (eds), 76–127. Salt Lake UT: University of Utah.
- Misarti, Nicole & Maschner, Herbert D. G. 2015. The Paleo- to Neo-Aleut transition revisited. *Journal of Anthropological Archaeology* 37: 67–84. doi:10.1016/j.jaa.2014.12.004
- Potter, Ben. 2010. Archaeological patterning in Northeast Asia and Northwest North America: An examination of the Dene-Yeniseian hypothesis. *Anthropological Papers of the University of Alaska, New Series* 5(1–2): 138–167.
- Renfrew, Colin. 1987. *Archaeology and Language: The Puzzle of Indo-European Origin*. London: Jonathan Cape.
- Ross, Malcolm. 2003. Diagnosing prehistoric language contact. In *Motives for Language Change*, Raymond Hickey (ed.), 174–198. Cambridge: CUP. doi:10.1017/CBO9780511486937.012

- Smith, Silvia E. *et al.* 2009. Inferring population continuity versus replacement with a DNA: A cautionary tale from the Aleutian Islands. *Human Biology* 31(4): 407–426.
doi:10.3378/027.081.0402
- Steffian, Amy, Saltonstall, Patrick & Yarborough, Linda Finn. 2016. Maritime economies of the central gulf of Alaska after 4000 B.P. In *The Oxford Handbook of the Prehistoric Arctic*, T. Max Friesen & Owen K. Mason (eds), 303–322. Oxford: OUP.
- Turner, Lucien M. 2008. *An Aleutian Ethnography*. Raymond L. Hudson (ed.). Fairbanks AK: University of Alaska Press.
- Woodbury, Anthony. 1984. Eskimo and Aleut languages. In *Handbook of North American Indians 5: Arctic*, David Damas (ed.), 49–63. Washington DC: Smithsonian.

Lexical recycling as a lens onto shared Japano-Koreanic agriculture

Alexander Francis-Ratte

Furman University

Despite the existence of strong cognates in other realms of basic vocabulary, it remains unclear why Korean and Japanese share so few words for grain and agriculture. This paper proposes that pre-rice vocabulary has undergone a process of lexical recycling in Korean to refer to later rice-related practices. The observation that Korean words for ‘rice’ contain initial *p* suggests common derivations from pre-MK **po* ‘rice(?)’ that is relatable to Old Japanese *po* ‘a grain’. This paper uncovers important Japano-Koreanic cognates, including ‘buckwheat,’ ‘millet,’ and ‘rice plant’. This analysis also shows how linguists may retrieve early agricultural terminology that has been replaced by more advanced practices.

Keywords: Japano-Koreanic, proto-Korean-Japanese, rice, lexical recycling, historical linguistics

1. Introduction

Rice is the primary staple grain of Japan and Korea, and has an important place in Japanese and Korean traditional culture far outstripping that of other grains. For example, rice cakes made from pounded rice flour (Japanese *mochi*, Korean *ttök*) are essential elements of festivals celebrating the new year and other important events. Rice is also a feature of specific ritual or religious practices, such as the offering of washed rice (known as *araiyone* or *senmai*)¹ and *mochi* to deities in Japan, or the Korean shamanistic practice of dedicating rice cakes (known as *kosa'ttök*) to spirits (Jeremy & Robinson 1989; Lee 1981: 162).

Under the framework for Japano-Koreanic common origin put forth by Whitman (1985) and expanded by Unger (2009), Japanese and Korean share many

1. Washed rice offerings are also known as *kumashine*, *kashiyone*, or *okuma* in premodern Japanese.

convincing, phonologically unproblematic correspondences for basic vocabulary. Cognates include ‘fire’ (OJ² *pwi* ~ MK *pul*), ‘mountain’ (OJ *yama* ~ NK *yem* ‘rocks sticking out of water’), and ‘wood’ (OJ *kwi* ‘tree, wood’ ~ MK *kuluh* ‘stump’), among others (see also Whitman 2012 and Francis-Ratte 2016 for recent Japano-Koreanic cognates). Yet despite the outsized importance of rice in Japanese and Korean cultures, there are few if any obvious cognates to be found among Old Japanese and Middle Korean words for rice, or any other staple grain:

Table 1. Agricultural vocabulary (1)

Gloss	OJ	MK
riceplant	<i>ine</i> /ine/	<i>pyé</i> /pjə/
uncooked rice	<i>yone</i> /jone/ <i>komey</i> /kəmej/	<i>psól</i> /psal/
cooked rice	<i>ipi</i> /ipi/	<i>pap</i> /pap/
millet	<i>apa</i> /apa/	<i>cwoh</i> /tsoh/
buckwheat	<i>swoba</i> /so ^m ba/	<i>mwomilh</i> /momilh/

Not a single item in the vocabulary of grains in Table 1 displays a direct Japano-Koreanic correspondence. The absence of straightforward cognates for cultivars is surprising for two languages that share cognates in other realms. More curious still is the observation that there are strong etymologies for farming words such as ‘plot’ and ‘field’:³

Table 2. Agricultural vocabulary (2)

Gloss	OJ	MK
agricultural plot	<i>mati</i> /mati/	<i>math</i> /mat ^h /
farm field	<i>pata</i> /pata/ <i>patakey</i> /patakej/	<i>path</i> /pat ^h /
field ⁴	<i>ta</i> /ta/	<i>tulh</i> /tilh/

2. OJ = Old Japanese, ca. 8th century CE; EMJ = Early Middle Japanese, ca. 9–11th centuries CE; MK = (Late) Middle Korean, ca. 15–16th centuries CE; NK = Modern Korean, ca. 17th centuries to present. Old Japanese citations are from Omodaka et al. (1967), and Middle Korean citations are from Nam (1997).

3. These are unlikely to be borrowings; for example, OJ *mati* is unlikely to be a borrowing from Korean *math*, as the comparison is dependent upon a sound change (sonorant yodization) that took place before the differentiation of Japonic.

4. OJ *ta* means ‘paddy field’; MK *tulh* / *tulúh* refers to open fields but also means ‘wild’. The two forms are a phonological match provided that Korean *tulh* be analyzed as incorporating a locative suffix **-l(o)h*.

Cognates in Table 2 indicate that speakers of the source language from which Japanese and Korean diverged (henceforth “proto-Korean-Japanese” or pKJ)⁵ probably were agriculturalists, but may not have cultivated the range of crops that later Japanese and Korean speakers did. Thus in spite of convincing cognates in other realms of basic vocabulary, and the existence of cognates for ‘farm’ and ‘field,’ there are no straightforwardly identifiable cognates for grain cultivar crops. This is an exceedingly odd state of affairs.

Previous scholarship (Vovin 1998; Unger 2009; Whitman 2011) cites the absence of rice-related cognates to posit that proto-Korean-Japanese predates the adoption of wet (paddy) rice agriculture in Northeast Asia. According to Whitman (2011), this places proto-Korean-Japanese some time before 1500 BCE (the first appearance of paddy rice cultivation), and perhaps before 2500 BCE (the first appearance of non-paddy rice cultivation). And yet, the absence of not just rice vocabulary but any grain cultivars in the proto-Korean-Japanese lexicon indicates that our understanding of Japano-Koreanic agriculture is far from complete.

This paper addresses questions regarding agricultural vocabulary in Japanese and Korean in two ways. First, I buttress the theory of proto-Korean-Japanese by proposing hitherto undiscovered cognates in the realm of grain agriculture. Proposing phonologically unproblematic correspondences for ‘rice,’ ‘millet,’ ‘cereal,’ and ‘grain’ addresses a key implausibility in the theory of Japano-Koreanic common origin, namely the absence of cultivar words in a culture that must have practiced agriculture. Moreover, the existence of a cognate set for ‘rice’ or ‘rice plant’ suggests that Japanese and Korean may have diverged at a time when field rice (but not paddy rice) was already being cultivated in Northeast Asia alongside millet. Second, I show that despite later technological developments that have obscured the original relationships, historical linguistic methodologies can indeed be employed to reconstruct the language of the first farmers of Korea and Japan. I suggest that a pattern of “lexical recycling” can be detected in the Korean lexicon, whereby pre-technological, pre-rice words from a proto-Korean-Japanese stratum have been repurposed as post-technological, post-rice words in proto-Korean.

Section 2 discusses the origins of two Old Japanese rice words. Section 3 analyzes three words in Middle Korean relating to rice, and proposes an internal analysis that reveals strong etymologies with Japanese. Section 4 discusses the possibility of a relationship between MK *cwoh* ‘millet’ and OJ *swoba* ‘buckwheat’. Section 5 discusses the possibility of a pattern of lexical recycling in Korean, and rebuts the idea that the agricultural vocabulary analyzed in this paper are cases of borrowing. Section 6 concludes with implications for proto-Korean-Japanese and discussion

5. The term “proto-Korean-Japanese” should be considered synonymous with “proto-Japano-Koreanic” (pJK) and other similar terms used by other scholars.

of how historical linguistic methodology reveals insights into a pre-rice period of East Asian history.

Under the framework for Japano-Koreanic common origin (Francis-Ratte 2016), most phonemes have one-to-one correspondences between proto-Japanese (pJ) and proto-Korean (pK). Non-trivial correspondences relevant to this paper are listed in Tables 3 and 4.

Table 3. Non-trivial consonant correspondences

proto-Japanese		proto-Korean		pKJ
*s	(OJ <i>s</i>)	*ts	(MK <i>c</i>)	*ts
*N	(OJ prenasalization)	*ŋ/*h	(MK <i>h</i>) ⁶	*ŋ
*r (non-final)/*j (final)		*r	(MK <i>l</i>)	*r

Table 4. Non-trivial vowel correspondences

proto-Japanese		proto-Korean		pKJ
*ə	(OJ <i>o</i>)	*ə	(MK <i>o</i> ; <i>e</i> before <i>γ</i>)	*ə

2. Analysis of OJ *ine* ‘rice plant’ and *yone* ‘hulled, uncooked rice’

Before discussing Korean words for rice and grains and their correspondences to Japanese, it is important to first examine both the Korean and Japanese lexicons for possible internal reconstructions of rice-related vocabulary. In this section, I propose that two Old Japanese words relating to rice, OJ *ine* ‘rice plant’ and *yone* ‘hulled, uncooked rice’, are divergent forms from the same proto-Japanese/pre-proto-Japanese form *jənaɟ ‘rice-plant’. This form in turn derives from a compound of reconstructed *jə ‘rice’ and *naɟ ‘plant, root’ (cf. OJ *ne* ‘id.’); the relationship of this reconstructed *jə to Korean will be discussed in Section 3.

OJ *ine* ‘rice plant’ and *yone* ‘hulled, uncooked rice’ are not thought to be etymologically related (Omodaka et al. 1967; Martin 1987), and Vovin (1998:370) categorically denies any possibility of a relationship between the two. However, there is some evidence in the Japanese lexicon that hints that these two forms may once have been one and the same word. First, OJ *ine* ‘rice plant’ is attested in some

6. Francis-Ratte (2016: 31) proposes that alternations of Korean nasal *ng* with syllable-final *h/k* (e.g. NK *matang* ‘yard’, MK *math* ‘plot’) point to original *ŋ and a general merger of all voiced velars in pre-Middle Korean, a theory that is supported by cognates with Japanese. Thus, pK *ŋ is regularly reflected as MK *-h* in word-final position.

compounds where we expect *yone* ‘hulled rice,’ a curious observation that is difficult to explain if the two forms are not somehow related. For example, Japanese *urusine* ‘non-glutinous rice grain’ (cf. OJ *uru* ‘moist’) appears to be a compound with OJ *ine* ‘rice plant’. The reason for this excrescent *-s-* is unclear, but it appears to be introduced to prevent hiatus when the second compound element begins with a vowel; other well-known examples include OJ *parusame* ‘spring rain’ from *paru* ‘spring’ + *ame* ‘rain,’ and OJ *kwosame* ‘drizzle’ from OJ *kwo* ‘small’ + *ame* ‘rain’ (but note Early Middle Japanese *nagame* ‘long spell of rain’ with *naga* ‘long,’ not ***nagasame*). Japanese *urusine* refers to the rice *grain* used to make non-glutinous rice (that is, not used for rice cakes), so it is curious that this compound should contain OJ *ine* ‘rice plant’ as opposed to OJ *yone* ‘hulled rice grain’. The compound would be more logical in its construction had (s)*ine* meant *‘rice grain,’ not ‘rice plant’.

Even more curious is the Early Middle Japanese (EMJ) compound *kumasine* ‘washed offerings of rice to gods’. Like *urusine*, this word seems to be a compound containing *ine* ‘rice plant’.⁷ However, it seems particularly problematic that this compound should be formed with *ine* ‘rice plant,’ given that it is the rice *grain* that is washed and offered, not the stem and leaf of the plant. Moreover, synonyms of EMJ *kumasine* are formed with *yone* or *kome*, both ‘hulled rice grain,’ such as *kasiyone*, *kasigome*, and *araigome*. EMJ *kumasine* would be far more logical in its construction were (s)*ine* to be interpreted as *‘rice grain’ as opposed to ‘rice plant’.

Then there is the observation that *ine* and *yone* are suspiciously similar in both their meaning and their phonological form, separated only by a vowel *o* (pJ ?*ə). That both are <rice> words cannot be attributed to the common suffix *ne* ‘root, plant,’ since this suffix is present in many other non-rice plant words (e.g. *akane* ‘madder, *Rubia argyi*’). Therefore, the meaning <rice> can only be a contribution from the initial compound elements **i* and **yo*.

To account for the phonological similarity of OJ *ine* and *yone*, and the existence of compounds wherein (s)*ine* appears to mean *‘rice grain’ as opposed to ‘rice plant,’ I hypothesize that OJ *ine* ‘rice plant’ derives from the same proto-Japanese form as does OJ *yone* ‘hulled rice’. Assuming that *yone* ‘hulled rice’ and *ine* ‘rice plant’ both incorporate the common Japanese suffix *ne* ‘root, plant,’ I reconstruct OJ *yone/ine* < pJ *jə-naj ‘rice-plant,’ a proto-Japanese lexicalization that incorporates an earlier form *jə meaning ‘rice’.

By regular sound change, the expected reflex of *jə-naj ‘rice-plant’ is OJ *yone* ‘hulled rice’. Why then did these two forms phonologically diverge? I hypothesize

7. The first compound element *kuma* is probably an archaic nominal form meaning ‘bestowal’ derived from the same root as OJ *kubar-* ‘bestows’ (for an analysis, see Francis-Ratte 2016: 192). The fact that EMJ *kumasine* contains a very old derivation *kuma* implies *kumasine* is also a very old term in Japanese, and that its absence from the Old Japanese corpus may be accidental.

that (pre-)proto-Japanese *jə-naj ‘rice-plant’ underwent a slight phonological change to pJ *i-naj when prominence was given to the second compound element *naj ‘root, plant,’ a process that caused the de-emphasized glide-vowel combination of the initial syllable to condense to *i. In other words, OJ *ine* represents a lineage of pre-pJ *jə-naj ‘rice-plant’ emphasizing ‘rice-*plant*’ (with concomitant collapse of *jə > *i), whereas OJ *yone* represents a lineage emphasizing ‘*rice*-plant’ (with semantic bleaching of the suffix *-ne*). A hypothetical shift of *jə to *i is not wholly ad hoc, as other examples in Old Japanese show yV sequences alternating with a unitary front vowel:

- (1) OJ *yume* ‘dream’ ~ *ime* ‘dream’ ~ *i* ‘sleep’ (pJ *ju)
- (2) OJ *yu(duru)* ‘bow(string)’ ~ *i-(ru)* ‘shoots it’ (pJ *ju)
- (3) OJ *yo-(si)* ‘is good’ ~ *e-(si)* ‘is good’ (pJ *jər?)

These phonological alternations suggest a limited but definite sound change in pre-Japanese, where yV sequences condensed to a front vowel in as yet unknown environments. The alternation of OJ *ine* with *yone* fits this pattern, and points to pJ *jənaj. Given that the second element *naj of this pJ form *jənaj ‘rice plant’ is clearly identical to pJ *naj ‘plant, root’ (OJ *ne*), I reconstruct (pre)-proto-Japanese *jə as a word that denoted ‘rice’ or possibly ‘rice plant’. The possible relationship of this form *jə to Korean will be discussed in Section 3.

3. Analysis of Korean rice vocabulary

As noted in the Introduction to this chapter, there are no straightforward Japano-Koreanic correspondences for grain cultivars, including for rice. However, this section will show that an elegant internal analysis of Korean rice words unlocks a trove of Japano-Koreanic cognates in agricultural vocabulary.

It is an odd fact that three of the most important <rice> words in Middle Korean begin with the consonant /p/, namely MK *pyé* ‘rice plant,’ *psól* ‘uncooked rice,’ and *pap* ‘cooked rice.’ From the perspective of internal reconstruction, positing that initial *p* constitutes (part of) an etymological prefix would explain the semantic unity of these words as a result of a common prefix, which we can reconstruct as *po⁸ meaning ‘rice.’ This is to say, there are Korean-internal reasons for supposing that these three words relating to ‘rice’ all contain a common derivational prefix

8. Words must be minimally V or CV in Korean, so it is most reasonable to reconstruct a minimal vowel *o following *p-* that has undergone syncope.

**p(o)* denoting ‘rice’. This hypothesis points to pre-MK **p(o)-yé* ‘(rice)plant’, **p(o)-sól* ‘(uncooked) grain’, and **p(o)-ap* ‘(cooked) grain’ as shown in Table 5:

Table 5. Etymologies with pre-MK **po*

MK	<i>pyé</i> ‘rice plant’	<i>psól</i> ‘uncooked rice’	<i>pap</i> ‘cooked rice’
pre-MK	* <i>po-yé</i>	* <i>po-sól</i>	* <i>po-ap</i>
	* <i>yé</i> ‘(rice) plant’	* <i>sól</i> ‘(hulled) grain’	* <i>ap</i> ‘(cooked) grain’

By itself, the hypothesis that these three words contain a lost morpheme **p(o)* has power to explain their shared semantics, but is otherwise speculative. However, a striking piece of evidence for the hypothesis emerges when we compare this **po* ‘rice(?)’ as well the reconstructed pre-MK forms to which it has become fused, **yé*, **sól*, and **ap*, to Japanese. This simple hypothesis unlocks three phonologically perfect lexical matches with Old Japanese, detailed in Table 6 and discussed in Sections 3.1, 3.2, and 3.3.

Table 6. Etymologies with pre-MK **po* and Japanese cognates

MK	<i>pyé</i> ‘rice plant’	<i>psól</i> ‘uncooked rice’	<i>pap</i> ‘cooked rice’
pre-MK	* <i>po-yé</i>	* <i>po-sól</i>	* <i>po-ap</i>
	* <i>yé</i> ‘(rice) plant’	* <i>sól</i> ‘(hulled) grain’	* <i>ap</i> ‘(cooked) grain’
OJ	<i>po</i> ‘a grain’	(<i>wase</i> ‘early growth’)	<i>apa</i> ‘millet’
	<i>yo(ne)</i> ‘rice plant’		

3.1 Pre-MK **yé* ~ OJ *yone*

Separating MK *pyé* ‘rice plant’ into pre-MK **po-ye* ?‘(rice)-plant’ reveals a phonologically perfect correspondence between the pre-Middle Korean prefix **po* ?‘rice’ and OJ *po*⁹ ‘a grain, an ear’. Comparison of the two forms implies that a semantic shift has occurred, most likely in the Korean reflex (see Section 5). An etymological analysis of MK *pyé* as **po-yé* also unlocks a perfect correspondence between pre-MK **yé* ‘(rice) plant’ (pK *jə) and the initial syllable *yo-* of OJ *yone*¹⁰ ‘hulled rice’, which was analyzed in Section 2 as proto-Japanese *jə.

9. Following Unger (2007), I am unsure whether the OJ vowels *wo* /o/ and *o* /ə/ can be reliably distinguished before labial *p*; the transcription *po* denotes that either **po* or **pə* are possible precursors in pre-Old Japanese.

10. This hypothesis assumes reasonably that *ne* of OJ *yone* incorporates OJ *ne* ‘root, plant’ as a suffix.

3.2 Pre-MK **ap* ~ OJ *apa*

Separating MK *pap* ‘cooked rice’ into pre-MK **po-ap* ?‘rice-cereal’ reveals a correspondence of **ap* to OJ *apa* ‘millet’.¹¹ This etymology of MK *pap* is superior to that of Martin (1966), who compares *pap* directly to OJ *apa* by means of an ad hoc consonant correspondence of Korean /p/ that is unsupported in other cognates. At first glance, the semantic difference between MK *pap* ‘cooked rice’ and OJ *apa* ‘millet’ seems problematic. However, the hypothesis that MK *pap* ‘cooked rice’ incorporates the pre-MK prefix **po* for ‘rice’ indicates that its pre-MK nominal root **ap* may not have originally referred to cooked rice at all, but to cooked grains. Moreover, indirect correspondences of pre-rice and post-rice vocabulary are precisely what we expect if Japanese and Korean diverged before the advent of wet rice agriculture, as they likely did. I hypothesize that pKJ **apa* was a word for common grain cereals that were cooked for consumption, which in proto-Korean-Japanese culture was the grain harvested from foxtail millet (*Setaria italica*). After the adoption of wet rice in both cultures, this word was retained as a word for millet in pre-Japanese, but was repurposed with prefix **pə* in pre-Korean to refer to the new cereal of choice for cooking, namely rice.

3.3 Pre-MK **sól* ~ OJ *wase*

Separating MK *psól* ‘uncooked rice’ into pre-MK **po-sól* ?‘rice-(unknown)’ is more problematic. The second element **sól* of **po-sól* is of uncertain origin, and there is no perfectly corresponding form in Old Japanese that can be identified as a direct cognate to **sól* (we expect OJ ***se*). However, there are several interesting possibilities for the etymology of *psól*.

Vovin (2015) proposes that MK *psól* is actually a borrowing from a precursor to OJ *wase* ‘early growth,’ a possibly unparalleled example of a very early borrowing from Japanese into Korean. If correct, Vovin’s hypothesis would invalidate the hypothesis of separable *p-* in this word. This etymology is somewhat problematic. While there is ample evidence of Japanese-Korean language contact in the 1st millennium CE, there is no real evidence that pre-Japanese speakers on the Korean peninsula were in linguistic contact with pre-Korean speakers before the Yayoi Migrations in the first millennium BCE. But even if substantial contact did take place between these groups, it seems likely that agricultural practices and other technologies were transmitted from north to south on the Korean peninsula. Since

11. There is a very strong likelihood that Korean has undergone final vowel loss, which allows these forms to constitute a perfect phonological match.

pre-Japanese speakers almost certainly inhabited the south of the Korean peninsula before pre-Korean speakers did (Unger 2009, 2014), it is more likely that (pre-Japanese) inhabitants of the south appropriated agricultural practices from (pre-Korean) inhabitants of the north, not the other way around. It seems quite unlikely that pre-Korean speakers borrowed a word for ‘rice’ from pre-Japanese speakers who adopted this crop later. In virtually every other analysis of Japano-Koreanic lexical correspondences, Vovin proposes importation in precisely the opposite direction, namely from Korean into Japanese (e.g. Vovin 2010). I am unaware of other likely borrowings from pre-Japanese into pre-Korean, in rice agriculture or any realm of vocabulary.¹²

For these reasons, I deem Vovin’s (2015) theory of *psól* to be an implausible explanation. On the other hand, MK *psól* ‘uncooked rice’ could incorporate pre-MK **po* ‘rice’ and still be related to OJ *wase*. MK *psól* could be a compound of **pə* ‘rice; grain’ + **wasər* ‘growth, shoot’ (OJ *wase* ‘early growth’), via a truncation of **pə-wasər* ‘rice-shoot’ > **pasər*¹³ and finally to MK *psól*.¹⁴ Such a shift would assume a loss of initial **w* of **wasər*, which is phonetically natural following rounded /p/. This analysis of *psól* ‘uncooked rice’ is slightly more speculative than those proposed for other compounds of **po*, but the hypothesis of separable *p-* in Middle Korean <rice> vocabulary remains attractive for its explanatory power regarding MK *pyé* ‘rice plant’ and MK *pap* ‘cooked rice,’ and for the cognates it uncovers with Japanese.¹⁵

12. It appears that in this one case alone, the direction of importation has been reversed, because Vovin correctly understands that Japanese /w/ cannot correspond to Korean /p/ in a borrowing from Korean into Japanese. This is because Japanese has always possessed both /w/ and /p/ as contrasting phonemes, so any borrowing of Korean /p/ would never be reflected as Japanese /w/. One gets the sense in Vovin’s analysis that any relationship of the two forms other than borrowing has been precluded from the realm of possibility.

13. Note that Vovin (2015) also posits a weakening or minimalization of proto-Korean **a* > MK *o* > zero in the initial syllable of this form.

14. Alternatively, if OJ *wase* ‘early growth’ is a truncation of **wakase* (with OJ *waka* ‘young’), then pre-OJ **se* may have been a word unto itself meaning ‘shoot, plant’. Pre-OJ **se* forms a straightforward correspondence to the **sól* of MK *psól* ‘uncooked rice’.

15. Another lexicalized compound of pre-MK **po* ‘rice’ (from pKJ **a* ‘grain’) may be MK *psi* ‘seed, pit’. Connecting the *p-* of *psi* ‘seed’ to the prefix of <rice> words *pyé*, *psól*, and *pap* would imply that the meaning of the shared prefix **pə* in the oldest stratum of Korean vocabulary was not simply ‘rice’ but closer to ‘a small granule, a grain’. This links up to the hypothesis in Section 5 that its usage to denote ‘rice’ represents a later lexical stratum.

3.4 Pre-MK **po*: Conclusions

The analysis of separable **po* (pK **pə*) in Middle Korean words for <rice>, and the comparisons that this hypothesis enables to Old Japanese, point to at least three new proto-Korean-Japanese reconstructions in the realm of rice agriculture, as shown in Table 7:

Table 7. Etymologies with pre-MK **po* and pKJ reconstructions

pre-MK	* <i>po-yé</i> * <i>yé</i> ‘(rice) plant’	* <i>po-sól</i> * <i>sól</i> ‘(hulled) grain’	* <i>po-ap</i> * <i>ap</i> ‘(cooked) grain’
OJ	<i>po</i> ‘a grain’ <i>yo(ne)</i> ‘rice plant’	(<i>wase</i> ‘early growth’)	<i>apa</i> ‘millet’
pKJ	* <i>pə</i> ‘a grain’ * <i>jə</i> ‘rice, rice plant’	(* <i>wasə</i> ‘shoot’)	* <i>apa</i> ‘cereal; millet grain’

Although the lexicalized prefix **po* in MK *pyé*, *pap*, and *psól* must have meant ‘rice’ when these compounds were formed in Korean, its Old Japanese cognate *po* is not a word for rice, but rather a word meaning ‘a grain, an ear (of grain et cetera)’. I hypothesize that pKJ **pə* meant only ‘a grain, small grains’ and that its use as a prefix for *‘rice’ in Korean represents a semantic narrowing through lexical recycling (discussed in Section 5).

4. Analysis of OJ *swoba* and MK *cwoh*

In addition to rice and millet, buckwheat is an important pseudo-cereal in Japanese cuisine and culture. A direct comparison of OJ *swoba* ‘buckwheat, *Fagopyrum esculentum*’ with MK *mwomilh* ‘id.’ is out of the question, and the forms are clearly unrelated (MK *mwomilh* is a compound of *mwoy* ‘food, meal’ and *milh* ‘wheat’). However, the initial syllable of OJ *swoba* is phonologically similar to, and constitutes a match with MK *cwoh* ‘millet’. This observation raises the possibility of an etymological relationship of the two forms. MK *cwoh* has no other cognate in Japanese, and the analysis in Section 3 indicates that the Japanese word for ‘millet’ (OJ *apa*) may have originally designated a cereal, as opposed to the plant itself.

I propose that OJ *swoba* ‘buckwheat’ is divisible into pre-pJ **so(N)* + **pa*, the initial syllable **so(N)* being cognate with MK *cwoh* (pK **tsoh* or **tson*). I further propose that pKJ **tson* was a word for the millet plant, *Setaria italica*, and that this word was preserved in Korean as ‘millet’ but became lexicalized in a compound with pJ **pa* ‘leaf’ (OJ *pa* ‘id.’) to mean buckwheat plant (lit. ‘millet-leafed’) in the Japanese lineage.

pKJ *tsoŋ ‘millet plant’ > pre-pJ *soN-pa ‘millet-leaf’ > OJ *swoba* ‘buckwheat’

Why did this compound come to be in pre-proto-Japanese? Unlike millet, which possesses a single large panicle containing all of the harvestable grain, buckwheat grows grain-like seeds that are distributed on flowers throughout the plant, not on a single inflorescence. Thus although both buckwheat and millet are grown for cereal or pseudo-cereal, buckwheat resembles a proto-typical leafy, flowering plant far more than millet does. I hypothesize that the similarity of buckwheat to a prototypical flowering plant explains final *-ba* in the Japanese form, etymologically a suffixation of pJ *pa ‘leaf’ (OJ *pa*). This analysis points to pre-pJ *so or *soN¹⁶ as a nominal with which *pa ‘leaf’ combined to become pJ *soNpa ‘buckwheat’. Although the precise meaning of pre-pJ *so/*soN is unclear, I infer that this nominal must have referred to a cereal grain plant with less leafy characteristics than buckwheat. On the basis of the phonologically unproblematic comparison to MK *cwoh* ‘millet’, I reconstruct pre-pJ *soN, pKJ *tsoŋ as originally ‘millet plant’.

I hypothesize that in proto-Korean-Japanese and pre-proto-Japanese, *tsoŋ was a word for the millet plant, whereas *apa referred to the cereal grain that was harvested from millet and cooked (discussed in Section 3). In the Japanese lineage, *tsoŋ ‘millet plant’ was lost as an independent word, having been replaced by *apa as a word for both the millet plant and millet grain. In the Korean lineage, *tsoŋ ‘millet plant’ was preserved, but *apa ‘(millet) grain’ became lost as an unbound word, surviving only as a lexicalization in *pə-ap ‘rice-(cooked)grain’ where its original reference to ‘millet’ was bleached.

4.1 Agricultural vocabulary: Conclusions

On the basis of the comparison of reconstructed pre-MK *po ‘(prefix for rice words)’ and OJ *po* ‘a grain, a kernel’, I reconstruct pKJ *pə with an original meaning of ‘a grain’. The repurposing of *pə to a prefix for ‘rice’ in pre-Middle Korean can be considered a form of lexical recycling, whereby an earlier, more general word has been recycled to refer to a new, innovative item.

On the hypothesis of separable *po in Korean rice vocabulary, I further reconstruct pKJ *apa ‘(millet) grain/cereal’ (MK *pap*, OJ *apa*) and pKJ *jə ‘rice’ or ‘rice plant’ (MK *pyé*, OJ *yone*). A pKJ form *wasəŋ ‘sprout, shoot(?)’ is also reconstructible on the basis of MK *psól* ‘uncooked rice’ and OJ *wase* ‘early growth’, though the reconstruction is weaker than *pə, *apa, and *jə. I have also proposed that OJ *swoba*

16. Both reconstructions are possible. Reconstructing pre-pJ *soN with a final nasal implies direct compounding of *soN+pa > *soNpa; pre-pJ *so implies a *rendaku* compound of *so + pJ genitive *n(ə) + *pa.

‘buckwheat’ and MK *cwoh* ‘millet’ are relatable, from pKJ *tsoŋ ‘millet plant’. Both Old Japanese and Middle Korean words for buckwheat, OJ *swoba* and MK *mwomilh* (cf. MK *milh* ‘wheat’) appear to be internally derived compounds, so a pKJ word for ‘buckwheat’ is probably unrecoverable.

It is worth noting that the reconstruction of pKJ *jə ‘rice’ or ‘rice plant’ does not hinge on the etymological connection of OJ *yone* to OJ *ine* proposed in Section 2. The existence of OJ *yone* ‘hulled rice’ alone implies pJ *jə ‘rice’ that can be linked to MK *pyé* ‘rice plant’. In fact, OJ *ine* ‘rice plant’ could well be compared to MK *isak* ‘ear of grain,’ pointing to an alternative reconstruction of pKJ *i ‘rice grain’.

5. Discussion

Sections 5.1, 5.2, and 5.3 discuss issues and implications surrounding the Japano-Koreanic cognates proposed in Sections 3 and 4.

5.1 Borrowing of agricultural terminology?

Since it is impossible to positively disprove that lexical similarities are due to borrowing as opposed to being true cognates, we must consider whether the similarities discussed in this chapter might be later importations from Old Korean into pre-Old Japanese, not Japano-Koreanic cognates. The direction of importation, from Korean into Japanese, follows the direction of cultural transfer from Korea to Japan in the mid-first millennium CE. However, critical scrutiny of these proposed cognates makes it evident that borrowing out of Korean is not a satisfactory explanation for the lexical similarities presented in this chapter.

Borrowing from language X into language Y is most plausible when an item present in language X shows signs of being originally absent in language Y, or signs of being novel or otherwise out of place in the vocabulary or culture of language Y. An item that is productive and morphologically simplex in language X but lexicalized and bound in language Y may have been borrowed from language X into language Y. But borrowing of material that is synchronically *unproductive* at the time of borrowing (such as a lexicalized affix) is implausible, much less so for unproductive material whose etymological origins can only be traced by internal reconstruction (Winford 2003: 62).

So are these four Japano-Koreanic cognates likely to have been borrowed from Korean into Japanese? Two of the proposed cognates, *pə ‘a grain,’ *apa ‘cereal,’ have direct reflexes in Japanese but exist only as lexicalizations in Korean. At no period in the known history of Korean are *pə and *apa attested as free nouns. This does not comport with the importation hypothesis; if the Japanese forms were

borrowings from Korean, we expect Korean reflexes to be unbound and unlexicalized as well. Next, we can consider the *yo of OJ *yone*. The reconstructed form for ‘rice, rice plant’ (*jə) is lexicalized in both languages, though more transparently so in Japanese. Given that the form is lexicalized in both languages, there is no reason to suspect borrowing of the form in either language. Only one reconstructed form (*tsoŋ ‘millet plant’) has an unbound reflex in Korean (MK *cwoh*) and a lexicalized reflex in Japanese (OJ *swoba*), which comports with a loanword hypothesis. However, reflexes of Japanese *soba* can be found in both the Northern and Southern Ryukyuan languages (cf. Okinawa *suba*, Yaeyama *suba*). The existence of Northern and Southern Ryukyuan cognates means that ‘buckwheat’ (pJ *soNpa) is reconstructible for proto-Japonic, and cannot be deemed a likely borrowing from Old Korean. There is therefore no reason to think that borrowing explains any of the lexical similarities proposed in this chapter. The cognates proposed here can only be explained by sheer chance, which seems highly improbable, or common inheritance from the same ancestor language.

5.2 The chronology of Proto-Korean-Japanese

Citing the apparent absence of rice-related cognates in the proto-Korean-Japanese reconstruction, Vovin (1998), Unger (2009), and Whitman (2011) each argue that Japanese and Korean could not have diverged before the adoption of rice agriculture. This paper has departed from previous scholarship by presenting evidence of a reconstructed proto-Korean-Japanese word *jə relating to rice. The presence of a word for ‘rice’ or ‘rice plant’ in the proto-Korean-Japanese lexicon suggests that contrary to what most scholars believe, this group did know of the crop. Following the chronology in Whitman (2011), the presence of a word for ‘rice’ could indicate a post-2500 BCE split of Japanese and Korean (following the introduction of dry-field rice), or even a post-1500 BCE split of the languages (following the introduction of wet-field rice). On the other hand, Japanese and Korean do not seem to share any words for the cultivation, harvesting, or cooking of the grain itself, and *jə is the only such <rice> word reconstructible for pKJ. This absence militates against the idea that proto-Korean-Japanese culture cultivated the crop as a staple grain. So with the reconstruction of pKJ *jə ‘rice’ or ‘rice plant,’ there is both evidence for and against the idea that pKJ speakers cultivated rice.

Beyond archaeobotany, an argument against a late, post-1500 BCE split of Japanese and Korean (after wet-field rice) can be found in analyzing the chronology of Japanese and its relationship to the so-called “Koguryoan” language. “Koguryoan” is a language attested sparsely in toponyms recorded in the Korean historical record *Samguk Sagi*, and ostensibly represents a non-Sillan Korean language from before the Korean Three Kingdoms period (ended 668 CE). Unger

(2009) argues convincingly that Japanese-like elements in “Koguryoan” transcriptions really represent “para-Japanese,” the language of peninsular pre-Japanese peoples who never left Korea during the Yayoi Migrations (ca. 800–400 BCE), when pre-Japanese speakers migrated to Japan. This “para-Japanese” probably represents a peninsular sister lineage to the insular Japonic family that began to branch off during the early-to-mid 1st millennium BCE.

Although attestations of this language are extremely scant, the evidence suggests that Koguryoan and pre-Japanese share lexical innovations that already distinguish them from pre-Korean in some important ways. For example, reconstructions of Koguryoan numerals such as ‘3’ (密 ?*mit), ‘5’ (于次 ?*jutsi), ‘7’ (難隱 or 難 *nanV) and ‘10’ (德 ?*tək) look strikingly similar to Old Japanese *mi* ‘3’, *itu* ‘5’, *nana* ‘7’, *towo* ‘10’, and very little like Middle Korean numerals *seyh* ‘3’, *tasos* ‘5’, *nilkwup* ‘7’, *yelh* ‘10’ (Beckwith 2007).¹⁷ Crucially, the fusion of a second syllable *tu in ‘5’ (*i > *itu), the reduplication of *na in ‘7’, and the innovation of a word for ‘10’ from *‘double (hands)’ are key developments in the prehistory of Japanese numerals that already appear evident in Koguryoan numerals. This suggests that Koguryoan is a much closer sister to pre-Japanese than to pre-Korean (for proto-Korean-Japanese numeral etymologies, see Francis-Ratte 2016: 437–452). That is to say, numerals in the Japanese lineage have already undergone several key lexical changes by the time insular and peninsular Japanese begin to diverge in the early first millennium BCE, changes that never took place in the Korean lineage. Significant time must therefore have elapsed between the divergence of proto-Korean-Japanese and the divergence of insular-peninsular Japanese for such changes to have occurred. These observations, along with the general absence of rice agriculture terminology that we expect for a late split, conspire to indicate that a post-1500 BCE split of pre-Japanese and pre-Korean is implausibly early.

Without further evidence, it is difficult to judge when proto-Korean-Japanese may have split. But the presence of a word for ‘rice’ in the proto-Korean-Japanese lexicon may not necessarily indicate the cultivation of *Oryza* by proto-Korean-Japanese-speaking farmers. PKJ *jə could have referred not to *Oryza Sativa* but to ‘wild rice’ such as *Zizania Latifolia*, a.k.a. Manchurian wild rice (Japanese *makomo*, Korean *cwul*). Both *Zizania* and *Oryza* belong to the tribe *Oryzaceae*, and *Zizania* was once cultivated and gathered for its grain as well. *Zizania* bears strong morphological similarities to *Oryza*, but is native to and grows wild in East Asia (Simoons 1991: 165; Guo et al. 2015). It is plausible that pKJ *jə could have been a word for *Zizania*, and was later repurposed to refer to *Oryza* identically in

17. Reconstructions of Koguryoan numerals draw inspiration from Beckwith’s (2007) reconstructions, but should not be interpreted as an endorsement of his claim to have accurately pinpointed the phonetic realizations of these forms.

the two languages after the introduction of *Oryza* based on the close similarity of the two genres. At this point, rather than drawing a definitive conclusion on the basis of one reconstruction, the proto-Korean-Japanese form *jə meaning ‘rice, rice plant’ should be construed as one piece of evidence supporting a split of Korean and Japanese after the cultivation of dry field rice in 2500 BCE.¹⁸ Further linguistic research will likely help elucidate this question.

5.3 Lexical recycling as a general pattern in Korean

As hypothesized in Section 3, MK *pyé* ‘rice plant’, *psól* ‘uncooked rice’, and *pap* ‘cooked rice’ all likely contain a proto-Korean prefix *pə denoting ‘rice’. I have posited that pKJ *pə originally did not mean ‘rice’ but rather ‘a grain’ on the basis of comparing pKJ *pə with OJ *po* ‘an ear, a grain’. I propose that the shift of *pə from ‘grain’ to a prefix meaning ‘rice’ in Korean be considered a type of lexical recycling, whereby the word for a new staple ‘rice’ was expressed not with an entirely new word but by repurposing an older word possessing a similar (but pre-technological) meaning.¹⁹ Based on these agricultural etymologies, we can observe that Japanese generally preserves the pre-rice meaning of proto-Korean-Japanese agricultural terms (e.g. OJ *apa* ‘millet’, *po* ‘ear, grain’), whereas Korean has repurposed much of its pre-rice vocabulary into rice-related compounds (e.g. MK *pap* ‘cooked rice’). Lexical recycling of vocabulary in Korean is precisely the reason behind the absence of straightforward cognates between the two languages.

There is evidence outside of agricultural vocabulary that lexical recycling of pre-technological to post-technological words could be a general pattern in the pre-Korean lexicon. Francis-Ratte (2016: 337) proposes that OJ *isi* ‘rock’ (also *iswo*

18. A reviewer notes a potential problem in a 2500–1500 BCE split, namely the observation that Japanese and Korean appear to be too different in their lexicon and morphology to have diverged between 4500 to 3500 years BP. But many seismic cultural changes impacted Korea and Japan in the millennium preceding historical records, such as the adoption of wet rice agriculture, state organization, the migration of pre-Japanese to Japan, and Sinitification. These changes alone may account for the amount of lexical replacement that is hypothesized to have occurred in the two languages. As for morphological change, Francis-Ratte (2016) shows that there are more morphological correspondences between Japanese and Korean than previously believed. The evidence from lexical comparison of Japanese and Korean may not point to a very early divergence after all.

19. Compare Japanese *kome* ‘hulled rice’, which appears to be a borrowing Old Chinese *(C.)m[e]jɿ as opposed to a recycling of a pre-existing Japanese word (Sagart 2011; Baxter & Sagart 2014).

‘id.’), going back to proto-Japonic **esoj* < **e-soj*, is cognate with MK *swóy* ‘metal’.²⁰ Metallurgy is not thought to have been practiced in the Korean peninsula until the end of the Mumun period circa 400 BCE, meaning that proto-Korean-Japanese speakers could not possibly have possessed this technology. And yet, pJ **e-soj* ‘rock’ and MK *swóy* ‘metal’ (pK **soj*) once again demonstrate a correspondence of a pre-technological word in Japanese (‘rock’) to a post-technological term in Korean (‘metal’). Since reflexes of **esoj* are distributed widely in Japanese and Ryukyuan languages, borrowing of the form can be ruled out, so a cognate relationship of pJ **esoj* to pK **soj* is the most reasonable account. The shift in meaning of pKJ **soj* from ‘rock, ore’ to ‘metal’ in pre-Korean may constitute another example of lexical recycling, whereby words with similar but pre-technological meanings were repurposed into words for new practices and materials. This raises the possibility that lexical recycling of pre-technological to post-technological vocabulary may be a general pattern in the Korean lexicon, one which may help us to uncover further Januo-Koreanic cognates in future research.

6. Conclusion

This analysis has proposed four new Januo-Koreanic etymologies that are explained within a framework of lexical recycling, where pre-rice words have been repurposed into post-rice words in an earlier stage of Korean. This has given rise to correspondences between pre-rice vocabulary in Japanese (OJ *apa* ‘millet’, *po* ‘a grain/ear’, *wase* ‘early growth’) and post-rice vocabulary in Korean (MK *pap* ‘cooked rice’, *pyé* ‘rice plant’, *psól* ‘hulled rice’).

Just as scholars of the Indo-European family speak of a technological “wheel-line” that informs the chronology of the Indo-European language family, it may be useful to conceive of a “rice-line” in the history of East Asian languages, a chronological marker that helps us to understand the diachronic development of agricultural vocabulary. Although the repurposing of older words for new technologies is in itself not uncommon, it is significant that a pattern of lexical recycling can be identified in Korean, for rice agriculture and possibly more generally.

Four Januo-Koreanic etymologies with proto-Korean-Japanese reconstructions might not seem significant, but these etymologies relate Korean and Japanese words that possess disproportionate importance in the lives of prehistoric East

20. Proto-Japonic **esoj* is reconstructed on the basis of the alternation of OJ *isi* with *iswo-* (e.g. *iswonokami* ‘above the rock’) as well as evidence from Ryukyuan languages indicating original **e* with mid-vowel raising. Initial **e-* is almost certainly separable, given OJ *ipa* ‘boulder’ and the comparison to Korean *ye* ‘rocks beneath water’.

Table 8. Lexical recycling of pre-rice vocabulary

Pre-Rice Period	RICE	Post-Rice Period
pKJ *pə ‘a grain’	↓	OJ <i>po</i> ‘a grain, an ear’
pKJ *apa ‘millet (cereal)’	MK <i>p(-)ap</i> ‘cooked rice’	OJ <i>apa</i> ‘millet’
pKJ *ja ‘(dry) rice’	MK <i>p(-)yé</i> ‘rice plant’	OJ <i>yo(-ne)</i> ‘uncooked rice’
pKJ *wasər ‘shoot’	MK <i>p(-)sól</i> ‘uncooked rice’	OJ <i>wase</i> ‘early growth’
pKJ *tsoŋ ‘millet plant’	MK <i>cwoh</i> ‘millet’	OJ <i>swo(-ba)</i> ‘buckwheat’
Pre-Iron Period	Post-Iron Period	
pKJ *soj ‘rock, ore’	MK <i>swóy</i> ‘metal’	OJ (<i>i-</i>)/ <i>si/(i)-</i> ‘rock’

Asian peoples. Furthermore, discovering cognate relationships in words for ‘(millet) cereal,’ ‘millet plant,’ ‘a grain,’ and ‘rice’ shore up a key weakness in the theory of Janapo-Koreanic common origin, namely the absence of shared words for cultivars. The discovery of cognates in the pre-wet rice lexical stratum, plus the hypothesis of a pattern of lexical recycling in Korean, are precisely the kind of evidence that fits a Janapo-Koreanic common origin before the advent of paddy rice farming. That is to say, the etymologies proposed in this analysis go a long way towards creating a plausible narrative explaining how Japanese and Korean descend from a common source language, a narrative that has remained incomplete. It is certain that the advent of wet rice farming in Korea and Japan must have had a great impact on the agricultural vocabulary of these languages, so an analysis of this lexical stratum provides a test case for understanding how words change in the face of revolutionary technological innovations. The fact that historical linguistic methodology can be employed to reconstruct despite such seismic changes is a testament to the power of the Comparative Method.

References

- Baxter, William H. and Lauren Sagart. 2014. Baxter-Sagart Old Chinese Reconstruction, Version 1.1 (20 September 2014). <<http://ocbaxtersagart.lsa.umich.edu/BaxterSagartOCbyMandarinMC2014-09-20.pdf>>, accessed November, 2016.
- Beckwith, Christopher. 2007. *Koguryo, the Language of Japan's Continental Relatives: An Introduction to the Historical-Comparative Study of the Japanese-Koguryoic Languages with a Preliminary Description of Archaic Northeastern Middle Chinese*. Leiden: Brill.
- Francis-Ratte, Alexander. 2016. Proto-Korean-Japanese: A New Reconstruction of the Common Origin of the Japanese and Korean Languages. PhD dissertation, The Ohio State University.

- Guo, Longbiao *et al.* 2015. A host plant genome [*Zizania latifolia*] after a century-long endophyte infection. *The Plant Journal* 83(4): 600–9. doi:10.1111/tpj.12912
- Jeremy, Michael & Robinson, Michael Ernest. 1989. *Ceremony and Symbolism in the Japanese Home*. Manchester: Manchester University Press.
- Lee, Jung Young. 1981. *Korean Shamanistic Rituals*. The Hague: Mouton.
- Martin, Samuel E. 1966. Lexical evidence relating Korean to Japanese. *Language* 42(2): 185–251. doi:10.2307/411687
- Martin, Samuel E. 1987. *The Japanese Language Through Time*. New Haven CT: Yale University Press.
- Nam, Kwangwu. 1997. *Kwo'e sacen (Dictionary of Classical Korean)*. Seoul: Donga Publishing.
- Omodaka, Hisataka *et al.* 1967. *Jidaibetsu kokugo daijiten: Jōdai hen (Great Dictionary of Japanese by period: Ancient period)*. Tōkyō, Japan: Sanseidō.
- Sagart, Laurent. 2011. How many independent rice vocabularies in Asia? *Rice* 4: 121–133. doi:10.1007/s12284-011-9077-8
- Simoons, Fredrick. 1991. *Food in China: A Cultural and Historical Inquiry*. Boca Raton FL: CRC Press.
- Unger, J. Marshall. 2007. Dating the collapse of the *kō-otu* distinctions and the status of /mo/. In *Current Issues in the History and Structure of Japanese*, Bjarke Frellesvig, Masayoshi Shibatani & John Charles Smith (eds), 339–50. Tōkyō: Kuroshio.
- Unger, J. Marshall. 2009. *The Role of Contact in the Origins of the Japanese and Korean Languages*. Honolulu HI: University of Hawai'i Press.
- Unger, J. Marshall. 2014. No rush to judgment: The case against Japanese as an isolate. *NINJAL Project Review* 4(3): 211–230.
- Vovin, Alexander. 1998. Japanese rice agriculture terminology and linguistic affiliation of Yayoi culture. In *Archaeology and Language II*, Roger Blench & Matthew Spriggs (eds), 366–378. London: Routledge. doi:10.4324/9780203202913_chapter_15
- Vovin, Alexander. 2010. *Koreo-Japonica: A Re-Evaluation of a Common Genetic Origin*. Honolulu HI: University of Hawai'i Press.
- Vovin, Alexander. 2015. On the etymology of Middle Korean psar 'rice'. *Türk Dilleri Araştırmaları* 25: 229–238 (Festschrift für Uwe Bläsing).
- Whitman, John. 1985. *The phonological basis for the comparison of Japanese and Korean*. PhD Dissertation, Harvard University.
- Whitman, John. 2011. Northeast Asian linguistic ecology and the advent of rice agriculture in Korea and Japan. *Rice* 4: 149–158. doi:10.1007/s12284-011-9080-0
- Whitman, John. 2012. The relationship of Japanese and Korean. In *The Languages of Japan and Korea*, Nicholas Tranter (ed.), 24–38. London: Routledge.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden MA: Blackwell.

The language of the Transeurasian farmers

Martine Robbeets

Max Planck Institute for the Science of Human History, Jena

The Farming Language Dispersal Hypothesis makes the radical and controversial claim that many of the world's major language families owe their present-day distribution to the adoption of agriculture by their early speakers. Especially for regions such as Northern Asia, where farming is only marginally viable, this claim has been seriously called into question. This paper investigates to what extent agriculture impacted the dispersal of the Transeurasian language family, i.e. the genealogical grouping consisting of the Turkic, Mongolic, Tungusic, Koreanic and Japonic languages. For this purpose, I establish the internal family structure of Transeurasian, reconstruct cultural vocabulary and situate the Transeurasian languages in time and space. Assessing the cultural reconstructions and mapping the tree topology, time-depth and homeland on the demographic transitions visible in the archaeological and genetic record, I find indications that proto-Transeurasian was spoken by people gradually adopting farming and that its dispersal was indeed driven by agriculture.

Keywords: Transeurasian, Farming Language Dispersal Hypothesis, genealogical relatedness, homeland, Neolithic

1. Introduction

In this chapter, I use linguistics as a window on early human and agricultural expansion in North and East Asia. My aim is to investigate to what extent agriculture impacted the ancestral proto-Transeurasian language and its early dispersals. The term “Transeurasian” refers to a large group of geographically adjacent languages, given in Figure 1. They stretch from the Pacific in the East to the Baltic and the Mediterranean in the West and include up to five different linguistic families: Japonic, Koreanic, Tungusic, Mongolic, and Turkic (Johanson & Robbeets 2010: 1–2). I distinguish “Transeurasian” from the more traditional term “Altaic”, which I reserve for the linguistic grouping consisting of Tungusic, Mongolic and Turkic languages only.



Figure 1. The Transeurasian languages (generated with WALS tools)

There is an ongoing controversy about the genealogical relatedness of these languages. In my research so far, I have shown that the majority of Transeurasian etymologies proposed in support of inheritance are indeed questionable. However, rather than proposing a wholesale rejection of Transeurasian, I have argued that there is nonetheless a core of reliable etymologies that enables us to classify Transeurasian as a valid genealogical grouping. The evidence (Robbeets 2005, 2015) consists of an inventory of regular consonant and vowel correspondences, common lexical etymologies including basic vocabulary and shared verb morphology.

New questions emerge from the assumption that proto-Transeurasian was an actual spoken language ancestral to the Japonic, Koreanic, Tungusic, Mongolic, and Turkic languages. What populations corresponded to the speakers of proto-Transeurasian? Where and when did these people originally live? When did the language family separate into its main branches? What triggered the expansion of the daughter languages? In which directions did the dispersals go? And, when, how and why did the daughter languages move to their present locations? In this chapter, I will argue that the speakers of proto-Transeurasian were familiar with millet cultivation and gradually developed farming during the Neolithic in the West Liao River region of Northeast China. I will suggest that the eastward linguistic expansions of the Transeurasian languages were initially driven by the spread of agriculture.

For some linguists, researching agricultural expansions in Northern Asia sounds as promising as looking for plants on Mars. With regard to “Altaic”, Heggarty and Beresford-Jones (2014:4) for instance, argue that “Northern Asia is home to environments where farming is either not viable at all or only marginally so ... In

this sense, these regions fall by definition outside the scope of the farming/language dispersal hypothesis”. In this chapter, I will show that it is a misconception to assume that certain subsistence patterns, such as nomadic pastoralism or hunting-gathering, have always prevailed in the Transeurasian region. I will argue that the family structure, homeland, time-depth and vocabulary of proto-Transeurasian leave room for a hypothesis that correlates the origin and spread of the language family with the Neolithic transition to farming in Southern Manchuria.

To this end, I will apply the different methods and principles for determining the time, location and cause of linguistic dispersals discussed in the introduction of this volume to the case of the Transeurasian languages. The following section searches for a plausible homeland for the Transeurasian family, using the diversity hotspot principle. Section 3 proposes a tree topology and a time estimate for the nodes in the Transeurasian family on the basis of Bayesian phylogenetic inference. Section 4 maps the tree topology, homeland and time-depth on demographic transitions in the Southern Manchurian Neolithic. Section 5 reconstructs cultural vocabulary for proto-Transeurasian. By way of conclusion, Section 6 summarizes the main arguments for identifying the speakers of proto-Transeurasian with the first farmers in the region and for associating the spread of their language with farming dispersals.

2. The diversity hotspot principle

A loose principle that can help us in locating the original homeland of a language family is the “diversity hotspot principle”. It 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.

From Chinese historical records such as the *Shiji* ‘Records of the Grand Historian’ (109–91 BC), the *Sanguoji* ‘Records of the Three States’ (284 AD) and the *Houhanshu* ‘History of the Later Han’ (5th century AD), we can infer that the Turkic, Mongolic, Tungusic, Koreanic and Japonic languages have all spread to their present-day locations from an area comprising Korea, southern Manchuria and Inner Mongolia. Therefore, even critics of the affiliation of the Transeurasian languages, such as Janhunen (1996) situate the original speech communities of the individual families in the compact area represented in Figure 2. Although the contemporary focus of diversity may diverge, the greatest linguistic diversity in recorded history, and therefore perhaps the location of the Transeurasian homeland, is in the West Liao River region in southern Manchuria.

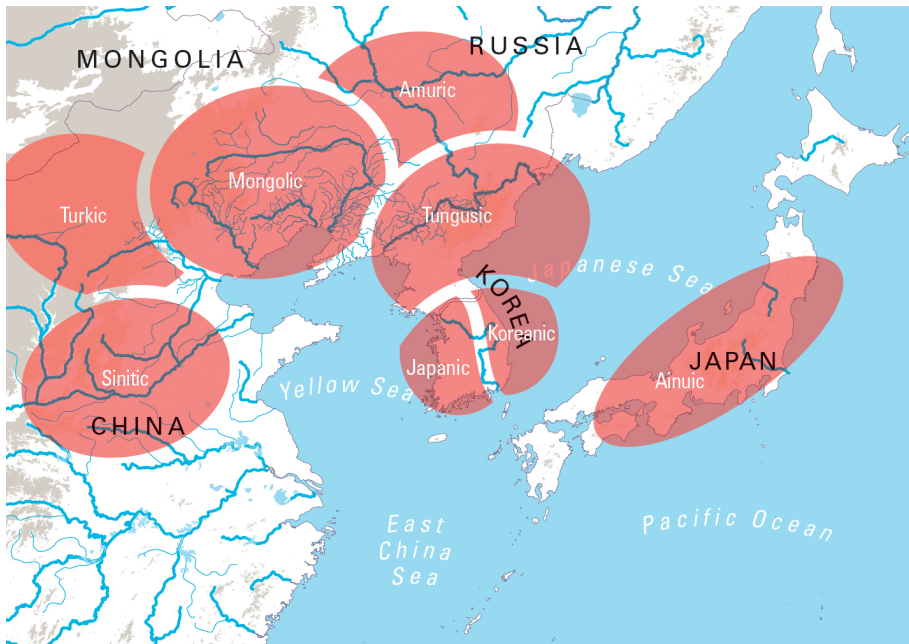


Figure 2. The ethnic groups of prehistorical Manchuria in the last millennium BC according to Janhunen (1996: 216): 1. Sinitic; 2. Turkic; 3. Mongolic; 4. Amuric; 5. Tungusic; 6. Koreanic; 7. Japonic; 8. Ainuic

3. Bayesian phylolinguistics

Although Bayesian phylogenetic inference cannot establish genealogical relatedness between a set of languages, it can be useful to double-check the internal structure of a language family reached by applying classical historical linguistics. Additionally, Bayesian inference can provide us with absolute dates for the nodes in the family and give us an idea of the robustness of our inferences. In a forthcoming study with Remco Bouckaert (Robbeets & Bouckaert forthcoming), we performed a preliminary Bayesian phylolinguistic analysis on the Transeurasian etymologies represented in the Leipzig-Jakarta basic vocabulary list (Tadmor et al. 2010). We used an alternative coding principle, whereby we started from a reconstructed proto-Transeurasian basic item and coded 1 for the presence of a cognate in a daughter language and 0 for the absence of a cognate, irrespective of whether the meanings were identical or not. Taking into account time calibrations for 4 lower nodes in

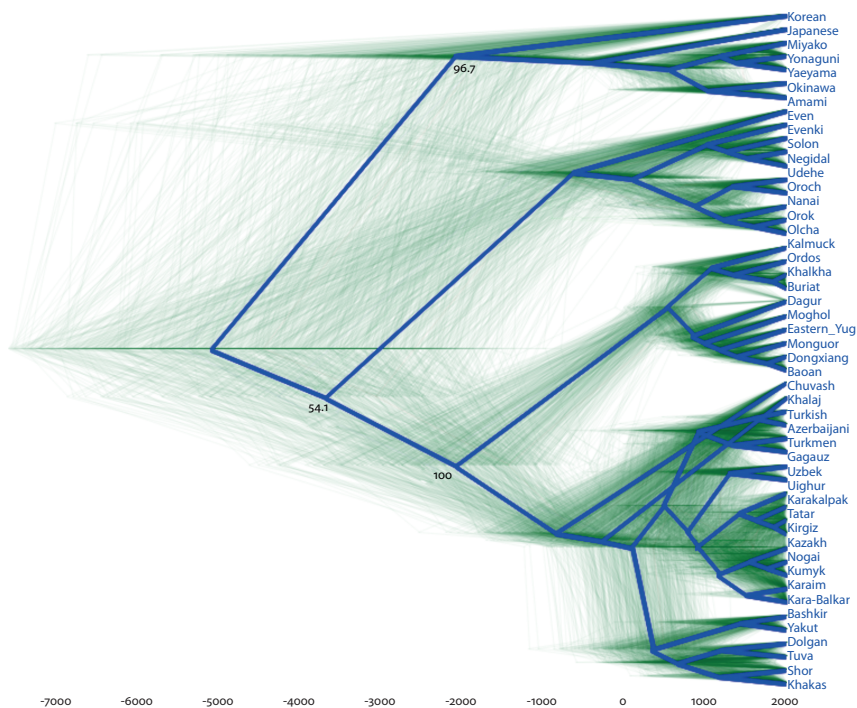


Figure 3. Densi Tree of the Transeurasian family (Robbeets & Bouckaert forthcoming)

the Transeurasian family we ran a Bayesian algorithm on the data. The preliminary result is captured in the Densi Tree, given in Figure 3.¹

In addition to proposing an internal structure for the Transeurasian family, the Bayesian analysis also provides us with estimates for the absolute time depth

1. The Bayesian tree confirms the classification proposed in Robbeets (2015) on the basis of the classical comparative method, except for the position of Tungusic vis-à-vis the other branches. I previously classified it in a unity with Turkic and Mongolic whereby Turkic – rather than Tungusic – branched off first. In contrast to the Bayesian method, which seeks a tree that explains the observed data by quantifying how likely it is that they have been produced by a certain evolutionary process, the classical method is a parsimony method, which seeks a tree that explains the dataset by minimizing the number of changes required to produce the observed state. Thus, the classical comparative method is based on shared innovations: it prefers trees that place innovations where they create the greatest amount of diversity. In the case of Transeurasian the innovations can be phonological (e.g., the loss of voicing distinction in Japanese and Korean, maintenance in Altaic, but loss of certain word-initial voice distinctions in Turkic), syntactic (e.g., the change from 2-way to 3-way distinction in Japanese and Korean demonstratives) or morphological (e.g., the original Transeurasian negative pTEA **ana-* is replaced by **ə-* in Altaic and again by **-mA-* in Turkic).

of the root and the primary nodes in our tree. The estimates are given in Table 1, along with their credible intervals. The observation that the time depth of the root coincides with the start of millet cultivation in Northeast China's West Liao River Region is striking, to say the least.

Table 1. Bayesian time estimates for the primary splits in the Transeurasian family

Node	Time depth	95% HPD credible interval
proto-Transeurasian	5700 BC	6800–4200 BC
proto-Altaic	4600 BC	6100–2800 BC
proto-Japono-Koreanic	3300 BC	5500–1300 BC
proto-Mongolo-Turkic	2800 BC	4800–800 BC

4. Linking demographic pulses to language dispersals

Recently, the archaeobotanists Stevens and Fuller (forthcoming) identified the following three phases in the development of agriculture in Southern Manchuria: (1) the establishment of millet agriculture (6500–4500 BC); (2) the eastward spread of millet agriculture (4500–3000 BC) and (3) the integration and spread of rice and millet agriculture after 3000 BC. It is inviting to map these three phases in the development of agriculture with linguistic stages in the Transeurasian family tree.

4.1 The establishment of millet agriculture

Millet cultivation began around 6200 BC in the Xinglongwa culture (6200–5400 BC), one of the earliest farming cultures in northeast China. There is early evidence for the cultivation of millets, notably large quantities of broomcorn millet (*Panicum miliaceum*) and small amounts of foxtail millet (*Setaria Italica*) (Zhao 2011:301). There is a continuity of cultivation tradition with the ensuing Zhaobaogu (5400–4500 BC) and Hongshan cultures (4500–2900 BC). In contrast to the millet-focused subsistence in the Yellow River Region, the Xinglongwa people in the West Liao River Region subsisted on a broad-spectrum strategy, using various wild and cultivated plants, including roots, beans, and nuts (Shelach 2000; Hunt et al. 2008; Weber & Fuller 2008; Zhao 2011; Liu et al. 2012; Liu et al. 2016). The small size of the recovered millet grains indicates that cultivation was still in a pre-domestication stage. It took almost two millennia for millet to become fully domesticated. The environmental conditions in the West Liao River region are extremely vulnerable to climatic changes. The strengthening of monsoon around 6200 BC increased precipitation and contracted dunefields, facilitating cultivation and leading to the

expansion of early Neolithic cultures such as Xinglongwa and Zhaobaogou (Jia et al. 2017). In my hypothesis, the people depending on broad-spectrum subsistence spoke proto-Transeurasian and the first-order linguistic split between Altaic and Japonic took place towards the end of the domestication process. Figure 4 shows the location of the Xinglongwa culture and thus the presumed homeland of proto-Transeurasian.



Figure 4. The Xinglongwa culture and the establishment of millet agriculture

4.2 The eastward spread of millet agriculture

By the time of the so-called Hongshan culture (4500–2900 BC), millet agriculture diffused eastwards, first, to the Liaodong peninsula and later to the Russian Far East. Kuzmin (2013:8) places the appearance of millet cultivation in the Primorye around 2700 BC in the context of the early Zaisanovka cultural complex (4800–1500 BC), but evidence for agriculture is lacking for the adjacent Boisman culture (4825–2470 BC). In the forest steppe area of the southern Primorye, natural conditions such as open spaces and a drier climate were more favourable for millet cultivation than in the inhospitable forested areas of the north. As the Hongshan population levels were too low to have created resource scarcity (Peterson & Drennan 2011: 106; Drennan & Dai 2017: 464), the spread of millet was not driven

by a population boost, but rather by climate change. Around 2800 BC a weakening of the monsoon and reduction in precipitation led to a major demographic decline and the collapse of the Hongshan culture (Jia et al. 2017). This climate change also affected the maritime-adapted cultural complexes of the Primorye's coast, through cooling, landscape changes and falls in sea level, which disrupted the traditional subsistence base of local hunters and fishermen (Vostretsov 2006). The region between the Liao River and the southern Primorye of the Russian Far East had been in a state of active contact, exchanging obsidian, since before the Neolithic. Therefore, the Hongshan populations could easily spread their millet agriculture and impressed pottery once the climate change called for a shift in subsistence regime.

Wang et al. (2016) have recently established genetic continuity between ancient DNA from 7 individuals from the Neolithic Boisman culture and speakers of most contemporary Tungusic languages. They find that contemporary Ainu and Nivkh speakers reflect the original Boisman genome but contemporary Tungusic speakers reflect Boisman genes that have been admixed with an additional component. This may indicate that the genetics of modern Tungusic speakers reflect the past admixture of local Nivkh genes with the genes of incoming Transeurasian farmers.

Therefore, as illustrated in Figure 5, I suggest identifying the Hongshan people with the speakers of Altaic, the outlying Hongshan culture on the Liangdong Peninsula with the Japano-Koreanic language and the people who adopted millet-agriculture in the Russian Far East with Tungusic speakers. From the Liaodong peninsula, millets were spread overland to the Korean peninsula in the fourth millennium BC (Ahn 2010; Ahn, Kim & Hwang 2015: 2; Crawford & Lee 2003: 2; Lee 2011). It is conceivable that the people who introduced millet agriculture to Korea were the speakers of proto-Koreanic. The split between proto-Japonic and proto-Koreanic thus occurred on the Liaodong Peninsula and not on the Korean Peninsula. The early date of the Japano-Koreanic split (3300 BC) in the Bayesian estimation above is consistent with the date of the importation of millet agriculture in Korea (ca. 3500 BC).



Figure 5. The Hongshan culture and the eastward spread of millet agriculture

4.3 The integration and spread of rice and millet agriculture

After 3000 BC, rice was added to the agricultural package in the Liaodong – Shandong interaction zone. According to Kim (2003), the millet cultivators on the Korean peninsula had returned to nomadic hunting-gathering by the second millennium BC, perhaps due to another wave of climatic cooling. Archaeobotanical studies such as Bale (2001), Miyamoto (2009) and Ahn (2010) show that wet-rice agriculture came to the Korean peninsula in the late second millennium BC (1300–1000 BC) via the Shandong and Liaodong peninsulas. The second transition from foraging to farming on the Korean peninsula involved not only a cultural shift, but most probably also a linguistic one: the people who brought wet-rice agriculture to Korea may have spoken proto-Japonic. In the first millennium BC the rice and millet farmers arrived via the Korean Peninsula in Japan, where they established the Yayoi culture (900BC–300AD) (Crawford & Shen 1998; Crawford & Lee 2003).

The archaeological evidence is supported by Kanzawa-Kiriyama’s (2016) study using nuclear genome sequencing of two Jomon (14,000–900 BC) individuals. They confirm the mainstream “dual structure model”, originally proposed by Hanihara (1991) and recently supported by Jinam et al. (2012) and Jeong et al. (2016), describing the Mainland Japanese population as an admixture of native Jomon genes and

incoming Yayoi genes from farmers coming from the Korean peninsula. I associate the spread of integrated rice and millet agriculture through Korea to Japan with the spread of the Japonic language. This is illustrated in Figure 6.



Figure 6. The Yayoi culture and the integration of rice and millet agriculture

4.4 Demography mapped on linguistic phylogeny

Mapping the above demographic processes on the Transeurasian tree, we find the correlations visualized in Figure 7. Proto-Transeurasian is associated with a gradual development of millet cultivation, the first-order split in the family with the full domestication of millet, the separation of Koreanic and Tungusic with the eastward spread of millet, and proto-Japonic is associated with later migrations driven by integrated rice and millet agriculture.

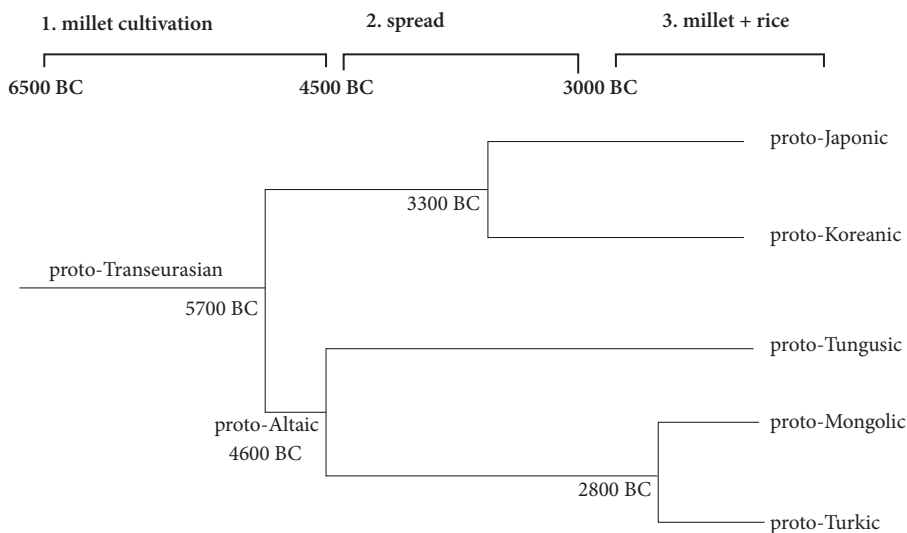


Figure 7. Mapping the agricultural development in Northeast Asia on the language tree of Transeurasian

5. Cultural reconstruction

Cultural reconstruction enables us to study human prehistory by correlating our linguistic reconstructions with information from archaeology about the cultural and natural environment in which the speakers of the proto-language likely operated. This method is also known as “Linguistic paleontology”, “Wörter und Sachen” or “Linguistic archaeology”. Reconstructed vocabulary associates proto-Transeurasian with broad-spectrum subsistence including millet cultivation. In addition to evidence for cultivated fields, seed and consumable plants such as a millet-like crop, nuts and roots, I reconstruct subsistence activities such as “sowing”, “grinding”, “kneading”, “weaving”, “sewing”, “making rope” and indirect evidence for pottery production. Interestingly, proto-Transeurasian lacks maritime vocabulary and terms for rice agriculture, while Japano-Koreanic reflects coastal subsistence terms but still lacks rice vocabulary (Robbeets 2017).² Therefore, cultural reconstruction

2. Francis-Ratte (this volume) reconstructs pJK **ɣə* ‘(dry) rice’, suggesting that Japanese and Korean may have diverged at a time when field rice was already being cultivated in Northeast Asia while paddy rice was not introduced yet. However, there is only a single cognate set relating to rice and it is rather dubious as the participating cognates are based on a morphological segmentation of MK (*p-*)*yé* ‘rice plant, kernel of rice (unhusked)’ and OJ *yo(-ne)* ‘uncooked rice’.

indicates that the time-depth of both proto-Transeurasian and proto-Japano-Koreanic preceded the integration of rice agriculture starting around 3000 BC. Additionally, it indicates an inland location for proto-Transeurasian in contrast to the homeland of proto-Japano-Koreanic, which seems situated on the coast.

5.1 Economic plants and cultivation

(1) pTEA **pata* ‘field for cultivation’

- a. Turkic: pTk **(p)ati* ~ **(p)ata* ‘field irrigated for cultivation’ (pTk **-z* collective suffix, pTk **(A)g* place suffix?)

OT (Karakhanid) *atiz* ‘any strip of land between two dikes’, MTk. *atizla-* ‘to create an irrigation canal in a field’, Uig. *etiz* ‘watered field, boundary’, Tkm. *atiz* ‘watered field, boundary’, Shor *adis* ‘a measure for fields, 1/18 dessiatin (= ca. 607 square meters)’, Kirg. *adir* ‘hilly terrain’, Kaz. *atiz* ‘a plot of land, watered by irrigation canals and properly limited’; MTk. *atov* ‘1 island’, Tk. *ada* ‘1’, Tat. *ataw* ‘1’, Tkm. *a:da* ‘1’, Chu. *odā* ‘1’

- b. Koreanic: pK **patA* ‘(dry) field’ (pK **(i/Λ)k* place suffix)

K *path*, MK *path* ‘(dry) field, farm, patch, garden, position on a game board’

- c. Japonic: pJ **pata* ‘(dry) field’ (pJ **-ka* place suffix, pJ **-i* substantivizer)

J *hata* 2.4, OJ *pata* ‘(dry) field’ J *hatake* (3.7a~b), OJ *patake*₂ ‘field, farm, plantation, garden’, Shuri (Okinawa) *hataki*, Naze (Amami) *hataə*, Ishigaki (Yaeyama) *patagi*, Oura (Miyako) *patagi*, Yonaguni *hatagi*, pR **patake* ‘field, croft’

The Turkic word pTk **(p)ati* ~ **(p)ata* ‘irrigated field for cultivation’ can be reconstructed, considering pTk **(p)ati-z* ‘watered fields’ and pTk **(p)ata-g* ‘island’ as reflexes of the same etymon, whereby pTk **-z* represents a dual and collective suffix (e.g., in paired body parts such as OT *kö-z* ‘eyes’, *ti-z* ‘knees’, *agi-z* ‘lips’ and *kökü-z* ‘breasts’, ethnonyms such as OT *ogu-z* and *kirgi-z*, sets of more than one such as *iki-z* ‘twins’, *üc-üz* ‘triple’, *dörd-üz* ‘quadruple’ and undefined quantities such as OT *yultu-z* ‘stars’, *yüldi-z* ‘roots’) and pTk **(A)g* a petrified place suffix (e.g., pTk **o:t* ‘fire’ → *o:t-ag* ‘tent, dwelling place’). The alleged loss of the initial labial stop **p* cannot be confirmed since we lack a Khalaj cognate. The reconstruction of the final low vowel in pTk **(p)ata* is supported by the vowel in the Mongolic borrowing pMo **atar* ‘uncultivated land’. Contrary to Ramstedt (1949: 192–293), Poppe (1960: 51, 82), Menges (1984: 284), Starostin et al. (2003: 1127) and Savelyev (this volume), I do not think that the Mongolic form reflected in WMo *atar* ‘unploughed or fallow field’, Khal. *atar*, Bur. *atar* and Mgr. *atər* is a cognate. Indications of borrowing are the lack of initial *f*- in the Monguor form *atər*, which would be the expected reflex

of pMo **p*- (e.g. pMo **poro*- ‘to entwine’ in (8)) and the fact that the Mongolic form is unsegmentable in spite of the morphological complexity of the Turkic form. In Korean, non-rising low monosyllabic place nouns ending in *-k* or *-h* commonly are reductions from disyllabic forms with a place suffix **(i/ʌ)k* in the second syllable (Martin 1996: 44–45), e.g., MK *pask* ‘outside’ (< **paʃʌ-k*), *math* ‘yard’ (< **maʃʌ-k*), *alph* ‘front’ (< **alpʌ-k*), etc. The lack of aspiration in the derivation K *patwuk* ‘stone checkers (game)’ (< **pat tolk* ‘field stones’) may be indicative of the word for ‘field’ without place suffix. In Japonic, pJ **pata-ka-i* ‘field, plantation’ is probably derived from pJ **pata* ‘(dry) field’ by means of the place suffix pJ **-ka*, which occurs also in *oka* ‘hill’, *arika* ‘whereabouts’, *sumika* ‘residence’ etc. The sharing of a corresponding place suffix on the word for ‘field’ in Turkic, Koreanic and Japonic may indicate that the derivation goes back to proto-Transeurasian.

- (2) pTEA **pusu*- ‘to sprinkle with the hands’ ~ **pisi*- ‘sprinkle with the hands, sow’
 → **pisi* ‘what is sown’ > **pisi* ‘seed, seedling’ (pTEA **-i* deverbal noun suffix)
 → **pisi-ke* ‘major crop’ (pTEA **kA* plant suffix)

- a. Mongolic: pMo **hüsü*- ~ **hisü*-/*hesü*- ‘to sprinkle, throw out, jump around’
 → **hisi*/**hesi* ‘origin or base of a plant, shoot’ (pMo **i* deverbal noun suffix)

pMo **hüsü-r*- ~ **hesü-r*- / **hisü-r*- ‘to sprinkle, scatter; jump around’ (pMo **r*- intensive)

Middle Mongolian *üsür*- ‘1 to spout, squirt out (of water); 2 to jump, leap (intr.)’, Written Mongolian *üsür*- ‘1, 2’, Khalkha *üsre*- ‘to squirt; to jump, leap, skip’, Buriat *hür*- ‘to jump, leap’, Ordos *üsür*- ‘to jump, leap’, Kalmuck *ösr*- ‘to sprinkle (water), throw out sparks (of fire); jump or hop (of insects), to fly in the air’ (Ramstedt 1935: 301), Dagur *xesere*- ‘to jump’ (Martin 1961: 161), *xəsur*-, *xesurə*- ‘to sprinkle’, Eastern Yugur *husur*- ‘to jump’, Dongxian *usuru*- ‘to flow’, Monguor *fizuru*-, *fuzuru*- ‘to sprinkle, pour, cast (metal)’, Moghol *üsürü*- ‘to jump, leap’ (Ramstedt 1906)

pMo **hisi* / **hesi* ‘origin or base of a plant’

Middle Mongolian *nisi*, *hesi*, Written Mongolian *isi* ~ *esi* ‘1 foundation, basis, origin, source; 2 a stalk of grain, trunk of a tree, stem of a plant, shoot; 3 handle, grip’, Khalkha *iš* ~ *eš* ‘1 source, basis; 2 stem, stalk, trunk, underground stem; 3 handle, shaft’ (Bawden 1997), Buriat *eše* ‘1, 2, 3’, Kalmuck *iš* ‘1 beginning, source; 2 stalk (of plant), stem (of tree), 3 handle, grip’ (Ramstedt 1935: 210), Ordos *eši* ~ *iši* ‘1, 2, 3’, Baoan *ješi*, *heši* ‘handle, grip’, Dagur *xeš*, *xeši*, *heši* ‘handle, grip, knob’ (Martin 1961: 161), Eastern Yugur *šə* ‘handle, stem’, Kangjia *heši* ‘handle, grip’ (Nugteren 2011: 354)

- b. Tungusic: pTg **pusu*- ‘to spread’ ~ **pisi*- ‘to sprinkle with the hands’ / **pise*- ‘to spread out’ → **pise* ‘offspring’ (through pTg **i* deverbal noun suffix?)
 → **pisi-ke* ‘broomcorn millet’ (pTg **kA* plant suffix)

pTg **pusu-* ‘to sprinkle, to scatter’ ~ **pisi-* ‘to sprinkle with the hands’ / **pise-* ‘to extend out’

Manchu *fusu-* ‘to sprinkle (water), spew, spirt, squirt,’ *fuse-* ‘to propagate, to reproduce, to breed,’ *fisi-* ‘to sprinkle with the hands, to shake, to toss (one’s sleeves),’ *fise-* ‘to project, to jut out, to fork, to branch’ (Norman 2013), Sibe *fusu-* ‘to sprinkle,’ Even *hus-* ‘to sprinkle (with water), splash, sputter, disperse,’ Negidal *xusi-* ‘to sprinkle,’ Olcha *pisuri-* ‘to sprinkle,’ Orok *pisitči-*, *possoli-* ‘to sprinkle,’ Nanai *pisi-*, *fisi-*, *fuksu-* ‘to sprinkle’ (Cincius 1975–1977: 39, 42, 355)

pTg **pise* ‘offspring’

Manchu *fisen* ‘relation, offspring, progeny’ (Norman 2013), Okhotka dialect of Even *hesen* ‘seed, offspring, kin’ (Starostin et al. 2003)

pTg **pisi-ke* ‘broomcorn millet’

Manchu *fisihe* ~ *fisike* ‘glutinous millet, broomcorn millet (*Panicum milia-ceum*),’ *fisitun* ‘a ritual vessel for offering millet; bowl for grinding millet, carved out from a piece of wood’ (< *fisi* + *tetun* ‘utensil’) (Norman 2013), Olcha *pikse* ‘millet,’ Nanai *pikse* ‘millet,’ Kur-Urmi dialect *fixse* ‘millet’

- c. Koreanic: pK **pusu-* ‘sprinkle, scatter, wash, smash’ ~ pK **pisi-* ‘sprinkle, scatter, sow’ → **pisi* ‘what is sown’ (pK *-i deverbial noun suffix) > pK **psi* ‘seed, lineage’

→ *pisi-k* ‘major crop’ (pK *-k plant suffix) > **pski-* > **phi* ‘barnyard millet’

pK **pusu-* ‘to sprinkle, scatter, sow’ ~ **pisi-* ‘to sprinkle, scatter, sow’

K *pu:s-* ‘1 to pour, 2 to sow (tr.);’ K *pu:s-* ~ K *puswu-* ‘to smash, scatter, break,’ MK *poso-* ‘break, shatter,’ K *pusi-* ‘to wash, clean, rinse,’ MK *puswoy-* ‘to wash, clean, rinse (tr.);’ K *pusule tuli-* ‘to smash, to shatter into splinters (tr.);’ K *pusule ci-* ‘to crumble (intr.)’ (K *le tuli-/le ci-* causativity polarizer < pK *($\Lambda/\text{ɪ}$)- anticausative), K *pusul pusul* ~ *posul posul* ‘gently raining,’ K *pusik ha-* ‘to plant, extend’ (MK *-i-* transitivizer < pK *-i- causative); K *ppu:li-* ‘1 to sprinkle, rain slightly (intr.); 2 to sprinkle, shower, water (tr.); 3 to scatter, sow,’ K *ppuli* ‘a root (of a plant),’ MK *spu:li-* ‘to sprinkle’ (MK (*u*)*li-* transitivizer < pK *(*u*)- anticausative + **i-* causative), MK *spih-* ‘to sprinkle; slander,’ K *p:al-* ‘to wash, launder, wash out (tr.);’ MK *·spol-* ‘to wash (tr.)’ (pK *($\Lambda/\text{ɪ}$)- pluractional), MK *·spum-* ‘sprinkle, spout, spurt’ (pK **mi-* ~ *m Λ -* inclinational)

pK **psi* ‘seed, lineage’

MK *·psi*, K *ssi* ‘1 seed, kernel, 2 lineage, descent, breed,’ K *pye-psi* ‘rice seed’

pK **phi* ‘barnyard millet’

MK *·phi*, K *phi* ‘(Japanese) barnyard millet (*Echinochloa esculenta*)’

- d. Japonic: pJ **piyai* ~ **piyia* ~ **piye* ‘barnyard millet’
 J *hie*, OJ *pi₁ye* ‘(Japanese) barnyard millet (*Echinochloa esculenta*)’

In the Mongolic verbs, the semantic shift from ‘to sprinkle’ to ‘to jump’ can be explained by observing the semantics of the Kalmuck verb *ösr-* ‘to sprinkle (water), throw out sparks (of fire); jump or hop (of insects), to fly in the air,’ in which the common denominator is ‘to scatter of a set of small items.’ The deverbal noun of this verb has the primary meaning ‘what is scattered, sown.’ The semantic development in the nouns extends from ‘what is sown’ from ‘origin or base of a plant’ to any ‘origin, base’ and specializes from ‘origin, base of a plant’ to ‘stem of tree’ to ‘handle, grip.’

Given the lexicalization of a deverbal intensive suffix pMo **r-* in a number of Mongolic verb stems (e.g., WMo. *ayimu-* ‘to become confused, mixed up, go astray, be unintelligible (intr.)’ → *ayimur-* ‘to change for the worse, indulge in lustful pursuits, be seduced, be heavily confused (intr.)’, *ciki-* ‘to jam, stuff, press, push; stuff oneself, overeat (tr./intr.)’ → *cikir-* ‘to be unable to pass through or fit in, get stuck,’ *sibqa-* ‘to scrape out, scoop out, empty out (tr.)’ → *sibqar-* ‘to squeeze out, pour out to the last drop, empty out (tr.)’ and *jaki-* ‘to give instructions, to entrust, to give an order for, to ask to run an errand (tr.)’ → *jakir-* ‘to rule, govern, direct, subordinate, subject (tr.)’), we can reconstruct the bare root pMo **hüsü-* ~ **hisü-/hesü-* ‘to sprinkle, throw out, jump around.’ The noun **hisi/*hesi* ‘origin or base of a plant, shoot’ can be derived from the root **hesü-/*hisü-* by suffixation of the deverbal noun suffix pMo **i*, e.g., in WMo. *sönü-* ‘to be extinguished, go out (of fire), cease to be’ → *söni* ‘night, at night’ (Robbeets 2015: 462–463).

Monguor *fɨzuru-* ‘to sprinkle, pour, cast (metal)’ preserves a reflex of the high front vowel in pMo **hisür-*. The reconstruction of initial pMo **h-* is supported by the Buriat, Dagur, Eastern and Monguor verbs and by the Dagur, Kangjia and Baoan nouns. The antiquity of initial **h-* and its origin in pre-pMo **p-* is further supported by the borrowing of the term as pTg **pesin* ‘handle’ (in Manchu *fesin*, Sibe *fesən*, Evenki *hesin*, Even *hesin*, Negidal *xesin*, Olcha *pesi(n)*, Orok *pesi(n)*, Nanai *pesī*, Oroch *xesi(n)* and Udehe *xehi*). The observation that the Tungusic meaning is limited to ‘handle,’ which is secondary in Mongolic, is indicative of borrowing.

The Tungusic verbs reflect the meaning ‘to sprinkle, to scatter.’ The meaning ‘to sow’ is not attested, but the polysemy is observed in other Tungusic verbs, e.g., Sibe *swata-* ‘to sprinkle, sow’ (Kim et al. 2008: 150). The noun pTg **pisi* ‘what is scattered, what is sown’ can be derived from the verb **pisi-* ‘to sprinkle with the hands’ by suffixation of the deverbal noun suffix pTg **i*, reflected, for instance, in Even *tet-* ‘to dress oneself’ → *teti*: ‘garment, uniform’ and Evk. *usi-* ‘to bind’ → *usi*: ‘rope, belt’ (Robbeets 2015: 461–462). Although I cannot explain the final vowel in pTg **pise* ‘offspring,’ I think it concerns a nominalization of the same verb. The

semantic development probably went over ‘seed’ in a similar way as the polysemy in K *ssi* ‘1 seed, 2 lineage, descent,’ as discussed below. Starostin et al. (2003) gloss the word *hesen* from the Okhotka dialect of Even as ‘seed, offspring, kin,’ but I have not been able to trace that form back. Since the final nasal in Okhotka Even *hesen* and Manchu *fisen* is instable and frequently drops when inflectional suffixes are attached, I do not consider it part of the root.

The morphological complexity of Manchu *fisitun* ‘millet bowl’ suggests that pTg **pisi-ke* ‘broomcorn millet’ includes a petrified derivational suffix of the shape pTg **-kA*, found in the names of animals and plants, e.g., in pTg **tasa-ka* ‘tiger’ (e.g., Ma. *tasxa*, Jurchen *tasxa*, Solon *tasax*), pTg **kumi-ke* ‘louse’ (e.g., Evk./Even/Neg. *kumke* and Evk. *kumikēn* ‘insect’, Na. *kunke*, Ud. *kumuge*, Solon *xunḱē* and *xumīxe* ‘ant’), pTg **inū-ke* ‘dog, wolf’ (e.g., Evk. *nēkē* ‘sable’, Even *ḡōke* ‘male (of dog, wolf, fox)’, Sibe *juxə* ‘wolf’, Ma. *noxə* ‘wolf’, *nuxere* ‘puppy’) pTg **eb-ke* ‘heather’ (e.g., Evk. *ebkemkirē*, Neg. *epkexin*, Orok/Oroch *ewxexi*, Na. *opokta* ‘hawthorn’) and pTg **bolo-ka* ‘spiraea’ (Evk. *boloko*, Neg. *boloxokto*, Na. *boloqto*, Ud. *bolokto*).

In Korean we find two sets of reflexes: one set reflecting **u-* vocalism and, therefore, resisting vowel loss, and another set reflecting **i-* vocalism and, therefore, subject to vowel loss and subsequent initial *sp-* clustering in Middle Korean and *pp-* reinforcement in contemporary Korean. In line with Ramsey (1993: 438; 1997), I assume that Middle Korean verb stems with complex initials that are tonic and monosyllabic and have minimal vowels (MK *o, u, i*) are created through the loss of a first-syllable vowel. This internal analysis justifies the reconstruction of the first high front vowel in **pisi-* ‘to sprinkle, scatter, sow’ on the basis of MK *spu-li-* ‘to sprinkle’, MK *spih-* ‘to sprinkle; slander’, MK *·spol-* ‘to wash (tr.)’ and MK *·spum-* ‘sprinkle, spout, spurt’.

Korean has a number of defective converbs, recognizable by the converb ending *e/a* and preceded by an element (*u*)*l-*. They occur with the auxiliary verbs *ci-* ‘to become,’ which polarizes their intransitivity, and *ttuli-* ‘to make,’ which makes them transitive: e.g., K *wuk-* ‘to turn’ → *wukule ci-* ‘to curl up (intr.)’, *wukule ttuli-* ‘to make a dent in (tr.)’. The transitive analytic construction in (*u*)*l-e ttuli-* replaces an older and almost obsolete suffix in (*u*)*li-* that likewise adds transitive meaning and goes back to a synthetic form *l-i-*, where *i-* reflects the causative pK **i-*, e.g., K *wuk-* ‘to turn’ → *wukuli-* ‘to crouch, crush (tr.)’ (Robbeets 2015: 310–311). These suffixes take part in the derivation of K *pusule tuli-* ‘to shatter into splinters (tr.)’, K *pusule ci-* ‘to crumble (intr.)’ and K *ppu:li-*, MK *spu-li-* ‘to sprinkle; scatter; sow’ from pK **pusu-* ‘to sprinkle, scatter, sow’. Korean has further lexicalized two adverbial suffixes pK **l* and pK **k*, for instance, in the derivation of *santul* ‘light’, *santul santul* ‘in cool ripples’ and *santuk* ‘with a sudden chill’ from pK **santi-* ‘to be light, fresh, cool’ (Robbeets 2015: 469–470). They participate in the derivation of K *pusul pusul* ~ *posul posul* ‘gently raining’ and K *pusik ha-* ‘to plant, extend’ from pK **pusu-* ‘to

sprinkle, scatter, sow. Moreover, the pluractional marker pK $^{*(\Lambda/i)l-}$, indicating that an action is carried out multiple times, by multiple agents or on multiple objects (e.g., in MK $\cdot spo(l)$ - ‘to sip, inhale,’ MK $\cdot awo(l)$ - ‘to join together’ and MK $\cdot sko(l)$ - ‘to spread out, pave with (tr.)’ vs. MK $\cdot ski$ - ‘cloud up’), derives MK $\cdot spol$ - ‘to wash (tr.)’ from pK *pisi - ‘to sprinkle, scatter, sow.’ Finally, the inclinational marker pK $^{*-mi/\Lambda}$, e.g., K *mek*-, MK *mek*- ‘to eat; harbor (a feeling) (tr.)’ → K *mekum*-, MK *me-kwum*- ‘to hold in the mouth; to swallow, gulp down; harbor (a feeling/idea) (tr.)’ (Robbeets 2015: 250–251) explains the formation of MK $\cdot spum$ - ‘sprinkle, spout, spurt’ from this root.

In Korean and Middle Korean, we find the causative suffixes K *ki*, *hi*, *i*, MK $\cdot Ki$, $\cdot Gi$, $\cdot hi$ -, $\cdot i$ - that can be derived through velar lenition as allomorphs from pK *ki , e.g., MK *cec*- ‘to be wet’ → *ce-ci*- ‘to moisten (tr.)’ and MK *nep*- ‘to be wide’ → MK *ne-phi*- ‘to widen (tr.)’ (Robbeets 2015: 320–321). These suffixes take part in the derivation of MK *puswoy*- ‘to wash, clean, rinse (tr.)’ from pK *pusu - ‘to sprinkle, scatter, sow’ and of MK *spih*- ‘to sprinkle’ from pK *pisi - ‘to sprinkle, scatter, sow.’

In Middle Korean, we find MK $\cdot psi$ ‘seed’ in addition to MK $\cdot phi$ ‘barnyard millet.’ As hinted above, tonic monosyllabic, open stems with aspirate initials followed by a minimal vowel (*u*, *o*, *i*) can be derived from an originally disyllabic root with an initial minimal vowel, i.e., in this case, pK *pisi ‘what is sown, seed.’ I assume that the addition of a velar plant suffix caused the aspiration in the term for ‘barnyard millet,’ i.e. pK $^{*pisi-k}$ (what.is.sown-PLANT) > *pski > *phi .

I do not exclude the possibility that the Japanese verb *hisigu* ‘crush, smash’ (< $^{*pisi-nku}$ -) and the verbal adjective *hisasii* ‘long, long-continued’ (< $^{*pisa-si}$ -) are ultimately related to this etymon. This remains speculative, but the coincidence in meaning between J *hie*, OJ *pi₁ye* and the Korean form can hardly be coincidental. Since the vowel type (1 or 2) is not distinguished following glides in Old Japanese, there is no conclusive evidence for the reconstruction of the final vowel in OJ *pi₁ye* ‘barnyard millet.’ The possibilities are *piyai ~ *piyia ~ *piye . The correspondence between the palatal glide *y*- in Japanese and the *s*- in Tungusic and Korean is irregular, but a few etymological sets within Japanese seem to involve internal alternation between *s* ~ *t* (e.g., *hisasii* ‘long, long-continued’ ~ *hita*- ‘straight, unceasing,’ *hutagu* ‘close, stop up’ ~ *husagu* ‘close, stop up,’ OJ *si* ~ *ti* ‘wind, direction’ etc.) and between *t* ~ *y* (e.g., *itamu* ‘hurt’ ~ *yamu* ‘ail,’ *taku* ~ *yaku* ‘burn (tr.),’ *tatu* ~ *tayasu* ‘cut off (tr.)’ etc.) Thus we cannot exclude that pJ *piyai ~ *piyia ~ *piye ultimately derives from *pisai ~ *pisia ~ *pise .

The convincing power of this etymology follows from the shared peculiarities of the Mongolic, Tungusic and Koreanic reconstructions. First, there is a shared alternation between the vowels in the verb bases that corresponds regularly and reconstructs back to a *u - ~ *i - vowel alternation in proto-Transeurasian. Second, the peculiar polysemy of ‘to sprinkle’ and ‘to sow’ is shared by the Mongolic,

Tungusic and Koreanic proto-forms. This polysemy is recurrent throughout the Transeurasian languages, including verb roots that are not cognate to the root under discussion, such as Japanese *maku* ‘to sprinkle, scatter, strew, sow (seed),’ *hodokosu* ‘sprinkle, scatter, sow; give, perform, apply,’ Sibe *swata-* ‘to sprinkle, sow,’ Turkish *saç-* ‘to sprinkle, scatter, sow (seed),’ *ek-* ‘to sprinkle, scatter, drop, throw about, sow (seed),’ etc. The derivation of the word for a major field crop by way of a nominalization of the verb ‘to sow,’ as proposed for the Tungusic term for ‘broomcorn millet’ and the Korean term for ‘barnyard millet,’ is reminiscent of the development of proto-Turkic **tari-* ‘to cultivate ground’ into the deverbal noun Uzbek *tariq* ‘broomcorn millet’ (Savelyev, this volume).

Third, the nominal derivations with a corresponding deverbal noun suffix are shared, as well as the suffixation of a velar plant suffix, in Tungusic and Koreanic. The formally and functionally corresponding derivations suggest that the suffixes were productive at their most recent common ancestral stage and probably on their way to lexicalization in the individual protolanguages. Due to these shared peculiarities at the phonological, semantic and phonological level, this etymology provides a strong argument for cognacy, while it is unlikely to be the result of borrowing.

From the perspective of cultural reconstruction, it is informative that the semantic development from ‘sprinkle’ to ‘sow’ and the morphological derivation from ‘sow’ to ‘what is sown’ to ‘seed’ took place at the stage of proto-Transeurasian. This allows us to infer that sowing, and thus plant cultivation, was adopted and gradually developed by the speakers of proto-Transeurasian. We find a similar situation in Indo-European, where the derivation from PIE **seH₁-* ‘to sow (seed)’ to **séH₁mn̥* ‘seed’ can be reconstructed to the level of the ancestral language because both the verb roots and derived nouns are regularly corresponding and derived by way of a common deverbal noun suffix: e.g., in Germanic, Old English *sāwan* ‘to sow,’ Gothic *saian* ‘to sow’ and Old High German *sāmo* ‘seed’; in Romance, Latin *serō* ‘I sow’ and *sēmen* ‘seed’; in Slavic Old Church Slavonic *sejō* ‘to sow’ and *seme* ‘seeds’; in Baltic, Old Prussian *situn* ‘to sow’ and *simen* ‘seed,’ Lithuanian *sėti* ‘to sow’ and *sėkla* ‘seed,’ *sėmenis* ‘linseed’; in Celtic, Old Irish *sīl*, Welsh *hil* ‘seed’; in Sanskrit *śīra-* ‘plow’; and in Hitite *ishūwāi* ‘(he) sows’.

The common derivation from the verb ‘to sow’ as well as the shared combination of the two meanings ‘seed, millet’ in Tungusic and Korean seems to imply that some kind of millet was targeted for its seeds and existed as a major crop in the culture in which the ancestral language was spoken. Although there is no evidence for full domestication of barnyard grass in northeast China in the Neolithic period, it is known that it formed part of the diet. The narrow range of wild grasses recovered in Neolithic sites in dry farming contexts in northeast China indicates that people were selecting the wild ancestor of Japanese barnyard millet as opposed to other grasses (Bestel et al. 2014: 264). Seeds of barnyard millet were also

retrieved from early agricultural sites of the Zaisanovka culture in the Russian Far East (Kuzmin 2013).

- (3) pTEA **kuru* ‘nut used for starch production such as walnut, acorn, chestnut or pine nut’
- a. Tungusic: pTg **kuri* ‘pine cone, pine nut’ (pTg *-*ktA* collective for small items)
Ma. *xuri* ‘cone of coniferous trees’, Jur. *xuri* ‘cone of coniferous trees’, Evk. *korekta* ‘cedar nut’ (Menges 1983: 274), Na. *koriči* ‘water chestnut’, *korekta* ‘pine cone, cedar cone’
 - b. Koreanic: pK **kul* ‘oak < ? walnut’
K / MK *kwul* ‘oak’ in K *kwul pa:m* ‘acorn’ (K *pa:m* ‘chestnut’), MK *kwul pam* ‘bristletooth oak (*Quercus serrata*)’, K *kwul cham-namu* ‘oriental oak (*Quercus variabilis*)’ (K *cham-namu* ‘oak tree’), K *kwul phi* ‘oak bark’ (K *phi* ‘bark’), *kwul phi namu* ‘Walnut-like tree (*Platycaria strobilacea*)’
 - c. Japonic: pJ **kuru* ‘walnut, chestnut’
J *kuri* (2.3), OJ *kuri* ‘chestnut’, J *kurusu* ‘chestnut grove’, OJ *kuri/u-kuma* ‘Chestnut Corner’, J *kurumi*, MJ *kurumi* ‘walnut (*Juglans regia*)’ (MJ *mi* ‘fruit, nut’)
- (4) pTEA **xusi* ‘nut used for starch production such as walnut, acorn, chestnut or pine nut’
- a. Mongolic: pMo **kusi* ‘walnut’ (pMo *-*Ga(n)* diminutive, often in plant names, e.g. WMo. *čibaya(n)* ‘jujube’, *abuya* ‘marshmallow’ etc.)
WMo. *qusiga* ‘walnut, nut; testicles’, Khal. *xušga* ‘walnut’, Kalm. *xušg* ‘walnut’, Ordos *gūšigā* ‘walnut’, WMo. *qusi* ~ *qosi* ‘cedar, Siberian pine’, Khal. *xuš* ‘cedar, Siberian pine’, Kalm. *xoš* ‘cedar, Siberian pine’.
 - b. Tungusic: pTg **xusi* ‘acorn’ (pTg *-*ktA* collective for small items)
Ma. *usixa* ‘big nut’, Evk. *usikta* ‘oak tree’, Na. *xosaqta* ‘acorn’, Ud. *uhikta* ‘acorn’
 - c. Japonic: pJ **kusi* ‘chestnut’
OJ *kusi* ‘chestnut’

During the Neolithic, the West Liao River region consisted for 55% of trees, a mix of conifer and broadleaf trees, the latter category being predominantly oak (*Quercus*) and walnut (*Juglans*) and also some chestnut. Wild walnuts (*Juglans mandshurica Maxima*) are found on the floors of houses at the Xinglongwa site (Shelach 2000: 380). Analyzing starch residue on grinding stones Liu (2016) finds that people processed acorns and several plant roots for starch at least as frequently as millets. It is probably significant that it is precisely nuts such as walnut, acorn, chestnut or pine nut, which were targeted for their starch and consumed by Xinglongwa people, that turn up in the etymologies. Walnuts and acorns were also stored at early agricultural sites of the Zaisanovka culture in the Russian Far East (Kuzmin 2013).

- (5) pTEA **abu* ‘plant of the *Althaea* genus with roots rich of starch’
- a. Mongolic: pMo **abu* ‘marshmallow (*Althaea officinalis*)’ (pMo *-*Ga(n)* diminutive, often in plant names, e.g. WMo. *čibaya(n)* ‘juzube’, *qusiga* ‘walnut, nut’, etc.)
WMo. *abuya*, Khal. *avga* ‘marshmallow (*Althaea officinalis*)’
 - b. Koreanic: pK **apok* ‘marshmallow (*Althaea officinalis*)’
K *awuk*, MK *a-wok* ‘marshmallow, *Althaea officinalis*’, modern dialect forms *apuk*, *apok*, *akwuk*, *akwu*
 - c. Japonic: pJ **apupi* ‘hollyhock (*Althaea rosa*)’
J *aoi* (3.1), OJ *apupi*₁ ‘hollyhock (*Althaea rosa*)’

According to Liu (2016) roots and bulbs were targeted for their starch. The root of plants of the *Althaea* genus are also used medicinally.

5.2 Subsistence activities

- (6) pTEA **nap-* ‘to make rope’
- a. Tungusic: pTg **nap-* ‘to make rope’ (pTg *-*ki* resultative nominalizer; Robbeets 2015: 407)
Ulcha *lāxi*, Orok *lāpu*, Na. *lāpi*, Oroch *lappi* ‘tiers, straps (for skis)’
 - b. Koreanic: pK **nap-* ‘twist, spin’
K *nah-* ‘spin, weave, make yarn’, K *kkunapwul* ‘a string of cord’ < *kkun* ‘cord, string’ + **nap-* ‘twist, twine, spin’ + *-wul* deverbial nominalizer, Kyeylim Yusa phonogram EMK *na(h)* ‘string’
 - c. Japonic: pJ **nap-* ‘to make rope’ (pJ *-*a* deverbial nominalizer; Sakakura 1966: 286–303; Robbeets 2015: 156)
J *nau* (B), OJ *nap-* ‘twist, plait, weave (into rope)’, J *nawa* (2.3), OJ *napa* ‘rope’

The Tungusic words for ‘tiers, straps (for skis)’ can be derived with the resultative deverbial noun suffix pTg *-*ki* from an underlying verb **nap-* ‘to make rope’. Proto-Tungusic lacks initial liquids, except **l-* going back to original nasal **n-* assimilation before labial consonants (Poppe 1960: 74; Robbeets 2005: 69).

Twining can produce cloth, string or rope. Cords for making traps and nets have been found in a number of upper Paleolithic sites across the world (Tedlock 2009: 66; Soffer et al. 2000: 512–514). Therefore, twining is not necessarily linked to agriculture.

- (7) pTEA **nup-* ‘to sew’
- a. Tungusic: pTg **nup-* ‘to prick, pierce’
Neg. *lepu-* ‘to pierce’, Na. *lopqa-*, *loqpa-* ‘to prick’, Olcha *loqpa-* ‘to prick’, Orok *lüqqa-* ‘to prick’, Evk. *lupa-* ‘to prick’, *lupu:-* ‘to pierce’, Even *nūbas an-* ‘to prick’
 - b. Koreanic: pK **nwupi-* ‘to sew, quilt’
MK *nwu(-)pi-* ‘to quilt’, MK *nwu-pi* ‘quilting’
 - c. Japonic: pJ **nup-* ‘to sew, stitch’
J *nuu* B, OJ *nup-* ‘to sew, stitch, embroider’, pR **noCu-* ‘to sew’, Shuri (Okinawa) *no: yun*, Hirara (Miyako) *nu:* ‘to sew’, Igarashi (Yaeyama) *no:η* ‘to sew’, Yonaguni *nuy* ‘to sew’

The Tungusic verb stem is probably a compound of pTg **nup-* ‘to prick, pierce’ with a suffix **-ka-*, perhaps the allomorph of the inchoative suffix pTg **-xa-* in voiceless clusters (Robbeets 2015). Similar to the phonological environment in etymology (5) initial *l-* in the Tungusic languages is a secondary development from an original **n-*. Note that Even consistently retains the initial nasal here.

Sewing enters the archaeological record with leather clothing, and is generally older than weaving textiles. Therefore, it is not necessarily linked to agriculture.

- (8) pTEA **pɜ:rɔ-* ‘to weave’
- a. Turkic: pTk **pö:r-* ‘to plait, weave’
OT (Karakh.) *ör-* ‘to plait (hair or other fibers)’, MTk. *ör-* ‘1 to weave, plait, twist things together’, *örmek* ‘cloth woven from camel hair’, Kirg. *ör-* ‘1’, Kaz. *ör-* ‘1’, Nog. *ör-* ‘1’, Bash. *ür-* ‘1’, Karaim *ör-* ‘1’, Karakalpak *ör-* ‘1’, Tatar *ör-* ‘to plait, to knit, to darn, to interlace, to interweave, to build (a wall), to lay bricks or stones in a building’, Tk. *ör-* ‘1’, Az. *hör-* ‘1’, Tkm. *ö:r-* ‘1’, Gag. *yör-* ‘1’, Uz. *or-*, Uig. *ö(r)-*, Yakut *ör-*, *örü* ‘plaiting’, Dolgan *ör-* ‘to plait, bind together, wind’, *örü* ‘plaiting’, Khalaj *hīri-*, *hör-* ‘to plait’, Chu. *var* ‘best sort of flax’, *vēren* ‘cord, rope’
 - b. Mongolic: pMo **poro-* ‘to entwine’ in **poro-go-* ‘to wrap’ (**-ga-* causative) and **poro-ti-* ‘roll, rotate’ (**-ti-* intensive)
WMo. *oriya-* ‘1 to tie around, entwine, wrap, bandage, wind, roll (tr.)’, *oruɣa-* ‘1’, *orči-* ‘2 to turn around, roll, rotate’ (intr.), MMo. *hura-* ‘1’, *xorči-*, *horči-*, *orči-* ‘2’, *orčul-* ‘2’, Khalkha *oro:-* ‘1’, *orči-* ‘2’, Buriat *oō-* ‘1’, *oršo-* ‘2’, Kalmuck *ora:-* ‘1’, *orčə-* ‘2’, Ordos *oro:-* ‘1’, *orčin* ‘around’, Dong. *xoro-* ‘1’, Baoan *horə-*, Dagur *óre:-*, Shira-Yughur *horo:-*, Monguor *furo:-*, *xuro:-* ‘1’
 - c. Tungusic: pTg **poro-* ‘to spin, weave (nets)’
Evk. *horol-* ‘1 to spin, whirl, go around’, Neg. *xoyol-*, *xoyil-* ‘1’, Ud. *xo:li-* ‘1’, Sibe *forə-*, *foru-* ‘1’, Ma. *foro-* ‘to turn round, turn over’, *foringa-* ‘1’, Olcha *pori-* ‘to weave (nets)’, *porpun* ‘device for weaving nets’, *po:rɣu* ‘spindle’, Oroch *po:rpu*, *po:rɣu* ‘spindle’

- d. Koreanic: pK **ol* ‘unit of woven fibers, component of woven fabric’
 K *o:l*, MK **wol* ‘strand of rope, ply, warp’, K *olk-* ‘to tie up, bind, weave’
 (< pK **ol* ‘woven fabric’ + *-k* λ - inchoative; Robbeets 2015: 258)
- e. Japonic: pJ **orə-* ‘to weave’
 J *oru* A ‘weave’, OJ *oro₂s-* ‘deign to weave’, Shuri *qur-* ‘weave’

For Turkic, it is commonly assumed that word initial pTk **p-* developed over a bilabial fricative into *h-*, leaving only a trace in Khalaj *h-* and finally disappearing in most of the contemporary Turkic languages. Given the attestation of Khalaj *hör-* ‘plait’, it is legitimate to reconstruct pTk **pö:r-* ‘to plait, weave’. The initial labial stop pMo **p-* is regularly preserved in the peripheral Mongolic languages, notably as *f-* in Monguor *furo:-*, as *h-* in Shira-Yughur *horo:-* or Baoan *horə-* and as *x-* in Dongxiang *xoro-*, but it disappeared in the central Mongolic languages. The regular reflexes of pTg **p-* are Nanai/Olcha/Orok *p-*, Manchu *f-*, Evenki/Even *h-*, Negidal/Oroch/Udehe *x-* and Solon \emptyset (Benzing 1956: 981). Except for Oroch *po:rpu*, *po:rfu* ‘spindle’, which is probably a borrowing from Olcha, the cognates thus correspond regularly and suggest the reconstruction of an initial pTg **p-*. The expected reflex of pTEA **p-* is **p-* in proto-Japonic and proto-Korean (Robbeets 2005: 373). However, an initial labial stop sporadically drops before a (long?) rounded pJK **o(:)*, as it probably also did in the reflexes of pTEA **bɔ:l-* ‘to sit down, become, be’ in Japanese and Korean (Robbeets 2015: 159–163). Since Old Japanese makes no distinction between *o₁* (< **o*) and *o₂* (< **ə*) in initial position, I have opted for **o* in pJ **orə-* ‘to weave’ because it entails a regular correspondence (Robbeets 2015: 128). The root-final vowel of pJ **orə-* is an irregular fit, which may be due to vowel reduction in root-final position.

Whereas twining and sewing are not necessarily linked to agriculture, weaving certainly is. There are no pre-agricultural textiles in North and East Asia because weaving is labor-intensive and technologically complex, requiring a loom system. Only a society with food-surplus can invest in the technology and labor required (Barber 1995).

- (9) pTEA **suru-* ‘to grind’
- a. Turkic: pTk **sür(ü)-* ‘to rub, smear’ (pTk **-ti-* causative-passive; Robbeets 2015: 290–292)
 OT *sürt-* ‘1 to rub, smear (tr.)’, MTk. *sür-*, *sürüt-*, *sürt-* ‘1’, Tk. *sür-*, *sürt-*, Az. *sürt-*, Tkm. *sür-*, *sürt-*, Gag. *sürüt-*, Uz. *surt-*, Tuva *sür-*, Yakut *ür-*, Khak. *sürt-*, Kirg. *sür-*, *sürt-*, Kaz. *sürt-*, Nog. *sür-*, *sürt-*, Bash. *hür-*, *hürt-*, Balk. *sürt-*, Karaim *sürt-*, Kpak. *sür-*, *sürt-*, Kum. *sürt-*, Chu. *sér-*
- b. Tungusic: pTg **suru-* ‘to grind’
 Ma. *šuru-* ‘to grind, whet, sharpen’

- c. Japonic: pJ **sura*- ‘to grind, rub’
 J *sur*- (B), *sure*- (B) ‘to rub against each other’, OJ *sur*- ‘to grind, rub’, J *surari* ‘without trouble, smoothly’ (-*ri* adverbializer), *sura-sura* ‘without a hitch, smoothly’, Shuri *sir*- ‘rub, grind’, *šiyuŋ* ‘to rub’, Shodon *k’usryum*, Hirara *sipag^{zi}*, Ishigaki *sisun*, Kabira *suri*, Yonaguni *ccitun*, *cirun*, pR **suri*- ~ **ko-suri*- ‘to rub’

With only a Manchu cognate, the reflex of this word is poorly distributed in Tungusic. In a few cases Manchu displays a palatal sibilant *š*- rather than *s*- in correspondence with words with initial *h*- in Even and initial *s*- in the other Tungusic languages. There is no internal ground for this palatalization, such as a following high vowel. However, as it concerns only a few cases and since the palatalization is restricted to Manchu, Benzing (1956: 989–990) refrains from establishing a separate palatal sibilant **š*- in proto-Tungusic.

Liu (2016: 247) stresses the significance of grinding stones throughout the entire Neolithic period in the Liao River region of Northeast China, whereas they gradually disappear from the archaeological record in the Yellow River region after 5000 BCE when millet-based agriculture was intensified. The significance of ‘grinding’ for Xinglongwa people is corroborated by the reconstructions for ‘grinding’ in (9) and ‘crushing food to pulp’ in (10).

- (10) pTEA **niku*- ‘to crush, knead’
- a. Turkic: pTk **yik*- ‘to crush, demolish, destroy’
 Otk. *yik*- ‘1 to crush, demolish, destroy’, Karakhanid *yiq*- ‘1’, MTk. *yiq*- ‘1’, Tk. *yik*- ‘1’, Az. *yix*- ‘1’, Tkm. *yiq*- ‘1’, Gag. *yiq*- ‘1’, Tat. *yiq*- ‘1’, Kirg. *žiq*- ‘1’, Karaim *yiq*- ~ *yix*- ‘1’, Kaz. *žiq*- ‘1’, Nog. *yiq*- ‘1’, Bash. *yiq*- ‘1’, Kpak. *žiq*- ‘1’, Kum. *jiq*- ~ *jix*- ‘1’, Uz. *yiq*- ‘1’, Uig. *yiq*- ‘1’, Khak. *yuq*- ‘1’, Oirat *yiq*-, *điq*- ‘1’, Khalaj *yuq*- ‘1’, Chu. (dial.) *šāx*- ‘1’
 - b. Mongolic: pMo **niku*- ‘to knead, crush’
 WMo. *niqu*- ~ *nuqu*- ‘1 to knead (flour), mash, crumple, rub, press, massage’, *niquyur* ‘implement for kneading dough’, MMo. *nuqu*- ‘1’, Khal. *nuxa*- ‘1’, Bur. *ñuxa*- ‘1’, Kalm. *nuxə*- ‘1’, Ordos *nuxu*- ‘1’, Bao. *noğə*- ‘1’, Dag. *nogu*- ‘1’, Monguor *nuğu*- ‘1’, Mog. *nuqu*- ~ *noqu*- ‘to crush’, Dong. *nuqu*- ‘to hit with force’
 - c. Koreanic: pK **niki*- ‘to crush to a pulp, knead’
 K *iki*-, MK *niki*- ‘to crush to a pulp, mash, knead, beat water into flour’
- (11) pTEA **səmtu*- ‘to form a layer on the surface by oxidation’
- a. Tungusic: pTg **septu*- ‘to become rusty’
 Evk. *semtu*- ‘to become rusty’, *semtu* ‘rust’, *semtuce*: ‘rustle, rusty’, Neg. *semti* ‘rust’, Oroch *semtu*- ‘to become rusty’, *semtu* ‘rust’, Ud. *semtu*- ‘to become rusty’, Olcha *septu*- ‘to become rusty’, *septuče* ‘rust’, Orok *septu*

‘rust’, Na. *septu-* ‘to become rusty’, *septuce* ‘rust’, Ma. *sebde-* ‘to become rusty’, *sebden* ‘rust’

- b. Japonic: pJ **sampu-* ‘to become rusty, to form a layer on the surface by oxidation’

J *sabi* (2.3), OJ *sabi*₂ ‘rust, tarnish, patina’, J *sabiru* (B), OJ *sabi*₂₋ ‘to rust, form rust, to get rusty/old, to mature and perish after spawning (of fish)’, Shuri *sabi* ‘rust’

The cluster correspondence reflects a regular heterogonic cluster correspondence pTEA **m^(P)T-*, whereby the nasal and the stop have a different place of articulation, which results in the insertion of a parasitic stop (Robbeets 2015: 147). The nasal is lost in the continental Transeurasian languages (here pTg **-pt-*), whereas Japanese has lost the final stop (pJ **-mp-* > OJ *-b-*).

At the first glance, this etymology may be somewhat puzzling because it seems to imply familiarity with iron. A similar paradox is found in the reconstruction of proto-Austronesian, where PAN **Namat* ‘iron’, and **diNanɿ* ‘rust’ can be reconstructed at a time depth of 3500 BC, in spite of the fact that metallurgy appeared in South East Asia only about 3000 years later. However, Blust (2013) argued that knowledge of iron does not necessarily imply knowledge of metallurgy. The Austronesian terms may be related to early Neolithic hematite pottery production, whereby iron-rich clay was turned red through a process of oxidation. It is known that at the very beginnings of pottery production in Xinglongwa, the color of different wares was important. Many ceremonial items were reddish in color, while others were grey and black. The clays were composed of ferrous minerals such as hematite (Li 2016) and the colours were attained by oxidation of these clays invoked during firing. Pottery rather than metallurgy may be the context within which this etymology should be understood.

6. Conclusion

Starting from the assumption that the Transeurasian languages represent a valid genealogical grouping, I investigated the impact of agriculture on the ancestral vocabulary as well as on the primary dispersals of proto-Transeurasian. Applying different techniques situated at the intersection of linguistics and other disciplines such as archaeology and genetics, I reached the following conclusions:

1. Proto-Transeurasian, the language ancestral to the Turkic, Mongolic, Tungusic, Koreanic and Japonic languages, reflects a broad-spectrum subsistence strategy probably including some plant cultivation and yielding food surpluses.

2. The assumed location and time depth of proto-Transeurasian associate the ancestral language with the Xinglongwa culture, the first farming society in Northeast China in the 7th and 6th millennium BC.
3. The spread of the Transeurasian languages to their present-day locations is consistent with the spread of agriculture in Northeast Asia. However, agriculture did not necessarily cause language spread by boosting the farmer's demography and pushing them to search for new land. It also followed ecological stress caused by climate change, disrupting traditional resource bases and replaced previous subsistence strategies.

Cultural reconstruction indicates that the speakers of proto-Transeurasian targeted a millet-like crop for its seeds, sowed seeds and maintained fields for cultivation. Their food surpluses were sufficient to permit labor-intensive and technologically complex activities such as weaving. They were familiar with a process of oxidation, probably in connection with iron-rich clay in hematite pottery production. In contrast to the communities in the Yellow River Basin, the speakers of proto-Transeurasian relied intensively on grinding for their food-production. The starches involved in this process were not limited to millets, but were provided by various nuts such as walnut, chestnut, acorn and pine as well as roots. The reconstructed vocabulary therefore suggests a broad-spectrum subsistence strategy with some economic dependence on the cultivation of plants such as millets.

The lexical evidence is in line with the diversity hot-spot principle, locating the homeland of Transeurasian in the West Liao River region and Bayesian inference, estimating the time-depth of the family at ca. 5700 BC. The location and time depth indicate that proto-Transeurasian may be connected with the Xinglongwa culture (6200–5400 BC) in Southern Manchuria. This culture depended on a broad-spectrum subsistence strategy including millet cultivation.

Towards the end of the Xinglongwa culture, the population expanded quickly and millet agriculture started spreading eastwards. The resulting demographic processes can be mapped on the Transeurasian phylolinguistic tree to such an extent that the major splits in the language family seem to coincide with the time and the route of agricultural expansions in Northeast Asia. This indicates that the eastward spread of the Transeurasian languages may indeed have been driven by agriculture.

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 programme (grant agreement No 646612) granted to Martine Robbeets.

Abbreviations

Az.	Azerbaijani	Mog.	Moghol
Balk.	Balkar	MTk.	Middle Turkic
Bao.	Bao'an	Na.	Nanai (Goldi, Hezhe)
Bur.	Buriat	Neg.	Negidal
Chu.	Chuvash	Nog.	Noghay
Dag.	Dagur	OJ	Old Japanese
Dong.	Dongxiang (Santa)	OT	Old Turkic
EMK	Early Middle Korean	pJ	proto-Japonic
Evk.	Evenki (Tungus)	pK	proto-Koreanic
Gag.	Gagauz	pMo	proto-Mongolic
J	(contemporary, standard Tokyo) Japanese	pTEA	proto-Transeurasian
Jur.	Jurchen	pTg	proto-Tungusic
K	(contemporary, standard Seoul) Korean	pTk	proto-Turkic
Kalm.	Kalmuk	pR	proto-Ryukyuan
Kaz.	Kazakh	Tat.	(Volga) Tatar
Khal.	Khalkha	Tk.	Turkish
Kirg.	Kirghiz	Tkm.	Turkmenian
Kpak.	Kara-Kalpak	Ud.	Udehe
Kum.	Kumyk	Uig.	Uighur
Ma.	Manchu	Uz.	Uzbek
Mgr.	Monguor	WMo.	Written Mongolian
MJ	Middle Japanese	'1'	Same semantics as the first meaning given
MK	Middle Korean	'2'	Same semantics as the second meaning given
MMo.	Middle Mongolian		

References

- Ahn, Sung-Mo. 2010. The emergence of rice agriculture in Korea: Archaeobotanical perspectives. *Archaeological and Anthropological Sciences* 2: 89–98. doi:10.1007/s12520-010-0029-9
- Ahn, Sung-Mo, Kim, Jangsuk & Hwang, Jaehoon. 2015. Sedentism, settlements, and radiocarbon dates of Neolithic Korea. *Asian Perspectives* 54 (1): 113–143. doi:10.1353/asi.2015.0005
- Barber, Elisabeth. 1995. *The Mummies of Urumchi*. New York NY: Norton.
- Bawden, Charles. 1997. *Mongolian-English Dictionary*. London: Routledge.
- Benzing, Johannes. 1956. *Die tungusischen Sprachen. Versuch einer vergleichenden Grammatik*. Abhandlungen der geistes- und sozialwissenschaftlichen Klasse 11. Wiesbaden: Akademie der Wissenschaften und der Literatur.
- Bestel, Sheahan *et al.* 2014. The evolution of millet domestication, Middle Yellow River Region, North China: Evidence from charred seeds at the late Upper Paleolithic Shizitan Locality 9 site. *The Holocene* 24: 261.

- Blust, Robert. 2013. Formosan evidence for Early Austronesian knowledge of Iron. *Oceanic Linguistics* 52: 255–264. doi:10.1353/ol.2013.0004
- Cincius, Vera Ivanovna (ed.) 1975–1977. *Sravnitel'nyj slovar' tunguso-mančžurskich jazykov*. Vol. 1–2. Moscow: Nauk.
- Crawford, Gary W. & Lee, Gyoung-Ah. 2003. Agricultural origins in the Korean Peninsula. *Antiquity* 77: 87–95. doi:10.1017/S0003598X00061378
- Crawford, Gary W. & Shen, C. 1998. The origins of rice agriculture: Recent progress in East Asia. *Antiquity* 72: 858–66. doi:10.1017/S0003598X00087494
- Drennan, Robert & Dai, Xiangming. 2017. Chiefdoms and states in the Yuncheng Basin and the Chifeng region: A comparative analysis of settlement systems in North China. *Journal of Anthropological Archaeology* 29 (4): 455–468. doi:10.1016/j.jaa.2010.09.001
- Hanihara, Kazurō. 1991. Dual structure model for the population history of the Japanese. *Japan Review* 2: 1–33.
- Heggarty, Paul & Beresford-Jones, David. 2014. Farming-language dispersals: A worldwide survey. In *Encyclopedia of Global Archaeology*, Claire Smith (ed.), 1–9. New York: Springer.
- Hunt, Harriet *et al.* 2008. Millets across Eurasia: chronology and context of early records of the genera *Panicum* and *Setaria* from archaeological sites in the Old World. *Vegetation History and Archaeobotany* 17 (1): 5–18. doi:10.1007/s00334-008-0187-1
- Janhunen, Juha. 1996. *Manchuria: An Ethnic History* [Mémoires de la Société Finno-Ougrienne 222]. Helsinki: Suomalais-Ugrilainen Seura.
- Jeong, Choongwon, Nakagome, Shigeki & Di Rienzo, Anna. 2016. Deep history of East Asian populations revealed through genetic analysis of the Ainu. *Genetics* 202: 261–272. doi:10.1534/genetics.115.178673
- Jia, Xin *et al.* 2017. Spatial and temporal variations in prehistoric human settlement and their influencing factors on the south bank of the Xar Moron River, Northeastern China. *Frontiers of Earth Science* 11 (1): 137–147. doi:10.1007/s11707-016-0572-5
- Jinam, Timothy *et al.* 2012. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *Journal of Human Genetics* 57: 787–795. doi:10.1038/jhg.2012.114
- Johanson, Lars & Robbeets, Martine. 2010. Introduction. In *Transeurasian Verbal Morphology in a Comparative Perspective: Genealogy, Contact, Chance* [Turcologica 78], Lars Johanson & Martine Robbeets (eds), 1–5. Wiesbaden: Harrassowitz.
- Kanzawa-Kiriyama, Hideaki *et al.* 2016. A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *Journal of Human Genetics*, 1–9.
- Kim, Juwon, Dongho Ko, D. O. Chaok, Yonfeng Han, Lianyu Piao, & B. V. Boldyrev. 2008. *Materials of Spoken Manchu*. Altaic Languages Series 1. Seoul: Seoul National University Press.
- Kuzmin, Yaroslav V. 2013. The beginnings of prehistoric agriculture in the Russian Far East. Current evidence and concepts. *Documenta Praehistorica* 40: 1–12. doi:10.4312/dp.40.1
- Lee, Gyoung-Ah. 2011. The transition from foraging to farming in prehistoric Korea. *Current Anthropology* 52 (4): 307–329. doi:10.1086/658488
- Li, Tao. 2016. *Economic differentiation in Hongshan core zone communities (northeastern China): A geochemical perspective*. University of Pittsburgh doctoral dissertation.
- Liu, Li *et al.* 2016. Plant-based subsistence strategies and development of complex societies in Neolithic Northeast China: Evidence from grinding stones. *Journal of Archaeological Science: Reports* 7: 247–261. doi:10.1016/j.jasrep.2016.04.014

- Liu, Xinyi *et al.* 2012. The earliest evidence of millet as a staple crop: New light on Neolithic foodways in North China. *American Journal of Physical Anthropology* 149: 1–8. doi:10.1002/ajpa.22127
- Martin, Samuel. 1961. *Dagur Mongolian: Texts and Lexicon*. Bloomington: Indiana University Press.
- Martin, Samuel Elmo. 1996. *Consonant Lenition in Korean and the Macro- Altaic question*. Honolulu HI: University of Hawaii Press.
- Menges, Karl Heinrich. 1984. Korean and Altaic. A Preliminary Sketch. *Central Asiatic Journal* 28: 234–295.
- Norman, Jerry. 2013. *A Comprehensive Manchu-English Dictionary*. Harving-Yenching Institute Monograph Series 85. Cambridge, MA: Harvard University Press.
- Nugteren, Hans. 2011. *Mongolic Phonology and the Qinghai-Gansu Languages*. PhD dissertation, Leiden University.
- Peterson, Christian & Drennan, Robert. 2011. Patterned variation in regional trajectories of community growth. In *The Comparative Archaeology of Complex Societies*, Michael E. Smith (ed.), 88–137. Cambridge: CUP. doi:10.1017/CBO9781139022712.008
- Poppe, Nicholas. 1960. *Vergleichende Grammatik der altaischen Sprachen, Teil 1: Vergleichende Lautlehre* [Porta Linguarum Orientalium, Neue Serie 4]. Wiesbaden: Otto Harrassowitz.
- Ramsey, Samuel Robert. 1993. Some remarks on reconstructing earlier Korean. *Language Research* 29: 433–441.
- Ramsey, Samuel Robert. 1997. The invention of the Korean alphabet and the history of the Korean language. In *The Korean Alphabet. Its History and Structure*, Young-Key Kim-Renaud (ed.), 131–143. Honolulu: University of Hawai'i Press.
- Ramstedt, Gustaf John. 1906. *Mogholica. Beiträge zur Kenntnis der Monghol-Sprache in Afghanistan*. Journal de la Société Finno-Ougrienne. Helsinki: Suomalai-Ugrilainen Seura.
- Ramstedt, Gustaf John. 1935. *Kalmückisches Wörterbuch*. Lexica Societatis Fenno-Ugricae 111. Helsinki: Suomalai-Ugrilainen Seura.
- Ramstedt, Gustaf John. 1949. *Studies in Korean Etymology* [Mémoires de la Société finno-ougrienne 95]. Helsinki: Suomalai-Ugrilainen Seura.
- Robbeets, Martine. 2005. *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* [Turcologica 64]. Wiesbaden: Harrassowitz.
- Robbeets, Martine. 2015. *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages* [Trends in Linguistics Studies and Monographs 291]. Berlin: Mouton de Gruyter. doi:10.1515/9783110399943
- Robbeets, Martine. 2017. Austronesian influence and Transeurasian ancestry in Japanese: A case of farming/language dispersal. *Language Dynamics and Change* 7.
- Robbeets, Martine & Bouckaert, Remco. Forthcoming. Bayesian phylolinguistics and the classification of the Transeurasian languages.
- Sakakura, Atsuyoshi. 1966. *Gokōsei no kenkyū* [A study of word formation]. Tokyo: Kadokawa shoten.
- Shelach, Gideon. 2000. The earliest Neolithic cultures of northeast China. *Journal of World Prehistory* 14: 363–414. doi:10.1023/A:1011124209079
- Soffer, Olga, Adovasio, James M. & Hyland, David C. 2000. The “Venus” figurines. Textiles, basketry, gender, and status in the Upper Paleolithic. *Current Anthropology* 41: 511–537.
- Starostin, Sergej, Dybo, Anna & Mudrak, Oleg. 2003. *Etymological Dictionary of the Altaic Languages*. Leiden: Brill.

- Tadmor, Uri, Haspelmath, Martin & Taylor, Bradley. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2): 226–246. doi:10.1075/dia.27.2.04tad
- Tedlock, Barbara. 2009. *The Woman in the Shaman's Body: Reclaiming the Feminine in Religion and Medicine*. London: Random House.
- Vostretsov, Yuri E. 2006. Turning points in the cultural evolution of prehistoric Primorye. *Archaeology, Ethnology & Anthropology of Eurasia* 27(3): 25–32. doi:10.1134/S1563011006030030
- Wang, Chuan-Chao *et al.* 2016. Reconstructing population history in East Asia. Presentation held at the *7th International Symposium on Biomolecular Archaeology (ISBA7)*. The Oxford University Museum of Natural History, Oxford. 14th–16th September.
- Weber, Steven A. & Fuller, Dorian Q. 2008. Millets and their role in early agriculture. *Pragdhara* 18: 69–90.
- Zhao, Zhijun. 2011. New archaeobotanic data for the study of the origins of agriculture in China. *Current Anthropology* 52: 295–306. doi:10.1086/659308

Farming-related terms in Proto-Turkic and Proto-Altaic

Alexander Savelyev

Max Planck Institute for the Science of Human History, Jena

Historical sources from different times describe Turkic, Mongolic and Tungusic traditional economies as based on pastoralism, with agriculture playing only a minor role among their subsistence strategies. Cultural reconstruction as used by historical linguists may provide additional inferences about the relative importance of farming and pastoralism in these lineages. This paper focuses on the origin of agricultural and pastoralist terms in Proto-Turkic and their parallels in the other branches of Altaic, i.e., Mongolic and Tungusic. I show that the majority of the Turkic pastoralist lexicon has a secondary nature, being formed due to contact, derivation or lexical recycling. At the same time, farming-related terms in Turkic are mostly unborrowed and underived and a few of them have reliable Altaic connections. The very limited number of agricultural terms reconstructible to Proto-Altaic as compared to the preceding Proto-Transeurasian period can be attributed to a loss of farming-related lexicon over time after the break-up of Altaic.

Keywords: Proto-Turkic, Proto-Altaic, agriculture, pastoralism, cultural reconstruction

1. Introduction

The term “Altaic” as used in this paper refers to a grouping of three relatively well-described language families, i.e., Turkic, Mongolic and Tungusic. For a long time, the question of whether these families are genetically related has provoked a lively discussion among scholars, and it currently remains one of the most controversial issues in historical linguistics. All experts in the field, regardless of their position on the above question, agree that the relationships between the language families are extremely complicated due to extensive lexical borrowing, primarily from Turkic to Mongolic and from Mongolic to Tungusic. Some linguists, so-called

“Anti-Altaicists,” such as G. Clauson (1956), G. Doerfer (1963–1975) and A. Vovin (2005), believe that all the similarities between the three groupings can be explained either through multiple contacts or by pure coincidence. Their opponents, known as “Altaicists,” claim that it is nevertheless a genetic relationship that underlies striking lexical, morphological and structural similarities between Turkic, Mongolic and Tungusic, and that this proposition can be supported by a set of phonological correspondences, a list of cognates including some basic vocabulary items, and a number of shared grammatical units (see, e.g., Ramstedt 1952; Poppe 1960; Starostin et al. 2003 and Robbeets 2015 for different versions of Proto-Altaic grammar). Many of the contemporary proponents of Altaic unity, such as Menges (1975, 1984), Miller (1996), Starostin et al. (2003) and Robbeets (2005, 2015) argue that, coupled with the Japano-Koreanic branch, Altaic forms a larger family for which, following Johanson & Robbeets (2010), I use the term “Transeurasian.” In line with these authors, my study is based on the assumption that the Transeurasian languages can be traced back to a single ancestor and that there are close affinities within the Altaic group.

The Altaic languages provide a curious and rather peculiar case in terms of cultural reconstruction, particularly with regard to the question of what subsistence patterns can be assigned to the speakers of their ancestral language. Archaeological and historical sources from different times describe Turkic, Mongolic and Tungusic traditional economies as based on pastoralism, with agriculture playing only a minor role among their subsistence strategies (see, e.g., Golden 1992; Kljaštornyj & Sultanov 2009; Lane 2006; Turaev et al. 1997, 2001, 2003). In general, this can be confirmed by linguistic evidence, at least as far as we rely on etymological dictionaries of the respective language families (see, e.g., Sevortjan et al. 1974–2003 for Turkic, Sanžeev et al. 2015 for Mongolic and Tsintsius 1975 for Tungusic), all listing many more pastoralist terms than agricultural terms. However, the question remains as to whether there are correlations between these pastoralist and agricultural terms between the language families under discussion and if so, whether they are the result of language contact or inheritance.

This paper presents a comparative study of farming-related terms that can be reconstructed for two proto-languages, Proto-Turkic and its proposed ancestor Proto-Altaic. While Proto-Turkic cultural reconstruction has already attracted some attention from scholars (e.g., Tenišev et al. 2006), Proto-Altaic has hardly been discussed in this respect. To a certain extent, this can be attributed to the fact that it is not commonly accepted to distinguish between Proto-Transeurasian and Proto-Altaic reconstructions as the internal structure of the Transeurasian family itself is under discussion. To give one example, Starostin et al. (2003:235) argue that Proto-Transeurasian split into Turko-Mongolic, Tungusic and Japano-Koreanic around the 6th millennium BC. This classification leaves no room for “Proto-Altaic” as a

linguistic entity. However, the idea that has received a much broader acceptance among Altaicists is that Japonic separated from Transeurasian first and can be thus clearly distinguished from Altaiic, that is, Turkic, Mongolic and Tungusic (Miller 1996; Dybo 1997; Robbeets this volume). For example, Dybo argues for Proto-Altaiic (“continental Proto-Altaiic,” according to the author’s terminology) as a language that divided directly into Turkic, Mongolic and Tungusic, emphasizing an essentially even distribution of triple and paired lexical matches between the three branches. The question definitely requires further examination using the methods of phylogenetic linguistics (Saveljev forthcoming). In the meantime, I will follow preliminary Bayesian estimates by Robbeets (this volume), which are based on shared basic vocabulary items. They point to a binary split of Proto-Transeurasian into Proto-Altaiic and Proto-Japonic-Koreanic at approximately 5700 BC, with a subsequent split of Proto-Altaiic into Turko-Mongolic and Tungusic at approximately 4600 BC. For its part, Turko-Mongolic divided into Turkic and Mongolic at approximately 2800 BC. In this context, and given that closer genetic affinities generally imply more numerous lexical matches, below I focus on Turkic in the context of the other Altaiic branches, leaving aside the Japonic and Koreanic branches.

Only a few papers deal with the issues of the Proto-Altaiic homeland and cultural reconstruction of Proto-Altaiic as compared to those of Proto-Transeurasian. Robbeets (2015, 2017) associates the Proto-Altaiic and the Proto-Turko-Mongolic speech communities with the Neolithic Hongshan culture (ca. 4500–2900 BC) in the West Liao River Basin (Manchuria), which is thought to have relied on millet farming in combination with pig raising (Nelson 2001; Guo 1995). Robbeets hypothesizes that the Proto-Altaiic economy, as well as the preceding Proto-Transeurasian one, was in part based on cultivation of crops, with gradual domestication towards the Hongshan period, putting forward both linguistic and archaeological evidence in favor of this assumption. S. Starostin (2008) connects the Proto-Transeurasian homeland to the Yangshao culture (5000–2000 BC) along the central Yellow River, which is often associated with Proto-Sino-Tibetan. Dybo (1997) does not directly address the problem of localization and archaeological affiliation of Proto-Altaiic but assumes that, based purely on historical linguistic evidence, the Proto-Altaiic speakers were nomadic pastoralists rather than agriculturalists. This assumption contradicts archaeological evidence, since Proto-Altaiic as dated by historical linguists existed long before the advent of the first pastoralists (3000 BC), not to mention nomadic herders (between 1200 and 700 BC), on the eastern steppes (Taylor et al. 2017; Janz et al. 2017). Janhunen (2015), who is a critic of the Altaiic proposal, argues that the similarities between the three families should be primarily attributed to prehistoric mutual influence, which implies that Proto-Turkic, Proto-Mongolic and Proto-Tungusic speakers have long lived in close contact with each other. Quite interestingly, Janhunen places their homelands in the southern part

of the Mongolian-Manchurian border zone, also referring to the possibility of the Hongshan affiliation of Mongolic and/or Tungusic.

This paper addresses the following questions:

1. Can we reconstruct agricultural vocabulary for Proto-Turkic in addition to a lexicon of pastoralism? If so, what are the characteristics of the agricultural vocabulary in Proto-Turkic?
2. Can the identification of Proto-Turkic with the Xiongnu by previous scholars be corroborated by the investigation of pastoralist and agricultural vocabulary?
3. What are the origins of pastoralist and agricultural vocabulary in Proto-Turkic? Can the terms be shown to be internally coined or borrowed from non-Transeurasian languages?
4. Are there any similarities between Turkic agricultural and/or pastoralist terms and those in Mongolic and/or Tungusic? Is it possible to distinguish borrowing versus inheritance in these words? Is there a tendency for pastoralist vocabulary to be attributed to borrowing, while agricultural vocabulary may be a residue of inheritance from Proto-Altaiic, or vice versa?

My contribution has the following structure. In Section 2, I give an overview of the contemporary views of the Proto-Turkic homeland, historical affiliation and cultural reconstruction. In Section 3, I discuss the set of pastoralist terms in Proto-Turkic, marking probable borrowings and morphological derivatives. In Section 4, I apply the same procedure to the Proto-Turkic agricultural vocabulary. Then I discuss possible Altaic connections for Proto-Turkic pastoralist (Section 5) and agricultural (Section 6) vocabulary. I conclude with some inferences regarding the results of this study.

2. Proto-Turkic: Its homeland and historical background

The Turkic peoples are known to be traditionally nomadic or semi-nomadic pastoralists, which can be confirmed by various written sources from at least the second half of the first millennium AD onwards (for example, a herding lifestyle including horse riding is reflected in Old Turkic runic texts, such as the 8th-century Kul Tigin inscription from the Orkhon river valley in Mongolia). For those Turkic-speaking peoples that were described as agriculturalists rather than pastoralists in the past few centuries, such as the Chuvash in the Volga Basin, a relatively recent shift from nomadism to sedentarism has been attested.¹ The majority of traditional

1. Aḥmad ibn Faḍlān, who was a member of an embassy of the Abbasid Caliph to the Volga Bulgars, the ancestors of the modern Chuvash, in 922, witnessed that they lived in tents and their

Turkic societies practiced agriculture only as a secondary activity. Needless to say, one cannot automatically extrapolate such a situation to the Proto-Turkic period. However, one can provide some insights into the issue by integrating linguistic data with historical and archaeological evidence. To do so, it is first necessary to outline the contemporary views of the Proto-Turkic homeland and the probable historical affiliation of the Proto-Turkic speech community.

It is generally agreed among historians and linguists that the starting point of the Turkic migrations was located in the eastern part of the Central Asian steppe (see, e.g., Golden 1992; Kljaštornyj & Sultanov 2009; Menges 1995: 55). Turkologists use various definitions for describing the Proto-Turkic homeland, but most indicate more or less the same region. While Janhunen (1996: 26, 2015: 293) locates the Proto-Turkic homeland fairly precisely in Eastern Mongolia, Róna-Tas (1998: 88), in a rather general manner, places the last habitat of the Turkic speakers before the disintegration of the family “in West and Central Siberia and in the region south of it.” The latter localization overlaps in large part with that proposed by Tenišev et al. (2006), who associate the Proto-Turkic *urheimat* with the vast area stretching from the Ordos Desert in Inner Mongolia to the foothills of the Sayan-Altai Mountains in Southern Siberia. Such a vague localization seems to be quite compatible with the association of at least late Proto-Turkic speakers with nomadic herders. From a historical linguistic viewpoint, the region under discussion appears to be the most probable habitat for a language that is assumed to have been in contact with Old Chinese, Old East Iranian and possibly Tocharian (and, according to some scholars (see Dybo 2007), at the same time reaching the languages far to the north-west, such as Proto-Yeniseian, Proto-Samoyedic and Proto-Ugric). An attempt at verifying the homeland by examining archaeological and paleobotanical evidence, as well as the Proto-Turkic roots referring to natural environment, has also been made (Tenišev et al. 2006).

A few noteworthy proposals on the depth of Proto-Turkic, i.e., the time of its primal split into the Bulgar and Common Turkic branches, vary from the 5th century BC (Róna-Tas 1998, based on contact linguistics) to the period between 120 BC and the beginning of the first millennium AD (Mudrak 2009, based on glottochronological analysis of Turkic morphology and historical phonology) to the period between the 1st century BC and the 1st century AD (Dybo 2007, based on contact linguistics and lexicostatistics).

The proposals regarding the Proto-Turkic homeland can be seen in the context of the possible Proto-Turkic affiliation with the Xiongnu, a nomadic group that lived north and northwest of China in the first centuries before and after the common

staple foods were different cereals along with horse meat, which may point to a semi-nomadic lifestyle.

era. Several dozen words used by the Xiongnu were recorded in Old Chinese texts such as *Shiji* (or the *Records of the Grand Historian*) and the *Book of Han*, and based on these few words, contemporary scholars have speculated on what language the Xiongnu may have spoken. Various hypotheses were put forward during the 20th century, yet the assumption that the Xiongnu, or at least some of them, were affiliated with Turkic-speaking groups has gained the widest acceptance among scholars (Ramstedt 1922; Basin 1948; Gabain 1949; Šervašidze 1986). This affiliation is based on direct linguistic evidence, i.e., comparing the Xiongnu words in Old Chinese texts with Proto-Turkic, supplemented by historical data that connects the Xiongnu and the subsequent Turkic peoples. Recently, the most reliable Xiongnu words that are comparable with reconstructed Proto-Turkic stems have been outlined by Dybo (2007). Janhunen (2015) also recognizes this affiliation. In short, although we can never exclude that the Xiongnu were a multi-ethnic confederation, it is very likely that their core was Turkic-speaking.²

Different historical and archaeological sources give clues about the subsistence patterns of the Xiongnu. Old Chinese histories (including *Shiji*) emphasize that the Xiongnu were nomadic pastoralists that bred different kinds of domestic ungulates, namely horses, cattle, sheep and camels (Watson 1961). On the other hand, there are multiple indications in Chinese chronicles (including *Shiji*, *Hou Hanshu* (or the *Book of the Later Han*) and notes on the Han annals by Yen Shi-ku) that the Xiongnu were familiar with agriculture, including millet farming (Bičurin 1950; Davydova & Šilov 1953; Davydova 1985). The written sources, however, do not indicate clearly whether it was the Xiongnu themselves or their Chinese captives who were involved in agricultural activities. From an archaeological perspective, although there is about 1000 years of nomadic life in Mongolia beforehand, the Xiongnu period is the first time we have any evidence of agriculture in the region. Agricultural tools and millet grains dating to this period have been found, as well as some isotopic evidence for millet consumption (William Taylor, p.c., Jena, May 2017). It is commonly agreed that the Xiongnu economy was based on pastoralism and had an agricultural component. However, the question of how important the latter was remains open (see Wright et al. 2009; Kradin & Kang 2011; Machicek 2011; Spengler et al. 2016 for further discussion). Given all these observations, it is interesting to examine whether historical linguistic analysis of Turkic subsistence terms can support the association of Proto-Turkic with the Xiongnu.

2. Dybo (2007) shows that the Turkic affiliation is valid, first of all, for the late Xiongnu, while some early “Xiongnu” words may have belonged to an Eastern Iranian (Khotan Saka?) language. There is also a hypothesis by Pulleyblank (1962), which was supported by Vovin (2000, 2002), that the Xiongnu were a Yeniseian-speaking people. An agnostic view of the linguistic affiliation of the Xiongnu is presented in Doerfer (1973).

3. Pastoralist vocabulary in Proto-Turkic

Below I list some of the most relevant Turkic pastoralist terms. To give a more detailed picture, I distinguish between Proto-Turkic and Common Turkic levels. The former label is used when a root occurs in both major subdivisions of the family: the “Standard” Turkic languages, like Turkish, Uyghur, Kazakh etc., and the very specific Bulgar branch, which is represented by its only living language, Chuvash, as well as rather poor lexical data from the extinct Bulgar dialects preserved mainly as loanwords in Hungarian. The label “Common Turkic” means that the word is not attested in Bulgar and hence should be technically attributed to the time after the split of Proto-Turkic. However, due to scarcity of evidence from the Bulgar branch, it is common practice in the field to equate such roots with the Proto-Turkic ones unless a source of borrowing into Turkic has been established.

Table 1. Proto-Turkic pastoralist vocabulary

Semantic group	Proto-Turkic	Common Turkic
goat	* <i>geče</i> (~ <i>geči</i>) ‘(she-)goat’ * <i>teke</i> ‘he-goat’ * <i>oglag</i> ‘kid’	* <i>ečkü</i> ‘(she-)goat’ * <i>erkeč</i> ‘gelded he-goat’
sheep	* <i>sarık</i> ‘sheep’ * <i>Koč</i> ‘ram’ ³ * <i>tokli</i> ‘lamb’	* <i>Koń</i> (~ * <i>Koyn</i>) ‘sheep’ * <i>Koři</i> ‘lamb’

(continued)

3. Here and throughout this paper, capital letters in reconstructed Proto-Turkic forms represent a phoneme the exact characteristics of which are unclear because of a lack of data from relevant Turkic branches. In the case of the capital K and T, the question is whether we should reconstruct a voiced or an unvoiced stop, which are usually distinguished in Oghuz and Sayan reflexes if present (Illič-Svityč 1963; Dybo 2005; Tenišev et al. 2006; see Robbeets 2004 for a different view on the question). For vowels, such as A, what is unclear is whether a short or a long vowel should be reconstructed – an opposition that preserved in Yakut and Turkmenian and can be supported by additional data from the Bulgar and Oghuz branches (Dybo 2007: 52–53).

Table 1. (continued)

Semantic group	Proto-Turkic	Common Turkic	
cattle	* <i>ingek</i> ‘cow’		
	* <i>būka</i> ‘bull’		
	* <i>ökür</i> ‘bull, ox’		
	* <i>dāna</i> ‘(two-years-old) heifer’		
	* <i>buŕa-gu</i> ‘calf’		
		* <i>sigir</i> ‘cattle’ ⁴	
		* <i>ud</i> (~ * <i>od</i>) ‘cattle’ ⁵	
horse	* <i>at</i> ‘(riding) horse’, * <i>adgır</i> ‘stallion’		
	* <i>ulala</i> ‘(small) horse’		
	* <i>elgek</i> ‘donkey’		
	* <i>Kulum</i> ‘foal’		
	* <i>yügen</i> ~ * <i>üygen</i> ‘bridle’		
	* <i>çdner</i> ‘saddle’		
			* <i>beye</i> ‘mare’
			* <i>yunt</i> ‘horse, (mare)’
			* <i>yilkı</i> ‘herd of horses’
			* <i>bün-</i> ~ * <i>bın-</i> ‘to mount a horse’
		* <i>kōlek</i> ‘young of camel’ ⁶	
camel			* <i>debe</i> ‘camel’ ⁷
			* <i>bugu-ra</i> ‘camel stallion’
			* <i>ingen</i> ‘female camel’
			* <i>botu</i> ‘young of camel’
			* <i>tōrum</i> ‘camel colt’
			* <i>Kōm</i> ‘camel’s pack-saddle’

4. Being absent in Chuvash and among the Bulgar borrowings in Hungarian, the root may still be traced back to Proto-Turkic in view of its probable attestation in Danube Bulgar, see Mudrak (2005).

5. Chu. *vil’əχ* < *və^wy-ləχ* ‘cattle’ goes back to PTK **od* and may be compared to CT**ud* ‘ox, bull’, assuming a vowel alternation in Proto-Turkic.

6. The word is reflected in CT**kōšek* ‘young of camel’. Its otherwise unattested Bulgar cognate has been borrowed in Hungarian with a more generic meaning: *kölyök* ‘young of an animal, kid, puppy, lad’ (Róna-Tas & Berta 2011: 586–588). Reconstructing a pastoralist meaning for Proto-Turkic is thus not very reliable.

7. Chu. *tə^wve* ‘camel’ is most probably an early Kypchak borrowing, see Dybo (2010: 58–59).

Table 1. (continued)

Semantic group	Proto-Turkic	Common Turkic
pig	* <i>γAsna-k</i> ~ * <i>γAsna-g</i> ‘pig’ ⁸	* <i>doγur</i> ‘pig’ * <i>čōčka</i> ‘young pig’
dairy	* <i>sag-</i> ‘to milk’ * <i>ayran</i> ‘a k. of salty yoghurt’ * <i>dorak</i> ‘a k. of cheese or quark’	* <i>yogurt</i> ‘curdled milk’ * <i>Katik</i> ‘fermented milk product’ * <i>Kumür</i> ‘alcohol milk drink’ * <i>Kürüt</i> ‘a k. of dried quark, cheese’
technology	* <i>göpe-ne</i> ‘haystack’	* <i>kidir</i> ‘felt’ * <i>aran</i> ‘shed, stable’

As can be seen, Proto-Turkic had a sophisticated system of names for domestic animals (horses, cattle, pigs, goats and sheep), distinguishing age and sex, which is quite typical of a nomadic pastoralist speech community. It should come as no surprise that, in some cases, synonymic names, e.g., for horses, are reconstructed, as they may also have been involved in a kind of semantic distribution. The lack of camel-related vocabulary in the Bulgar branch does not necessarily mean that it was absent in Proto-Turkic, since the Bulgar tribes would have lost the tradition of camel breeding (and hence the related vocabulary) at some point after migrating to Eastern Europe in the first centuries AD. It is also indicative of a pastoralist subsistence strategy that we can reconstruct some pastoralism-related verbs (‘to milk’, ‘to mount a horse’) and a good number of names for dairy products.

Many attempts have been made to explain the Proto-Turkic names for domestic animals as borrowings (often from an Indo-European language, see, e.g., Gamkrelidze & Ivanov 1984), but few of them appear to be plausible. The most widespread view is that some of the Proto-Turkic pastoralist roots originate from an Eastern Iranian language, probably Khotan Saka, cf. pTk **dāna* ‘heifer’ < Khot. *dīnū*, pIr **dainu-kā* ‘cow’ (Bailey 1979: 159; Rastorgueva & Edelman 2003: 447; Dybo 2007: 116–117), pTk **dora-k* ‘a k. of cheese’ < Mİr. **tura-ka*, cf. Av. *tūiri-* ‘curdled

8. The root is preserved only in the Bulgar branch (Chu. *sisna* instead of expected *šisna*, which is a result of late contamination with *sis-* ‘to defecate’) but is very likely to be archaic. With the same meaning, it was borrowed from different Bulgar dialects into Hungarian (*disznó*) and Mari (*sōsna*, *sasna*). No external source for the probable borrowing into Bulgar has been proposed so far.

milk', Khot. (?) *ttūra* 'cheese' (Bailey 1979: 132; Dybo 2007: 117) and, somewhat less reliable due to phonological complications, pTk **ečkü* 'goat' as compared to pIr **aža-* 'goat' (Rastorgueva-Edelman 2000: 292–293; Dybo 2007: 123–124). Beyond that, a Tocharian source has been proposed for pTk **ökür* 'bull, ox', cf. PToch **okso* 'cow, ox' < pIE **uk^wse-* ~ **uk^wso-* (which, however, has been rejected in Doerfer 1963–1975, 1: 539).

As far as genuine Turkic pastoralist terms are concerned, some of them can be easily interpreted as derivatives of a non-agricultural Turkic root, with derivation going back to the Proto-Turkic period. This is, for instance, the case for the following terms:

pTk **ogl-a-g* 'kid', which is traditionally explained as a derivative of pTk **ogul* 'son, child' (Róna-Tas & Berta 2011: 638–642), but differently in (Tenišev et al. 2001: 430), suggesting derivation from **ogla-* 'to shout, to make a racket';

pTk **Kür(i)-t* 'a k. of dried quark, cheese', a common derivative of **Kür(i)-* 'to dry';

pTk **yogurt* 'curdled milk', presumably derived from *yogur-* 'to knead' or a homonymous verb meaning 'to thicken, condense' (Levitskaja et al. 1989).

For almost every root mentioned in this section, etymological parallels in Mongolic, and some in Tungusic, have been proposed previously (see Appendix 1 for supplementary information). Lexical connections between the three branches of Altaic in the domain of pastoralism, with special attention to the distinction between borrowing and inheritance, are further discussed in Section 5.

4. Agricultural vocabulary in Proto-Turkic

It is commonly known that the agricultural component in the Proto-Turkic vocabulary is much smaller than the pastoralist one. Nevertheless, linguistic data clearly show that the Proto-Turkic speakers were familiar with this subsistence pattern as well. The most compelling agricultural terms as reconstructed for Proto-Turkic are the following.

Table 2. Proto-Turkic agricultural vocabulary

Semantic group	Proto-Turkic	Common Turkic
cereals	* <i>darıg</i> ‘corn (millet?)’ * <i>ügür</i> ‘millet’ ⁹ * <i>arpa</i> ‘barley’ * <i>bugday</i> ‘wheat’ ¹⁰	* <i>Konak</i> ‘millet’
grain production	* <i>urug</i> ‘seed’ * <i>ebin</i> ‘grain, (seed)’ *(<i>i</i>) <i>un</i> ‘flour’	* <i>tögi</i> ‘millet groats’ * <i>etmek</i> ‘bread’
pulses	* <i>burçak</i> ‘bean, pea’	* <i>yasmik</i> ‘lentils’
vegetables	* <i>sogan</i> ‘onion’	
tools and technology	* <i>or-</i> ‘to reap, to harvest (a crop)’ > * <i>orlag</i> ‘sickle’ * <i>kētmen</i> ‘hoe, mattock’ * <i>sa(r)pan</i> ‘plough’ * <i>ek-</i> ‘to sow’	* <i>tirmak</i> ‘harrow’ * <i>kerki</i> ‘adze, mattock’ * <i>TArī-</i> ‘to cultivate (ground)’

In Common Turkic, there are several agriculture-related derivatives of a non-agricultural root, e.g., **tög-i* ‘millet groats’ < **tög-* ‘to crush, to husk (e.g. grain)’, **yas-mik* ‘lentils’ < **yas-* ‘to be(come) flat’, **tirma-k* ‘harrow’ < **tirma-* ‘to scratch’. Despite the lack of cognates in the Bulgar branch, it is still possible that some of the derivatives go back to the Proto-Turkic period. Either way, these words cannot be considered as very archaic, but such non-derived verbs as **ek-* ‘to sow’, **or-* ‘to reap, to harvest (a crop)’ and **TArī-* ‘to cultivate (ground)’, as well as the names for cereals, definitely point to a tradition of agriculture in the Proto-Turkic community.

9. Starostin et al. (2003: 1548) reconstruct pTk **yügür* that has the meanings ‘millet’, ‘sorghum’, ‘corn, maize’ and ‘a kind of buckwheat’ across the individual languages. However, it is questionable whether one can bring together forms pointing to an initial **y-*, such as Tat. *jögärä* and Kaz. *žügeri*, and those with an underlying initial vowel (OTurk. *üjür*, Chu. *vir*, etc.). It is interesting that the Turkic forms denoting millet almost never start with a **y-*; this etymology should be kept separately from *y-*forms that denote other crops.

10. Note also the Turkic word for ‘oats’ (Chu. *sə^wlə^w*, Turkm. *süle*, Kaz. *sulī*, *süli* etc.), which, however, demonstrates vowel irregularities and may well be a *Wanderwort* borrowed in different Turkic languages after the family’s split.

It seems essential to discuss in more detail the Turkic names for millet, given the traditionally important role of this crop in the region in question. Three roots, **ügür*, **darıg* and **Konak*, meet formal requirements to be regarded as possible terms for millet in Proto-Turkic. Of them, **ügür* appears to be the most probable candidate for having denoted a kind of millet in the proto-language – it occurs in Chuvash and Yakut, two non-contiguous languages that both separated very early from the main Turkic stock, and it is also attested as ‘millet’ in Old Uyghur texts. Based on the reflexes in the modern Turkic languages, it seems plausible that the Proto-Turkic meaning of the root was ‘broomcorn millet (*Panicum miliaceum*)’. Later on, most Common Turkic languages replaced **ügür* with **darıg* to denote broomcorn millet. In Chuvash, the latter root is represented as *tirə* ‘cereal, corn’, with a more conservative meaning given to the probable derivation of **darıg* from the verb **TArı-* ‘to cultivate (ground)’ (i.e., originally ‘that which is cultivated’). In Common Turkic, one can suggest a semantic development of ‘corn’ → ‘broomcorn millet’, implying that the latter was the primary crop produced by the speakers of Common Turkic. The third root, **Konak*, occurs mainly in Central Asia, particularly in the Karluk branch of Turkic. Its original meaning can be reconstructed as ‘foxtail millet (*Setaria italica*)’ based on the reflexes in modern Turkic languages (along with sporadic ‘sorghum’, ‘maize’ and ‘broomcorn millet’) and Old Uyghur. Despite the old attestation, there is still a question as to whether **Konak* ‘foxtail millet’ can indeed be reconstructed to the time prior to the split of Proto-Turkic, given that there is no trace of the root in the Bulgar branch and in view of its narrow distribution in general (see Appendix 1 for details).¹¹

For all the above terms for cereals, parallels in the other branches of Altaic have been previously proposed. However most of them are rather dubious. For example, pTk **arpa* ‘barley’ is phonologically compatible with pMo **arbai* ‘barley’ and Manchu *arfa* ‘barley, oats’, which was long ago interpreted as a Proto-Altaic root (Ramstedt 1952: 90; Poppe 1960: 87). Alternatively, the Turkic form may be regarded as a loan from an Eastern Iranian reflex of pIr **arbusā* ‘barley’, assuming a subsequent chain borrowing from Turkic to Mongolic and from Mongolic to Manchu. Robbeets (2017: 28) points out that the latter scenario is more consistent with the historical background of barley cultivation in ancient Central and East Asia. Another cereal name of dubious origin is represented by pTk **bugday* ‘wheat’. An Altaic etymology involving pTg **murgi* ‘barley’ has been proposed by Starostin (cited in Dybo 1997), but the correspondence between pTk *-gd- and pTg *-rg- is quite irregular. Róna-Tas and Berta (2011: 188) regard pTk **bugday* as “an old *Kulturword*,” possibly of Indo-European or Chinese origin, but with “no clear

11. An Altaic etymology has been proposed for the root (Starostin et al. 2003: 698), which would consequently confirm its Proto-Turkic status, but the comparison is phonologically problematic.

evidence for either” (see Robbeets 2017: 30–31 for discussion on the connections between pTg **murgi* ‘barley’ and similar forms in Japonic-Koreanic, Indo-European and Old Chinese).

In sum, the originality of the Proto-Turkic terms for broomcorn millet (*Panicum miliaceum*) and foxtail millet (*Setaria italica*) and the equivalence of millet with ‘that which is (generally) cultivated’ contrast with the borrowed nature of the words for ‘barley’ and ‘wheat’. This may indicate that millets were among the original crops cultivated by the speakers of Proto-Turkic.

5. Altaic connections of Proto-Turkic pastoralist vocabulary

An attempt at tracing the Altaic origins of Turkic cultural terms is complicated by the fact that it is easy to confuse cognates with later borrowings because of the intensive contacts between the branches of Altaic. Therefore, it is necessary to place tight constraints when estimating the previously proposed Altaic comparisons that involve evidence from Turkic (see the most comprehensive collection in Starostin et al. 2003). In this regard, I sift out the etymological proposals that seem overly permissive semantically and, on the other hand, apply stricter criteria for phonological correspondences, drawing on the idea of Transeurasian phonology provided in Robbeets (2015). Below I discuss parallels between the main pastoralist terms as reconstructed to Proto-Turkic and Mongolic/Tungusic terms, distinguishing between probable fragments of inherited Proto-Altai lexicon and borrowings.

As far as the Turkic pastoralist vocabulary is concerned, there is a remarkable group of meanings that falls in part within the restrictions and appears to have reliable Altaic parallels, namely, terms for bovine and equine domestic animals. See for example the following matches:¹²

pTk **beye* ‘mare’ < pA **bej-* ‘a k. of ungulate animal’ > Tung. **bejū-* ‘an ungulate animal’;

pTk **Kulum* ‘foal’ < pA **kul-* ‘a k. of small equine’ > pMo **kulan* ‘donkey’;

pTk **sigir* ‘cattle’, cf. pTk **sigun* ‘(male) deer’ < pA **sig-* ‘deer, horned ungulate’ > pTg **sig-* ~ **seg-* ‘wild deer’, ? pMo **siyenek* ~ **seyenek* ‘(2-years-old) he-goat’;

pTk **būka* ‘bull, ox’ < pA **muxa-* ‘male’ > pTg **muxa-* ‘man; male’.

12. For some of the roots presented here and elsewhere in the paper, etymological matches in the other branches of Transeurasian have been previously proposed, but I do not quote them here because of their unreliability.

These examples, excluding the last one, present the interesting semantic development of ‘wild animal’ → ‘domestic animal’. We can assume that this change reflects a shift in subsistence patterns from Proto-Altaic to Proto-Turkic, resulting in the adaptation of hunting terms for the needs of a pastoralist society. It is notable in this respect that agricultural societies in North East China that can be associated with Proto-Altaic, such as the Hongshan, produced millet, but they obtained their protein sources from hunting in the wild (Nelson 1994).

Another question relates to the possible inheritance from Proto-Altaic to Proto-Turkic in the realm of animal husbandry. Under the approach I described above, almost all the Altaic comparisons referring to this field appear to fail on formal grounds. In fact, the only reliable case where borrowing does not appear to be the most likely explanation as compared to inheritance is the parallel between pTk **tōrum* ‘young camel (or calf or goatling)’, pMo **toruy* ‘young pig’ (but Ord. *torō* ‘young donkey’) and pTg **tora-kī* ‘boar (male of a pig)’. In this comparison, phonological correspondences are perfect, and the fact that none of these forms are morphologically identical serves as additional evidence for inheritance rather than borrowing, especially since the Turkic word is indeed borrowed in Mongolic as WMo. *torum*, Kalm. *torm* ‘young camel’. Based on Mongolic and Tungusic, the original meaning ‘pig’ can be reconstructed to Proto-Altaic, implying a shift to ‘camel’, but also to ‘goat’ and ‘calf’ in the Turkic branch. Interestingly, domestic pigs are found along with dogs in early farmer sites in North East China as early as 6000 BC (Larson et al. 2010).

An additional interesting match may correspond to a period after the split of Proto-Altaic, as it involves only the Turkic and Mongolic branches: pTk **sag-* and pMo **saya-*, both meaning ‘to milk’.¹³ Inheritance is more likely than borrowing in this case, given the relatively low borrowability of bare verb roots and the typology of verbal borrowing across the Transeurasian languages, which involves formal accommodation rather than direct insertion (Robbeets 2015). Thus **sag-* ‘to milk’ may be reconstructed to Proto-Turko-Mongolic (4600–2800 BC).

Many pastoralist terms shared by Turkic and Mongolic are universally accepted (and relatively late) Common Turkic loans in Mongolic, e.g., Turk. *teke* > Mong. *teke* ‘he-goat’, Turk. *buqa* > Mong. *buqa* ‘bull’, Turk. *buyura* > Mong. *buyura* ‘camel stallion’, Turk. *torum* > Mong. *torum* ‘camel colt’. For its part, Mongolic donated a great deal of its pastoralist terms to Tungusic (see, e.g., Rozycki 1994).

However, there are also a number of Turko-Mongolic parallels in pastoralist vocabularies that are traditionally considered as cognates in Altaic studies but

13. Tung. **saji-ža* (~-g-) ‘sieve’, which is proposed in (Starostin et al. 2003:1198) as a cognate for the Turko-Mongolic comparison, cannot be regarded as reliable in view of very different semantics.

cannot be regarded as reliable due to irregular phonology and are more likely to be early borrowings, probably from Proto-Turkic (Pre-Proto-Bulgar, according to Janhunen and some other authors) into Proto-Mongolic. For example, this could be the case for the following roots:

pTk **buŕa-gu* ‘calf’ > pMo **biragu* ‘id.’;

pTk **Koč* ‘ram’ > pMo **kuča* ‘id.’;

pTk **Kofī* ‘lamb’ > pMo **kurigan* ‘id.’;

pTk **Kumīr* ‘alcohol milk drink’ > pMo **kimur* ‘fermented milk with water’;

pTk **ayran* ‘a k. of salty yoghurt’ > pMo **ayirag* ‘id.’;

pTk **ingek* ‘cow’ > pMo **ūniyen* ‘id.’;

pTk **ökür* ‘bull, ox’ > pMo **(h)üker* ‘id.’

The above examples can be compared to the following pairs where phonological irregularities are supplemented by an unexplainable difference in syllable structures: pTk **tokli* ‘lamb’ > pMo **tugul* ‘calf’, pTk **sarik* ‘sheep’ > pMo **serke* ‘gelded goat’. Occasionally, it is morphological evidence that suggests borrowing, cf. pTk **koñ* (~ **Koyñ*) ‘sheep’ and pMo **koni-n* with an unstable *n* that may originally have functioned as a “class” marker (Janhunen 2012).

A rather difficult case is the parallel between pTk **elġek* ‘donkey’ and pMo **elžigen* ‘id.’. It demonstrates the non-trivial correspondence pTk **ġ* ~ pMo **lž*, which is characteristic of Proto-Altaiic. However, the contact scenario is more likely (see Rozycki 1994: 67, involving Manchu *eihen* ‘id.’ as part of the borrowing chain, probably from Turkic into Mongolic and from Mongolic into Tungusic, and recent discussion in Parpola & Janhunen 2011: 90–94). According to Chinese historical records, domestic donkeys could be found, though quite rarely, in northern China around 2000 BC, but no evidence allows them to be traced back to an earlier period (Han et al. 2014). One more noteworthy comparison is between pTk **at* ‘horse’ to pMo **aduyu* ‘id.’. Although it is technically possible to reconstruct pA **at-* ‘horse’, the unexplainable segmentation of the Mongolic form is indicative of borrowing in this case, perhaps from a morphologically complex Turkic form. Archaeological evidence indicates that horses did not appear in the Western Liao river valley until the Lower Xiajiadian period (2000–1500 BC), which is at least 1000 years later than the Hongshan period (see Robbeets 2017: 32 for the horse in East Asia and the borrowing of another horse term). Given this, it is still preferable to attribute the lexical parallel to a later contact between the branches.

To sum up, most pastoralism-related terms in Proto-Turkic seem to be of secondary origin. Some of them are transmitted as loanwords from a non-Transeurasian language or developed through internal derivation as shown in Section 3. In

other cases, they can be shown to have developed from a term for the original wild predecessor in Proto-Altaic (e.g., ‘deer’ > ‘cattle’). The only reliable case where the term for a domestic animal in Turkic goes back to such a term in Proto-Altaic is **tor(u)-* ‘pig’. It is striking that ‘pig’ is the only name for a domestic animal that can be reliably reconstructed to Proto-Altaic, as it is an animal that is associated with the agricultural societies in Northeast Asia and not with nomadic pastoralism. All this evidence seems to suggest that the Turkic people shifted from a subsistence pattern involving pig raising, millet cultivation and wild animal hunting to a pattern based on horse-riding pastoralism.

6. Altaic connections for Proto-Turkic agricultural vocabulary

Compared to the Proto-Turkic pastoralist lexicon, its agricultural vocabulary is limited and, consequently, one would not expect to find many such terms derived from Proto-Altaic. Yet, a few interesting correlations are worth discussing.

The only plausible parallel that is present in all three branches of Altaic is represented by pTk **TArī-* ‘to cultivate (land)’, pMo **tari-* ‘to sow, to plant, to plough’ and pTg **tari-* ‘to cultivate’. It is often thought that the Turkic word was borrowed into Written Mongolic as *tari-*, from which it entered Tungusic, i.e., Evk. *tari-* ~ *tare-*, Solon *tari-*, Manchu *tari-*, Nanai *tari-*, Ulcha *tari-* ‘id.’ (Doerfer 1963: 244–245; Rozycki 1994: 203). However, it can be argued that this is in fact a Proto-Altaic agricultural term (pA **tari-* ‘to cultivate land’). In addition to the arguments mentioned for **saq-* ‘to milk’ in Section 5, a chain borrowing scenario for a naked verb root is cross-linguistically rather uncommon (Robbeets 2015). The inherited status of the root can be further supported by the fact that the representations of **tari-* in each family are involved in productive derivational processes (cf. such derivatives as pTk **dārīg* ‘corn’ > ‘millet’, pMo *tariyan* ‘crops’ and Evk. *tariyan* ‘bread’).

A less striking comparison involves pTk **or-* ‘to reap, harvest, mow’ and pTg **oro-kta* ‘(dry) grass, hay’ (Starostin et al. 2003: 1063–1064), where **-kta* is a collective suffix. The correlation would be more direct if we assume that the Tungusic form is of verbal origin (**oro-* ‘to graze, pasture, mow?’), cf. maybe pTg **oro-n*, PL. *oro-r* ‘domesticated reindeer’. Even if the hypothesis on Altaic connections does not stand up to scrutiny, it is still interesting that the Turkic verb for harvesting has a very simple morphological structure and does not appear to be derived or borrowed. For a similar case, one can look to pTk **ek-* ‘to sow’, which has no reliable Altaic connections established, but must have belonged to non-derived and non-borrowed lexicon of Proto-Turkic. It is also telling that the main Turkic names for millets, **ügür* ‘broomcorn millet’ and **Konak* ‘foxtail millet’, have a quite different historical background as compared to those for other cereals. While **arpa*

‘barley’ and **bugday* ‘wheat’ are often regarded as wanderwörter, there are no clear indications that the Turkic names for millets were borrowed from outside. Moreover, **Konak* itself may have been borrowed into Written Mongolian as *qonuy* ‘millet’ (Starostin et al. 2003: 698).

As for the other agricultural terms in Proto-Turkic, few of them can be reliably connected to the other branches of Altaic. Even look-alikes that appeared as a result of early borrowing are much less numerous in the field of agriculture as compared to pastoralism. A rare reliable example of such borrowing is the case of pTk **burčak* ‘bean, pea’ and pMo **buγurčag* ‘id.’. The forms are undeniably related, but they hardly can be explained in terms of genetic affinities. Thus borrowing (possibly from Mongolic to Turkic, given that the Mongolic form is more complex) is very likely. This can be compared to the parallel between pTk **sogan* ‘onion’ and pMg **songina* ‘id.’, where the exact direction of borrowing, probably involving other East Asian languages, is unclear (Starostin et al. 2003: 1303).

In some cases, such as that represented by the parallel between pTk **urug* ‘seed’ and pMg **(h)üre* ‘id.’, the difference between the Turkic and Mongolic form is such that the resemblance may just be coincidental.

To summarize, I have investigated the origin of Proto-Turkic agricultural and pastoralist vocabularies. While there are indications that the majority of the Turkic pastoralist vocabulary is internally coined, borrowed from a non-Transeurasian language, inherited from names for wild predecessors or fragments of agricultural vocabulary, I found less indications for the secondary nature (i.e., borrowing, derivation or lexical recycling) of agricultural terms, such as ‘millet’. Basic agricultural activities, such as ‘to harvest’, ‘to sow’ and ‘to cultivate’ also seem to be unborrowed and underived. Except for the verb ‘to cultivate’, the word for ‘pig’ (see Section 5) and a vague connection for ‘to harvest’, I did not reveal reliable Altaic connections for Turkic agricultural words. However, agricultural core-vocabulary seems to preserve more Altaic cognates than the lexicon of pastoralism does, although the latter is far better represented in Turkic. Further, the Turkic pastoralist vocabulary has a more secondary nature than the agricultural one. In general, the very limited number of agricultural terms reconstructible to Proto-Altaiic as compared to the preceding Proto-Transeurasian period (see Robbeets 2017; this volume) can be attributed to a loss of farming-related lexicon in the daughter languages over time after the break of Altaic; they may have lost the words along with the tradition after climate change and shift to pastoralism.

7. Conclusions

In this study, I have provided a historical linguistic discussion of the subsistence-related activities that can be assigned to the Proto-Turkic speakers and to their Proto-Altaiic predecessors. I established that, along with a rich and complex pastoralist vocabulary, a number of agricultural terms can also be reconstructed to Proto-Turkic. The Turkic names for ‘barley’ and possibly ‘wheat’ may be borrowings in Proto-Turkic, but millet seems to be very prominent given that it is referred to as “that what is cultivated (= the main crop)”. It is likely that two kinds of millet, broomcorn and foxtail, were distinguished linguistically by the speakers of Proto-Turkic. The Proto-Turkic agricultural vocabulary also includes terms for such basic activities as ‘to sow’, ‘to harvest’ and ‘to cultivate’, and all seem to be archaic.

This study can support the identification of Proto-Turkic with the Xiongnu, as the proportion of pastoralist to agricultural terms in Proto-Turkic is consistent with what we know about the agricultural component in the Xiongnu archaeological record.

Subsistence-related terms in Proto-Turkic differ in their origins. Some of them are borrowed from a non-Transeurasian language, such as pTk **dāna* ‘heifer’ and **arba* ‘barley’, and some are internally coined.

Both pastoralist and agricultural vocabularies in Proto-Turkic are in part similar to those in Mongolic and Tungusic languages. However, while the similarities between the pastoralist terms are almost exclusively due to borrowing, agricultural vocabularies of the branches seem to share a few items inherited from Proto-Altaiic. In most cases, it was possible to distinguish between borrowing and inheritance due to linguistic indications, such as phonological and semantic differences, morphological complexity in one language but not in the other, etc. In general, we found no Altaiic reconstructions pointing to pastoralism in the Proto-Altaiic period, while a few Proto-Altaiic etymologies are reconcilable with an agricultural lifestyle.

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 programme (grant agreement No 646612), granted to Martine Robbeets.

Abbreviations

Chu	Chuvash	pJ	Proto-Japonic
CT	Common Turkic	pK	Proto-Koreanic
Kalm.	Kalmyk	pMo	Proto-Mongolic
Kaz.	Kazakh	pTk	Proto-Turkic
Khot.	Khotan Saka	pToch.	Proto-Tocharian
K	Korean	pTg	Proto-Tungusic
Ord.	Ordos Mongolian	Tkm.	Turkmenian
pA	Proto-Altaiic	Uig.	Uighur
pIE	Proto-Indo-European	WMo.	Written Mongolian
pIr	Proto-Iranian		

References

- Bailey, Harold W. 1979. *Dictionary of Khotan Saka*. Cambridge: CUP.
- Bazin, Louis. 1948. Un texte proto-turc de 4e siècle. *Oriens* 1: 208–219.
- Bičurin, Nikita (Iakinf). 1950. *Sobranie svedenij o narodah, obitavših v Srednej Azii v drevnie vremena*, Vol. 1. Moskva-Leningrad.
- Clauson, Gerard. 1956. The case against the Altaic theory. *Central Asiatic Journal* 2: 181–187.
- Davydova, Antonina & Šilov, Valentin. 1953. K voprosu o zemledelii u gunnov. *Vestnik drevnej istorii* 2: 193–195.
- Davydova, Antonina. 1985. *Ivolginskij kompleks (gorodišče i mogil'nik): Pamjatnik hunnu v Zabajkalje*. Leningrad: Izd-vo LGU.
- Doerfer, Gerhard. 1963–1975. *Türkische und mongolische Elemente im Neupersischen* (Vols. 1–3). Wiesbaden: Franz Steiner.
- Doerfer, Gerhard. 1973. Zur Sprache der Hunnen. *Central Asiatic Journal* 1: 1–50.
- Dybo, Anna. 1997. K kul'turnoj leksike praaltajskogo jazyka. *Balto-Slavjanskije issledovanja* 1988–1996: 164–177.
- Dybo, Anna. 2005. Dental'nye vzryvnye v prattjurkskom. *Aspekty komparativistiki* 1: 49–82.
- Dybo, Anna V. 2007. *Lingvističeskie kontakty rannih tjurkov. Leksičeskij fond. Prattjurkskij period*. Moskva: Vostočnaja literatura.
- Dybo, Anna V. 2010. Vokalizm rannetjurkskih zaimstvovanij v vengerskom. In *Gedenk von E.A. Helmsky. Finnisch-Ugrische Mitteilungen* 32–33: 83–132.
- von Gabain, Annemarie. 1949. Review of Louis Bazin “Un texte proto-turc de 4e siècle”. *Der Islam* 20: 244–246.
- Gamkrelidze, Tamaz V. & Ivanov, Vjačeslav V. 1984. *Indoevropskij jazyk i indoevropcy: Rekonstrukcija i istoriko-tipologičeskij analiz prajazyka i protokul'tury*. Tbilisi: Izd-vo Tbilisskogo un-ta.
- Golden, Peter B. *An Introduction to the History of the Turkic Peoples: Ethnogenesis and State-formation in Medieval and Early Modern Eurasia and the Middle East*. Wiesbaden: Harrassowitz, 1992.
- Guo, Da-Shun. 1995. Hongshan and related cultures. In *The Archaeology of Northeast China. Beyond the Great Wall*, Sarah Nelson (ed.), 21–64. London: Routledge.

- Han, Lu *et al.* 2014. Ancient DNA provides new insight into the maternal lineages and domestication of Chinese donkeys. *BMC Evolutionary Biology* 14: 246. doi:10.1186/s12862-014-0246-4
- Illič-Svityč, Vladislav. 1963. Altajskie dental'nye: *t, d, δ*. *Voprosy jazykoznanija* 6: 37–56.
- 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 *Recent Advances in Tungusic Linguistics*, Andrej L. Malchukov & Lindsay J. Whaley (eds), 5–16. Wiesbaden: Harrassowitz.
- Janhunen, Juha. 2015. Reconstructing the language map of prehistorical Northeast Asia. *Studia Orientalia Electronica* 108: 281–304.
- Janz, Lisa, Odsuren, D. & Bukhchuluun, D. 2017. Transitions in palaeoecology and technology: Hunter-gatherers and early herders in the Gobi desert. *Journal of World Prehistory* 30(1). doi:10.1007/s10963-016-9100-5.
- Johanson, Lars & Robbeets, Martine (eds). 2010. *Transeurasian Verbal Morphology in a Comparative Perspective: Genealogy, Contact, Chance*. Wiesbaden: Harrassowitz.
- Kljaštornyj, Sergej & Sultanov, Tursun. 2009. *Gosudarstva i narody Evrajskikh stepej: Ot drevnosti k Novomu vremeni*. Sankt-Peterburg: Peterburgskoe Vostokovedenie.
- Kradin, Nikolaj N. & Kang Inuk. 2011. Sovremennye issledovanija po arheologii hunnu v Evrazii. *Russian Studies*: 309–349.
- Lane, George. 2006. *Daily Life in the Mongol Empire*. Westport CT: Greenwood Press.
- Larson, Greger *et al.* 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Science* 107(17): 7686–7691. doi:10.1073/pnas.0912264107
- Machicek, Michelle L. 2011. *Reconstructing diet, health and activity patterns in early nomadic pastoralist communities of Inner Asia*. Ph.D. dissertation, University of Sheffield, Sheffield.
- Menges, Karl H. 1975. *Altajische Studien, II. Japanisch und Altajisch. Abhandlungen für die Kunde des Morgenlandes* 41(3).
- Menges, Karl H. 1984. *Korean and Altaic. A Preliminary Sketch* [Central Asian Journal 28].
- Menges, Karl H. 1995. *The Turkic Languages and Peoples: An Introduction to Turkic Studies*, 2nd edn. Wiesbaden: Harrassowitz.
- Miller, Roy A. 1996. *Languages and History. Japanese, Korean and Altaic*. Bangkok: White Orchid & Oslo: The Institute for Comparative Research in Human Culture.
- Mudrak, Oleg A. 1993. *Istoričeskie sootvetstvija čuvaških i tjurkskih glasnyh. Opyt rekonstrukcii i interpretacii*. Moskva: DIAMOND.
- Mudrak, Oleg A. 2005. Zametki o jazyke i kul'ture dunajskih bulgar. *Aspekty komparativistiki* I: 83–106.
- Mudrak, Oleg A. 2009. *Klassifikacija tjurkskih jazykov i dialektov s pomoščju metodov glototihronologii na osnove voprosov po morfologii i istoričeskoj fonetike*. Moskva: RGGU.
- Nelson, Sarah. 1994. The development of complexity in prehistoric North China. *Sino-Platonic Papers* 63: 1–17.
- Nelson, Sara. 2001. Hongshan. In *Encyclopedia of Prehistory, Vol. 3: East Asia and Oceania*, Peter N. Peregrine & Melvin Ember (eds), 77–81. Berlin: Springer. doi:10.1007/978-1-4615-1189-2_14
- Parpola, Asko & Janhunen, Juha. 2011. On the Asiatic wild asses and their vernacular names. In *Linguistics, Archaeology and the Human Past* [Occasional paper 12], Toshiki Osada & Hitoshi Endo (eds), 59–124. Kyoto: Indus Project, Research Institute for Humanity and Nature.
- Poppe, Nicholas N. 1960. *Vergleichende Grammatik der Altaischen Sprachen, Teil 1: Vergleichende Lautlehre*. Wiesbaden: Harrassowitz.

- Pulleyblank, Edwin G. 1962. The consonantal system of Old Chinese. *Asia Minor* 9: 58–144, 206–265.
- Ramstedt, Gustaf J. 1922. Zur Frage nach der Stellung des Tschuwassischen. *JSFOu* 38 1: 1–34.
- Ramstedt, Gustaf J. 1952. *Einführung in die Altaische Sprachwissenschaft, 1: Lautlehre* [Mémoires de la Société Finno-Ougrienne 104]. Helsinki: Suomalais-Ugrilainen Seura.
- Rastorgueva, Vera S. & Edel'man, Džoj I. 2000, 2003. *Etimologičeskij slovar' iranskih jazykov. T. 1 (a –ā). T. 2 (b –d)*. Moskva: Vostočnaja literatura.
- Robbeets, Martine. 2004. Does Doerfer's Zufall mean 'cognate'? The case of the initial velar correspondence in Altaic. *Turkic Languages* 8: 146–178.
- Robbeets, Martine. 2005. *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* [Turcologica 64]. Wiesbaden: Harrassowitz.
- Robbeets, Martine. 2015. *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages* [Trends in Linguistics. Studies and Monographs 291]. Berlin: Mouton de Gruyter. doi:10.1515/9783110399943
- Robbeets, Martine. 2017. Proto-Trans-Eurasian: Where and when? *Man In India* 97 (1): 19–46.
- Róna-Tas, András. 1998. Reconstruction of Proto-Turkic and the genetic question. In *The Turkic Languages*, Lars Johanson & Eva Csató (eds), 67–80. London: Routledge/Curzon.
- Róna-Tas, András & Berta, Árpád. 2011. *West Old Turkic. Turkic Loanwords in Hungarian* [Turcologica 84]. Wiesbaden: Harrassowitz.
- Rozycki, William. 1994. *Mongol Elements in Manchu*. Bloomington IN: Research Institute for Inner Asian Studies.
- Sanžeev, Garma D., Orlovskaja, Maria N. & Ševernina, Zoja V. 2015. *Etimologičeskij slovar' mongol'skih jazykov, Tom I: A–E*. Moskva: IV RAN.
- Šervašidze, Ivan. 1986. *Formy glagola v jazyke tjurkskih runičeskij nadpisej*. Tbilisi: Mecniereba.
- Sevortjan, Ervand V. et al. (eds). 1974–2003. *Etimologičeskij slovar' tjurkskih jazykov* (Vols. 1–7). Moskva: Nauka.
- Spengler, Robert N., Ryabogina, Natalia, Tarasov, Pavel E. & Wagner, Mayke. 2016. The spread of agriculture into northern Central Asia: Timing, pathways, and environmental feedbacks. *The Holocene*: 1–14. doi:10.1177/09596836166641739
- Starostin, Sergej A. 2008. Altaic loans in Old Chinese. In *In Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics* [Routledge Studies in the Early History of Languages 5], Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin (eds), 254–262. London: Routledge.
- Starostin, Sergej A., Dybo, Anna V. & Mudrak, Oleg A. 2003. *Etymological Dictionary of the Altaic Languages*. Leiden: Brill.
- Taylor, William T. T. et al. 2017. A Bayesian chronology for early domestic horse use in the Eastern Steppe. *Journal of Archaeological Science* 81: 49–58. doi:10.1016/j.jas.2017.03.006
- Tenishev, Edhjam R. et al. 2001. *Sravnitel'no-istoričeskaja grammatika tjurkskih jazykov. Leksika*, 2nd edn. Moskva: Nauka.
- Tenishev, Edhjam R. et al. 2006. *Sravnitel'no-istoričeskaja grammatika tjurkskih jazykov. Prattjurkskij jazyk-osnova. Kartina mira prattjurkskogo etnosa po dannym jazyka*. Moskva: Nauka.
- Tsintsius, Vera I. (ed.). 1975. *Sravnitel'nyj slovar' tunguso-mančžurskij jazykov*. Leningrad: Nauka.
- Turaev, Vadim A. et al. 1997. *Istorija i kul'tura evenov. Istoriko-etnografičeskije očerki*. Sankt-Peterburg: Nauka.
- Turaev, Vadim A. et al. 2001. *Istorija i kul'tura oročej. Istoriko-etnografičeskije očerki*. Sankt-Peterburg: Nauka.

- Turaev, Vadim A. *et al.* 2003. *Istorija i kul'tura nanajcev. Istoriko-etnografičeskie očerki*. Sankt-Peterburg: Nauka.
- Vovin, Alexander. 2000. Did the Xiongnu speak a Yeniseian language? *Central Asiatic Journal* 44(1): 87–104.
- Vovin, Alexander. 2002. Did the Xiongnu speak a Yeniseian language? Part 2: Vocabulary. In *Altaica Budapestinensia MMII, Proceedings of the 45th Permanent International Altaistic Conference, Budapest, June 23–28*, 389–394.
- Vovin, Alexander. 2005. The end of the Altaic controversy: In memory of Gerhard Doerfer. *Central Asiatic Journal* 49 (1): 71–132.
- Watson, Burton (trans.). 1961. *Records of the Grand Historian of China. Translated from the Shiji, by Sima Qian*, 2 Vols. New York NY: Columbia University Press.
- Wright, Joshua, Honeychurch, William & Amartuvshin, Chunag. 2009. The Xiongnu settlements of Egiin Gol, Mongolia. *Antiquity* 83: 372–387. doi:10.1017/S0003598X00098495

Appendix 1. Forms underlying the reconstructed Proto-Turkic roots and their Altaic connections

**aran* ‘shed, stable’

Karakh. *aran* 1 (MK); Turkm. *aram* (dial.) 3; MTurk. *aran* 1 (Sangl.); Krm. *aran* 3; Tat. *aran* 1; Bashk. *aran* 1, Kaz. *aran* 3; Kum. *aran* 1; Nogh. *aran* 3; Yak. *araņas* (< **aran-gač*, Dimin.) 4, dial. *arān* ‘место, на котором стоит чум, палатка’; Dolg. *araņas* 4 (Starostin *et al.* 2003: 1123–1124).
‘shed 1, stable 2, fold 3, store-room 4’

**arpa* ‘barley’

OTurk. *arpa* (OUygh.), *abra* (late OUygh.); Karakh. *arpa* (MK, KB); Tur. *arpa*; Gag. *arpa*; Az. *arpa*; Turkm. *arpa*; Sal. *arfa* (CCЯ 292); Khal. *arpa*; MTurk. *arpa* (Sangl.); Uzb. *arpa*; Uig. *a(r)pa*; Krm. *arpa*; Tat. *arpa*; Bashk. *arpa*; Kirgh. *arpa*; Kaz. *arpa*; KBalk. *arpa*; KKalp. *arpa*; Kum. *arpa*; Nogh. *arpa*; Khak. *arba*; Oyr. *arba*; Chu. *orba*; Bulg. > Hung. *árpa* (Starostin *et al.* 2003: 313).

Probably an IE loanword, see Robbeets 2017. Cf. PMo **arbay* ‘barley’: MMong. *arbāi* (HY 8), *arbāi*, *ārbāi* (MA 104, 253); WMo. *arbai* (L 49); Kh. *arvay*; Bur. *arbay*; Kalm. *arwā*, *arwā*; Ord. *arwā*; Mog. *arfei*, *arfā*; Dong. *apa*; Ma. *arfa* ‘barley; oats’; OJap. *apa* ‘millet’.

**at* ‘(riding) horse’

OTurk. *at* (Orkh., Yen., OUygh.); Karakh. *at* (MK, KB); Tur. *at*; Gag. *at*; Az. *at*; Turkm. *at*; Sal. *at*, *ac*; Khal. *hat*; MTurk. *at*; Uzb. *ot*; Uig. *at*; Krm. *at*; Tat. *at*; Bashk. *at*; Kirgh. *at*; Kaz. *at*; KBalk. *at*; KKalp. *at*; Kum. *at*; Nogh. *at*; SUig. *a't*; Khak. *at*; Shr. *at*; Oyr. *at*; Tv. *a't*; Chu. *ut*; Yak. *at*; Dolg. *at* (Starostin *et al.* 2003: 317).

Probably a derivative of **at* is represented by **adg̃ir* ‘stallion’: OTurk. *adyir*; Karakh. *adyir*, *ayyir*; Chag. *ayyir*; Kirgh. *ayyir*; Alt. *ayyir*; Uzb. *ayyir*; Uigh. *ayyir*; S.-Yugh. *azyir*; Khak. *asxir*; MChul. *asqir*; Tuv. *asqir*; Tof. *asqir*; Yak. *at̃ir*; Dolg. *at̃ir*; Chu. *aỹer* (Tenišev *et al.* 2001: 442–443).

Cf. PMo **aduyu-* > MMong. *adusun* ‘horse(s)’, etc. (possibly < Turkic).

*ayran ‘a k. of salty yoghurt’

Karakh. *ayran* (MK); Tur. *ayran*; Az. *ayran*; Turkm. *ayran*; Uzb. *ayran*; Uig. *ayran*; Krm. *ayran*; Tat. *eyren*; Bashk. *ayran*; Kirgh. *ayran*; Kaz. *ayran*; KBalk. *ayran*; KKalp. *ayran*; Kum. *ayran*; Nogh. *ayran*; Khak. *ayran*; Oyr. *ayran*; Chu. *uyran*, dial. *uʻan*, *oren* (Anatri) (Starostin et al. 2003: 280).

Cf. PMo **ayirag* ‘id.’: MMong. *aiyirax* (HY 25); WMo. *ayiray* (L 21); Kh. *ayrag*; Bur. *ayrag*; Kalm. *āræg*; Ord. *āraq*; Dag. *airag* (possibly < Turkic).

*beye ‘mare’

OTurk. *be* (OUig. – YB); Karakh. *be* (MK, IM); MTurk. *beye* (Sangl.); *bej* (CCum.); Uzb. *biya*; Uig. *biya* (dial.); Krm. *biye*; Tat. *biye*; Bashk. *beya*; Kirgh. *bē*; Kaz. *biye*; KKalp. *biye*; Nogh. *biye*; SUig. *pie*, *pi*; Khak. *pī*; Oyr. *bē*; Tv. *be*; Yak. *bia* (Starostin et al. 2003: 335–336).

*botu ‘young of camel’

Karakh. *botu* (MK); Tur. *potak* (dial.); Az. *pota* ‘young of buffalo, bear’; MTurk. *bota* ‘child; young of animal’ (Abush., Sangl.); Uzb. *boʻta*; Uig. *bota*; Krm. *bota*; Tat. *buta*; Bashk. *buta*; Kirgh. *boto*; Kaz. *bota*; KKalp. *bota*; Nogh. *bota* (Starostin et al. 2003: 901–902).

*bugday ‘wheat’

OUig. *buyday*; Karakh. *buyday*; Chag. *buyday*; Tur. *buyday*; Turkm. *buyday*; Gag. *bo:day*; Az. *buyda*; Khal. *buyda*; Sal. *boyde*, *pojde*, *poyce*, *pohtar*; Kar. *boyday*, *buday*; KBalk. *buday*; Kum. *buday*; Tat. *boday*; Bashk. *boyday*; Nogh. *biyday*; KKalp. *biyday*; Kaz. *biyday*; Kirgh. *bu:day*; Oyr. *pu:day*; Uzb. *buydoy*; Uigh. *buyday*; Khak. *puyday*; Chul. *puday*; ? Chu. *pōri* ‘smelt’.

A Wanderwort of unclear origin (Róna-Tas and Berta 2011: 188).

*bugu ‘deer male’ > *bugu-ra ‘camel stallion’

OTurk. *buyu* 1 (13th c.), *buyura* 2 (Orkh.); Karakh. *buyra* 2 (MK); Tur. *buyur* 2, dial. *buyu* 1; Az. *buyur* 2; Turkm. *buyra* 2; MTurk. *buyu* 1, *buyra*, *buyur* 2 (Pav. C.); Uzb. *buyu* 1; Uig. *buyu* 1, (dial.) *buyra*, *boyra* 2; Kirgh. *būra* 2; Kaz. *burā* 2; KBalk. *bū* 1; KKalp. *buwra* 2; Nogh. *bora* 2; SUig. *pūrya* 2; Oyr. *burā* 2; Tv. *būra* 2, *būr* ‘male elk’; Yak. *būr* ‘male reindeer, male’; Dolg. *būr* ‘male reindeer’ (Starostin et al. 2003: 1102).

‘deer male 1, camel stallion 2’

*būka ‘bull’

OTurk. *buqa* (Orkh., OUygh.); Karakh. *buqa* (MK, KB); Tur. *boa*; Gag. *buya*, *buā*; Az. *buGa*; Turkm. *buGa*; MTurk. *buya* (Sangl.); Uzb. *buqa*; Uig. *buya*, *buqa*; Krm. *buya*; Tat. *buya* (dial.); Bashk. *buya*; Kirgh. *buqa*; Kaz. *buqa*; KBalk. *buya*; KKalp. *buya*; Kum. *buya*; Nogh. *buya*; SUig. *puqa*; Khak. *puya*; Shr. *puya*; Oyr. *buqa*; Tv. *buya*; Tof. *buxa*; Yak. *buga* (Starostin et al. 2003: 951).

Probably a Proto-Altaiic root, cf. pTg **muxa-* / **muxe-* ‘man 1, male 2’: Neg. *muxeti* 2; Man. *muḡan* 2; Nan. *moḡa(n)* 1, 2; Orch. *mueti* 2; Ud. *mugeti*, *mueti* 2.

*buʻa-gu ‘calf’

OTurk. *buzayu* (OUygh.); Karakh. *buzayu* (MK, IM); Tur. *buzayu*; dial. *buza-* ‘to bear a calf’; Osm. *buza-la-* ‘id.’; Gag. *buzā*; Az. *bizov*; Turkm. *buzaw*; Sal. *puzo*, *pūzi* (CCЯ 457); MTurk. *buzayu*, *buzay*, *buzaw* (Sangl., MA, Pav. C.); Uzb. *buzoq*; Uig. *mozay*; Krm. *bizuv*, *buzuv*; Tat. *bizaw*; Bashk. *biḏaw*; Kirgh. *muzō*; Kaz. *buzau*; KBalk. *buzow*; KKalp. *buzaw*; Kum. *buzaw*; Nogh. *buzaw*; Khak. *pizo*; Shr. *puza* (R); Oyr. *bīza*; Tv. *bizā*; Chu. *pōʻru* (Starostin et al. 2003: 353–354).

Cf. PMo **birayu* ‘calf (1 year old)’: MMong. *burā'u* (SH), *buru* (MA); WMo. *birayu* (L 106); Kh. *barū*; Bur. *burū*; Kalm. *burū*; Ord. *birū* ‘calf (2 year old)'; Mog. ZM *borɔyol* (20–8), KT *borwol* (20–6); Mongr. *burū* (SM 36) (probably < Turkic).

**burčak* ‘bean, pea’

OTurk. *burčaq* (OUygh.); Karakh. *burčaq* (MK); Tur. *burčak*; Gag. *borčaq*; Turkm. *burčaq*; MTurk. *burčaq* (Sangl.); Uzb. *burčəq*; Uig. *počaq*; Krm. *burčax*; Tat. *borčaq*; Bashk. *borsaqa*; Kaz. *buršaq*; KBalk. *burčaq*; KKalp. *buršaq*; Kum. *burčaq*; Nogh. *buršaq*; SUig. *pirčaq*; Shr. *mirčaq*; Oyr. *mirčaq*; Chu. *pəʷrza*, *pəʷrže* (Starostin et al. 2003: 380).

Cf. Mong. **burčag* ~ **buyurčag* ‘pea’ (probably Mongolic > Turkic).

**būn-* ~ **bīn-* ‘to mount a horse, ride on’

OTurk. *bin-* (Orkh.), *mūn-* (OUygh.); Karakh. *mūn-* (MK, KB); Tur. *bin-*; Gag. *pin-*; Az. *min-*; Turkm. *mūn-*; *mīn-* (dial.); Sal. *min-*, *mim-*, *miŋ-* (CCЯ); MTurk. *min-* (Sangl.); Uzb. *min-*; Uig. *min-*; Krm. *min-*; Tat. *men-*; Bashk. *men-*; Kirgh. *min-*; Kaz. *min-*; KBalk. *min-*; KKalp. *min-*; Kum. *min-*; Nogh. *min-*; SUig. *min-*; Khak. *mūn-*; Shr. *mūn-*; Oyr. *min-*; Tv. *mūn-*; Tof. *mūn-*; Chu. *minder* ‘pillow’; Yak. *mīn-*; Dolg. *mīn-* (Starostin et al. 2003: 1110).

**čōčka* ‘young pig’

Karakh. *čočuq* (MK) 1; Tur. *čožuk* 2; Gag. *čožuq* 2; Az. *čošGa* 1, 3; Turkm. *žōžuq* 1 (cf. *colloq. čōča* ‘camel’); MTurk. *čočya* 1 (Sangl.), (OKypch.) *čočqa* (Houts.) 1; Uzb. *žužuq* 2; Uig. *čošqa* 3; Krm. (K) *čočqa* 3, *čočuq jatayi* ‘afterbirth’; (T) *čočxa* ‘young boy (not a Karaim)’, (H) *cocka* 2; Tat. *čučqa* 3; Bashk. *sosqa* 3; Kirgh. *čočqo* 1; Kaz. *šošqa* 1; KBalk. *čočxa* 3; KKalp. *šošqa* 3; Kum. *čočqa* 3; Nogh. *šošqa* 3; Khak. *sosxa* 3; Shr. *šošqa* 3; Oyr. *čočqo* 3; Tv. *šošqa* 3 (Starostin et al. 2003: 1335).
‘young pig 1, child, boy 2, pig 3’

**dāna* ‘(two-years-old) heifer’

MKypch. *tana* 1; Chag. *tana* 2; Tur. *dana* 1; Gag. *dana* 2; Az. *dana* 1, Turkm. *tāna* 1; Sal. *tana* 3; Kar. *tana* 1; Kum. *tana* 2; KBalk. *tana* 2; Tat. *tana* 2; Nogh. *tana* 3; KKalp. *tana* 3; Kaz. *tana* 1; Kirgh. *tana* 3; Chu. *tīna* 3 (Dybo 2007: 116–117).

‘calf 1, calf (two-years-old) 2, heifer 3’

< probably East Iranian, see (Bailey 1979: 159).

**darīg* ‘corn’ > ‘broomcorn millet’

OTurk. *tariy* (OUygh.) 2, 3; Karakh. *tariy* (MK) 2, 3; MKypch. *tariy* 1; Tur. *dari* 1; Gag. *dari* 1; Az. *dari* 1; Turkm. *dari* 1; Sal. *dari*; MTurk. (MKypch.) *tari* (CCum., AH); Uzb. *tariq* 1; Uig. *teriq* 1; Kar. *tari*, *dari* 1; Tat. *tari* 1; Bashk. *tari* 1; Nogh. *tari* 1; Kaz. *tari* 1; Kirgh. *taru*: 1; KBalk. *tari* 1; Kum. *tari* 1; Khak. *tariy* 4; Chu. *tirə* 2; Bulg. > Hung. *dara* ‘grain, groats’ (Tenišev et al. 2001: 456–458; Starostin et al. 2003: 1356).

‘proso (broomcorn) millet 1, corn 2, cultivated land 3, sowing 4’

Possibly a derivative of **TArī-* ‘to cultivate (ground)’.

**debe* ‘camel’

OTurk. *tebe* (Orkh.), *teve* (OUygh.); Karakh. *teve* (*tevey*) (MK); Tur. *deve*; Gag. *devä*; Az. *devä*; Turkm. *düye*; Sal. *töye*, *töüvä*, *tüvi*; MTurk. *deve* (Pav. C.), *teve* (Abush., Pav. C.); Uzb. *tuya*; Uig. *tögä*; Krm. *tüye*, *deve*; Tat. *düyä*; Bashk. *düyä*; Kirgh. *tō*; Kaz. *tüye*; KBalk. *tüye*; KKalp. *tüye*; Kum. *tüye*; Nogh. *tüye*; SUig. *te*, *ti*; Khak. *tibe*; Oyr. *tō*, *tebe*; Tv. *teve*; Tof. *tebe* (Pac. ФИЛ); Chu. *təʷve*; Yak. *taba* ‘deer’; Dolg. *taba* ‘deer’ (Starostin et al. 1424–1425).

Cf. PMo **teme-yen* ‘camel’: MMong. *teme'en* (SH), *temeyen* (HY 11), *temē* (IM), *təmēn* (LH), *təmən*, *timen* (MA); WMo. *temege(n)* (L 800); Kh. *temēn*; Bur. *temē(n)*; Kalm. *temēn*; Ord. *temē(n)*; Mog. *temɔ* (Weiers); Dag. *temē* (Тод. Дар. 166, MD 223); S.-Yugh. *temen*; Mongr. *timēn* (SM 420), *təmēn* (possibly < Turkic).

**doŋur* ‘pig’

OTurk. *toŋuz* (OUygh.); Karakh. *toŋuz* (MK); Tur. *domuz*; Gag. *domuz*; Az. *donuz*; Turkm. *doŋuz*; Sal. *toŋas*; MTurk. *toŋuz* (Sangl.); Uzb. *toŋyis*; Uig. *toŋyuz*; Krm. *toŋyuz*, *domuz*; Tat. *dunŷiz*; Kirgh. *doŋuz*; Kaz. *doŋiz*; KBalk. *toŋyuz*; KKalp. *doŋiz*; Kum. *doŋyuz*; Nogh. *doŋiz*; SUig. *doŋiz* (Starostin et al. 2003: 1355).

**dorak* ‘a k. of cheese or quark’

Chag. *doraq*, Khal. *tuoraq*, Turkm. *doraq*, Tur. dial. *torak*, *dorak*; Chu. *tora*, *tɔwara*; Bulg. > Hung. *turo* ‘quark’ (Dybo 2007: 117);
< Mİr. **tura-ka* (Bailey 1979: 132).

**ebin* ‘grain, seed’

OTurk. *evin* (OUygh.); Karakh. *evin* (MK, KB); Tur. Osm. *evin*, Anat. *efin*; MTurk. *evin* (Qutb); Oyr. *ebin*; Chu. *avən šap-* ‘to flail’, *avən karti* ‘cornfloor’ > Mari (Low) *avən*, Mari (High) *en* (Starostin et al. 2003: 578).

**ečkü* ‘(she-)goat’

**ečki* (Sevortjan et al. 1974–2003, 3: 34–35), **äžikä* > **äčki* (Tenišev et al. 2001: 426–427), **ečkü* (Dybo 2007: 123)

OUig. *äčkü*; Karakh. *äčkü*; MKypch. *äčki*; Khal. *äčgü*, *äččü*; Kar. *äčki*; KBalk. *äčki*; Kum. *äčki*; Nog. *äški*; KKalp. *äški*; Kirgh. *äčki*; Alt. *äčki*; Uzb. *äčki*; Uigh. *öčkä*, *äčkü*; Khak. *öski*; Tuv. *öškü*; Tof. *öškü* (Tenišev et al. 2001: 426–427).

The root is often confused with another word for ‘(she-)goat’, **geče* (~ *geči*) (see).

**eđner* ‘saddle’

Karakh. *eđer* (MK); Tur. *eyer*; Gag. *yēr*; Az. *yähär*; Turkm. *eyer*; Sal. *eđer* (Kakuk); MTurk. *eđer*; Uzb. *egar*; Uig. *eğə(r)*; Krm. *yer*; Tat. *iyer*; Bashk. *eyär*; Kirgh. *ēr*; Kaz. *er*; KBalk. *iyer*; KKalp. *yer*; Kum. *er*; Nogh. *iyer*; SUig. *ezer*; Khak. *izer*; Shr. *ezer*; Oyr. *ēr*; Tv. *ezer*; Tof. *e'zer* (Рас. Фил 183); Chu. *yäner*; Yak. *ijir*; Dolg. *ijir* (Starostin et al. 2003: 506).

**ek-* ‘to sow’

OTurk. *ek-* (Late OUygh.) 1; Karakh. *ek-* (MK, KB) 1, 2; Tur. *ek-* 1; Gag. *ek-* 1; Az. *äk-* 1, 2; Turkm. *ek-* 1; Sal. *ex-* 1; Khal. *häk-* 1; MTurk. *ek-* (Abush., Sangl.) 1; Uzb. *ek-* 1; Uig. *ek-* 1; Krm. *ek-* 1; Tat. *ik-* 1; Bashk. *ik-* 1; Kirgh. *ek-* 1; Kaz. *ek-* 1; KKalp. *ek-* 1; Nogh. *ek-* 1; Chu. *ak-* 1 (Starostin et al. 2003: 1132).

‘to sow 1, to scatter 2’

**ełgek* ‘donkey’

OTurk. *ešgek* (OUygh.); Karakh. *ešgek*, *ešyek* (MK); Tur. *ešek*; Gag. *iešek*; Az. *eššäk*; Turkm. *ešek*; MTurk. *ešek* (Бор. Бад., Abush., Pav. C.); Uzb. *ešäk*; Uig. *ešäk*; Krm. *ešek*; Tat. *išäk*; Bashk. *išäk*; Kirgh. *ešek*; Kaz. *esek*; KBalk. *ešek*; KKalp. *ešek*; Kum. *ešek*; Nogh. *ešek*; Oyr. *ešte*; Chu. *ažak* (Starostin et al. 2003: 503).

Cf. PMo **elžigen* ‘donkey’: MMong. *elžigan* (HY 9), *ülžige* (IM), *ilžigen* (LH), *ilžigän* (MA); WMo. *elžige(n)* (L 311); Kh. *ilžig*, *ilžgen*; Bur. *elžege(n)*; Kalm. *elžyne*, *elžynə*; Ord. *elžige(n)*; Mog. *elžiyön*; Dong. *enžeye* (Тод. Дн.); Bao. *nžige* (Тод. Бн.); Ma. *eihen* (possibly Turkic > Mongolic > Tungusic).

**erkeč* ‘gelded he-goat’

OUIg. *ärkäč* 1; Karakh. *ärkäč* 1, 6; MKypch. *ärkäč* 1, 2; Tur. *ärkäč* 1, dial. *ärgäč*, *ürgäč*, *ärgäs* 2; Az. dial. *ärkäč* 2; Turkm. *ärkäč* 4; KBalk. *ärkäč* 3, 6; Kum. *ärkäč* 5; Kirgh. *ärkäč* 2, 6 (Tenišev et al. 2001: 428–429).

‘he-goat 1, gelded he-goat 2, (three years old) he-goat 3, (two years old) he-goat 4 (one year old) goat 5, bellwether 6’.

**etmek* ‘bread’

OTurk. *ötmek* (OUygh.); Karakh. *etmek* (MK), *epmek* (MK – Oghuz, Qypch.); Tur. *etmek*, *ekmek*; Gag. *iekmek*; Az. *äppäk*; Turkm. (dial.) *ekmek*, *epmek*; MTurk. *etmek*, *ötmek* (Pav. C.); Krm. *ekmek*, *etmek*, *ötmek*; Tat. *ikmäk*; Bashk. *ikmäk*; KBalk. *ötmek*; Kum. *ekmek*; Nogh. *ötpek*; Khak. *ipek*; Shr. *itpäk*; Oyr. *ötpök* (Starostin et al. 2003: 594).

Cf. PMo **ide*- ‘to eat’.

**geče* (~ *geči*) ‘(she-)goat’

**geči* (Sevortjan et al. 1974–2003, 3: 34–35), **käči* (Tenišev et al. 2001: 426–427), **geče* (Dybo 2007: 123)

Chu. *Kaĵa-ga* 1; Bulg. > Hung. *kecske* 1; Turkm. *geči* 1; Tur. *keči*, dial. *geči* 1, 2; Az. *keči* 1, 2; Gag. *keči* 1; Karakh. *käči*; MTurk. *käči*; Tat. *käjä* 1; Bashk. *käzä* 1; Uzb. dial. *geji* 1 (Tenišev et al. 2001: 426–427).

‘(she)goat 1, he-goat 2’

The root is often confused with another word for ‘(she-)goat’, **ečkü* (see).

**göpe-ne* ‘haystack’

Tur. *geben*; Tat. *kübe*; Bashk. *kübe*; Kum. *keben*; Tv. *xöpēn*; Chu. *koba* (Starostin et al. 2003: 723).

Delabialization of **ö* in some languages is secondary.

**ingek* ‘cow’ and **ingen* ‘female camel’

OTurk. *ingek* (Orkh., OUygh.) 1, *ingen* 2 (OUygh.); Karakh. *ingek* 1, *ingen* 2 (MK); Tur. *inek* 1; Gag. *inek* 1; Az. *inäk* 1; Turkm. *inek* 1, *inen* 2; MTurk. *inek* 1 (AH), *inen* 2 (Pav. C.); Uzb. *inäk*, *inäy* 1 (dial.); Uig. *inäk* 1, (dial.) *ingan*, *inġan* 2; Krm. *inek* 1; Tat. *inäk* 1 (dial.); Kirgh. *inek* 1, *inġen* 2; Kaz. *inek* 1, *inġen* 2; KBalk. *inek*, *iynek* 1; KKalp. *inġen* 2; Kum. *inek* 1; SUig. *inek*, *enek* 1; Khak. *inek* 1; Shr. *inek*, *näk* 1; Oyr. *inek*, *iynek* 1; Tv. *inek* 1, *enġin* 2; Chu. *əne* 1; Yak. *inax* 1 (Starostin et al. 2003: 619).

‘cow 1, female camel 2’

Cf. PMo **üniyen* ‘cow’: MMong. *uni'en* (SH), *uneyen* (HY 11); WMo. *üniye(n)* (L 1010); Kh. *ünēn*; Bur. *ünē(n)*; Kalm. *ünē*, *ünēn*; Ord. *ünē(n)*; Mog. *ünä*; Dag. *unē*, (Тод. Дар. 171) *ünē*; Bao. *unaŋ*; S.-Yugh. *nin*; Mongr. *unē* (SM 472) (probably < Turkic).

**(i)un* ‘flour’

Karakh. *un*; MKypch. *un*; Chag. *un*; Tur. *un*; Gag. *un*; Az. *un*; Khal. *hu:n*; Turkm. *u:n*; Kar. *un*; KBalk. *un*; Kum. *un*; Tat. *on*; Bashk. *on*; Nogh. *un*; KKalp. *un*; Kaz. *un*; Kirgh. *un*; Uzb. *un*; Uigh. *un*; Khak. *un*; ? Chu. *śnəχ* (Tenišev et al. 2001: 471).

*kerki ‘adze, mattock’

Karakh. *kerki* (MK) 1, *kerey* (MK) 2; Tur. *kerki* 1; Az. *kärki*, *kerki* (dial.) 1; Turkm. *kerki* 1; MTurk. *kerki* (IM, AH) 1; Uig. *keke*, *kerke* (dial.) 1; Kirgh. *kerki* 1; Oyr. *kerki* 1; Tv. *keržek* ‘adze’; Chu. *karǝ* ‘chisel’ (Starostin et al. 2003: 791).

‘adze, mattock 1, razor 2’

*kētmen ‘hoe, mattock’

OTurk. *ketmen* (OUigh.); Karakh. *ketmen* (MK); Tur. *gedmen*; Az. *kätmän*; Turkm. *kätmen*; MTurk. *ketmen* (Sangl.); Uzb. *ketmən*; Uig. *kätmän*; Bashk. *kätmän*; Kirgh. *ketmen*-; Chu. *katmak* (Starostin et al. 2003: 810).

The root is usually derived from *gēt- ‘to notch’ but the Oghuz languages systematically distinguish *g- in ‘notch’ and *k- in ‘hoe’.

*kidir ‘felt’

OTurk. *kidiz* (OUygh.); Karakh. *kiđiz* (MK, KB); Tur. *kiyiz*, *keyiz* (dial.); Turkm. *kiz*; MTurk. *kiyiz* (IM, Abush., Qutb., Houts.); Uzb. *kigiz*; Uig. *kigiz*; Tat. *kiyez*; Bashk. *keyeđ*; Kirgh. *kiyiz*; Kaz. *kiyiz*; KBalk. *kiyiz*; KKalp. *kiyiz*, *kiygiz* (dial.); Kum. *kiyiz*; Nogh. *kiyiz*; Khak. *kis*; Oyr. *kiyis*; Tv. *kidis* (Starostin et al. 2003: 846).

Turk. > MMong. *kiyiz* (Шеп6ак 1997: 127).

*Koč ‘ram’

Turk. > Hung. *kos*; OUig. *qočqar*, *qočunjar*; Karakh. *qočnar*; Tur. *koč*, *kočkar*; Gag. *qoč*; Az. *Goč*; Turkm. *Goč*, *GočGar*; Sal. *qošqor*, *qošqur*; Khal. *Goč*; MTurk. *qoč*, *qočqar*; Uzb. *yoč* (dial.), *qočqar*; Uig. *qočqa(r)*, *qošqa(r)*; Krm. *qoč*, *qočqar*, *qočxar*; Tat. *qučqar* (dial.); Bashk. *qušqar*; Kirgh. *qočqor*; Kaz. *qošqar*; KBalk. *qocxar*; KKalp. *yoš*, *qošqar*; Kum. *qočqar*; Nogh. *qošqar*; SUig. *qužyar*; Oyr. *qočqor*; Tv. *qošqar* (Starostin et al. 2003: 711–712).

Cf. PMo *kuča ‘ram’: MMong. *xuča*, *qəča* ‘lamb’, *quča*; WMo. *quča*; Kh. *xuc*; Bur. *xusa*; Kalm. *xuca*; Ord. *Guča*; Dag. *koč*; Dong. *quža*; S.-Yugh. *quža*; Mongr. *xuž’a*, *xuža* (possibly < Turkic).

*Kōm ‘camel’s pack-saddle’

Karakh. *qom* (MK); Turkm. *Gōm*; MTurk. *qom* (MA); Uzb. *qum*; Bashk. *qum*; Kirgh. *qom*; Kaz. *qom*; KKalp. *qom*; Oyr. *qom*; Tv. *qom*; Tof. *xom* (Starostin et al. 2003: 717).

Turk. > WMo. *qom* (KW 184, Шеп6ак 1997, 139), whence Evk. *kōm*, Man. *qomo* (see TMC 1, 408, Doerfer MT 61).

*Konak ‘foxtail millet’

OUig. *qonaq* 1; Karakh. *qonaq* 1; Chag. *qonay*, *qonaq*; KKalp. *qonaq* 1; Kirgh. *qonoq* 1, 2; Uzb. *qunčq* 1, 3; Uigh. *qonaq* 3, 4, 5; Tuv. *xonaq* 2 (Tenišev et al. 2001: 458–459).

‘foxtail millet 1, a k. of weed 2, broomcorn millet 3, sorgho 4, maize 5’

*koń (~ *Koyñ) ‘sheep’

*qoni > *qoń (Tenišev et al. 2001: 431–432), *Koyñ (Dybo 2007: 43)

Tuv. *xoy*; Tofa *hoj*; Khak *xoy*; S.-Yugh. *xoy*; OTurk. *qoŷ*; Karakh. *qoy*; Argu (MK) *qon*; Chag. *qoy*; Uzb. *qoy*; Uigh. *qoy*; Khal. *qōon*; Tur. *koy(u)n*; Az. *Goy(u)n*; Turkm. *Goy(u)n*; Kar. *qoy*; Kum. *qoy*; KBalk. *qoy*; Tat. *quy* ‘fat-tailed sheep’; Bashk. *quy* ‘fat-tailed sheep’; Nog. *qoy*; Kaz. *qoy*; KKalp. *qoy*; Kirgh. *koy*; Oyr. *koy* (Dybo 2007: 43).

> Mong. *koni-n* > Tung. *konin* ‘id.’ (Janhunen 2012).

**Koři* ‘lamb’

OÜig. *qozi* (*quzi*); Karakh. *qozi* (*quzi*) (MK); Tur. *kuzu*; Gag. *quzu*; Az. *Guzu*; Turkm. *Guzi*; Sal. *qōza*; Khal. *quzi*; MTurk. *qozi*, *qozu*; Uzb. *quzi*; Üig. *qoza*; Krm. *qozu*; Tat. *quzi* (dial.); Kirgh. *qozu*; Kaz. *qozi*; KBalk. *qozu*; KKalp. *qozi*; Kum. *qozu* (dial.); Nogh. *qozi*; Süig. *quzi*, *qozi*, *qoza* (Starostin et al. 2003: 809).

Cf. PMo **kurigan* ‘lamb’: MMong. *quriqa(n)*, *qurixan*, *qəriyan*; WMo. *quriyan*, *quraya(n)*, *qurya(n)*; Kh. *xurgan*; Bur. *xurğa(n)*; Kalm. *xuryñ*; Ord. *xurGa*; Mog. *qoryan*; Dong. *quyan*, *Guyan*; Bao. *GorGañ*; S.-Yugh. *χurGan*; Mongr. *xorGa*, *xuroG* (possibly < Turk.).

**kōlēk* ‘young of camel’

Tur. *kōšek*, *gōšek* (dial.) 1; Az. *koşak* 1; Turkm. *kōšek* 1; MTurk. *kōšek* (AH) 1; Uzb. *kūšek* (dial.) 1; Bashk. *kölökey* 2; KKalp. *kōšek* (dial.) 1; Kum. *kiley* (dial.) 3 (Starostin et al. 2003: 717); Bulg. > Hung. *kōlyök* ‘young of an animal, kid, puppy, lad’ (Róna-Tas & Berta 2011: 586–588).

‘young of camel 1, calf 2, cub 3’.

Cf. PMo **gölige* ‘pup, young dog or cat’: WMo. *gölüge*, *gölige* (L 386); Kh. *gölög*; Bur. *gülge(n)*; Kalm. *gölgə*; Ord. *gölögö*; Dag. *gulug*, *gulgū* (Тод. Дар. 133); S.-Yugh. *gələg*; Mongr. *gor-go* (SM 143), *gulgo*.

**Kulum* ‘foal’

OTurk. *qulun* (Yen.); Karakh. *qulun* (MK); Tur. *kulun*; Az. *Gulun*; Turkm. *Gulun*; MTurk. *qulun*, *qulum* (Pav. C.); Uzb. *qulun* (dial.); Üig. *qulun* (dial.); Tat. *qolin*; Bashk. *qolon*; Kirgh. *qulun*; Kaz. *qulin*; KKalp. *qulin*; Nogh. *qulin*; Süig. *qulun*, *qulum*, *qulim*, *qolun*; Khak. *xulun*; Oyr. *qulun*; Tv. *qulun*; Chu. *xə^wm*; Yak. *kulun* (Starostin et al. 2003: 735).

May be a Proto-Altaic root, cf. PMo **kulan* ‘ass’: MMong. *qulan* (SH), *qulan* (MA); WMo. *qulan*, *külen* (L 984); Kh. *xulan*; Bur. *xulan*; Kalm. *xulñ*, *xulñ*; Ord. *xulan*.

**Kumir* ‘alcohol milk drink’

Karakh. *qimiz* (MK, KB); Tur. *kimiz*; Az. *Gimiz*; Turkm. *Gimiz*; MTurk. *qimiz* (Pav. C.); Uzb. *qimiz*; Üig. *qimiz*; Tat. *qimiz*; Bashk. *qomod*, *qimið*; Kirgh. *qimiz*; Kaz. *qimiz*; KKalp. *qimiz*; Nogh. *qimiz*; Khak. *ximis*; Sag. Koib. *xumis*; Oyr. *qimis*; Tv. *ximis*; Chu. *kə^wmə^s* < Kypch.; Yak. *kimis* (Starostin et al. 2003: 641).

Cf. PMo PMo **kimur* ‘fermented milk with water’: WMo. *kimur*, *kimurayan*; *kiram*, *kirma* (L 470) ‘boiled milk with water’; Kh. *xaram* ‘boiled water with milk’; Kalm. *kimr*, *kimrān*; Ord. *kirma* (possibly < Turkic).

**Kürüt* ‘a k. of dried quark, cheese’

OÜig. *qurut*; Karakh. *qurut*; MKypch. *qurut*; Chag. *qurut*; Tur. *kurut*; Az. *qurut*; Turkm. *qurt*; Tat. *qort*; Bashk. *qort*, *qorot*; Nogh. *qurt*; KKalp. *qurt*; Kaz. *qurt*; Kirgh. *qurt*, *qurut*; Oyr. dial. *qurut*, *quryut*; Uzb. *qurt*, *qurut*; Üigh. *qurt*; Khak. *xurut*; Tuv. *qurut*; Chu. dial. *kə^wrt* (Sevortjan et al. 1974–2003, 6: 170–171).

A derivative of **Kür-* ‘to dry’.

*oglak 'kid'

*oyılaq (Tenišev et al. 2001: 429–430), *oglaq (Dybo 2010: 83)

Bulg. (or Kypch.) > Hung. *olló* (ibid.); Karakh. *oylaq*; MKypch. *oylaq*, *oyalaq*, *oyulaq*, *ulax*; Tur. *odlak*; Gag. *olak*; Az. *oylay*, *oylaq*; Turkm. *owlaq*; Sal. *oylax*, *olax*; Kar. *ulaq*; KBalk. *ulaq*; Kum. *ulaq*; Nog. *ulaq*; Bashk. *ılaq*; KKalp. *ılaq*; Kaz. *laq*; Kirg. *ulaq*; Oyr. *ulaq*, *uwlaq*; Uzb. *ulq*; Uigh. *oylaq*; Khak. *oylax* 'young wild goat' (Tenišev et al. 2001: 429–430).

A derivative of *ogul 'son' or *ogla- 'to shout, to make a racket'.

*or- 'to mow, reap, harvest (a crop)'

Karakh. *or-* (MK) 1; Turkm. *or-* 1, 2; Kar. *or-* 1, 2; Kum. *or-* 1; KBalk. *or-* 1, 2; Kirgh. *or-* 1, 2; Kaz. *or-* 1; Nogh. *or-* 1; KKalp. *or-* 1, 2; Uigh. *or-* 1, 2; Uzb. *or-* 1, 2; Sal. *or-* 1, 2; Tat. *ur-* 1; Bashk. *ur-* 1; S.-Yugh. *ur-* 2, 3; Tur. *ora-* 1; Chu. *vir-* 1 (Sevortjan et al. 1974–2003, 1: 468).

'to reap, harvest (a crop) 1, to mow 2, to cut grass 3'

*ökür 'bull, ox'

OTurk. *öküz* (OUygh.); Karakh. *öküz* (MK); Tur. *öküz*; Gag. *yöküz*; Az. *öküz*; Turkm. *ökiz*, *öküz*; MTurk. *öküz* (Pav. C.); Uzb. *hökiz*; Uig. *öküz*, *höküz*; Krm. *öküz*, *ögüz*; Tat. *ugiz*; Bashk. *ugid*; Kirgh. *ögüz*; Kaz. *ögiz*; KBalk. *ögüz*; KKalp. *ögiz*; Kum. *ögüz*; Nogh. *ögiz*; SUig. *kus*; Chu. *və^wgə^wr*; Yak. *oyus*; Dolg. *ogus* (Starostin et al. 2003: 1168–1169).

Cf. PMo *hüker 'ox': MMong. *xuker* (SH), *xuger* (HY 10), *ukär* (MA); WMo. *üker* (L 1003); Kh. *üxer*; Bur. *üxer*; Kalm. *ükr* 'cow'; (KPC); Ord. *üker*; Mog. *ükär* (Weiers), ZM *okär* (20–4); Dag. *xukur* (Тод. Дар. 179), *hukure* (MD 166); Dong. *fugie(r)*; Mongr. *fugor* (SM 104), *xukur* (Minghe). Cf. also Evk. *hukur*; Evn. *höken*, *hökön*; Sol. *uxur* 'ox' (possibly Turkic > Mongolic > Tungusic).

*sag- 'to milk'

OTurk. *say-* (OUygh.); Karakh. *say-* (MK); Tur. *sā-*, dial. *say-*; Gag. *sā-*; Az. *say-*; Turkm. *saG-*; Sal. *sax-*; Khal. *sa:y-*; MTurk. *say-* (Pav. C.); Uzb. *say-*; Uig. *say-*; Krm. *sav-*; Tat. *saw-*; Kirgh. *sā-*; Kaz. *saw-*; KKalp. *saw-*; Kum. *sav-*; Nogh. *saw-*; SUig. *say-*; Khak. *say-*; Oyr. *sā-*; Tv. *say-*; Chu. *sə^wv-*; Yak. *ia-* (Starostin et al. 2003: 1198).

The root is likely to be genetically connected with PMo *saya- 'to milk': MMong. *sa'a-* (SH), *sa-* (MA 319); WMo. *saya-* (L 656); Kh. *sā-*; Bur. *hā-*; Kalm. *sā-*; Ord. *sā-*; Mog. *so-* (Weiers); ZM *sā-* (23–5b); Dag. *sā-* (Тод. Дар. 161, MD 204); Dong. *sa-*; Bao. *sā-*; S.-Yugh. *sā-*; Mongr. *s(w)ā-* (SM 356), *sāli* 'animal qu'on trait, femelle (brebis, chèvre)' (SM 321).

*sarik 'sheep'

Tat. *sariq* 1; Bashk. *hariq* 1; Kaz. *sariq* 2; KKalp. *sariq* 2; Chu. *sorəx* 1 (Starostin et al. 2003: 1283)
'sheep 1, a k. of tailless sheep 2'

Cf. PMo *serke 'gelded goat': WMo. *serke*; Kh. *serx*; Bur. *herxe*; Kalm. *serkə*; Ord. *serxe*; Dag. *selek*, *selke*; S.-Yugh. *serke* (possibly < Turkic).

*sa(r)pan 'plough'

Karakh. *saban* (MK); Tur. *saban*; Gag. *saban*; Az. *sapan*; Sal. *sovan* 'coxa' (CCЯ); MTurk. *saban* (IM, AH), *sapan* (Pav. C.); Uig. *sapan*; Krm. *saban*; Tat. *saban*; Bashk. *haban*; Kaz. *saban*; KBalk. *saban*; Kum. *saban*, *sarapan* 'plough breast'; Nogh. *saban*; Chu. *sorban* 'plough breast' (Starostin et al. 2003: 1216).

**sīgīr* ‘cattle’

Bulg. *šegor*, Karakh. *sīyīr* 1, 4, MKypch. *sīyīr*; Chag. *sīyīr* 3; Tur. *sīyīr* 1; Gag. *sīyīr* 1, 2; Az. *sīyīr* 1; Khal. *sīyīr* 6; Turkm. *sīyīr* 4, 6; Kar. *sīyīr* 1, *sīyīr* 4; KBalk. *sīyīr* 1, 4; Kum. *sīyīr* 4; Tat. *sīyīr* 4; Nogh. *sīyīr* 4; KKalp. *sīyīr* 6; Bashk. *hīyīr* 4; Kirgh. *sīyīr* 4; Oyr. *sīyīr* 4; Uzb. *sigir* 6; Uigh. *siyir* (Tenišev et al. 2001: 435–436).

‘cattle 1, herd 2, bull 3, cow 4, the year of cow 5’

Probably a Proto- Altaic root, cf. pTg **sig-* / *seg-* ‘wild deer’: Evk. *segžen*, dial. *sekserge* ‘wild deer’; Nan. *segži* ‘herd of wild swine’; Ud. *sigisa* ‘one year old maral’. Cf. also PJap. **sika* ‘deer’ and probably pMo **seyenek* (~ -i-) ‘he-goat (2 years old)’: WMo. *segenek* (L 684: *sejinug*); Kh. *sijneg*; Bur. *hīneg* ‘castrated he-goat; ox’; Kalm. *sīnək* (Starostin et al. 2003: 1243–1244).

**sogan* ‘onion’

OTurk. *soyun* (OUygh.); Karakh. *soyun* (MK); Tur. *soyan*; Gag. *suvan*, *suan*; Az. *soyan*; Turkm. *soyan*; Sal. *soyan*, *soyän*; Khal. *soyan*; MTurk. *soyan* (AH, IM, Pav. C.); Uig. *soyan*; Tat. *suyan*; Bashk. *huyan*; Kirgh. *soyan*, *soyon*; KBalk. *soxan*; Kum. *soyan*; Nogh. *soyan*; SUig. *soxan*; Chu. *soyan* (Starostin et al. 2003: 1303).

Cf. PMo **songina* ‘onion’: MMong. *soṅgina* (HY 8), *sunqinā* (MA); WMo. *songina* (L 727); Kh. *songin*; Bur. *hongino*; Kalm. *songina*; Ord. *songinoG*; Dong. *sunguna*; Mongr. *suṅGunoG* (possibly < Turkic).

**TArī-* ‘to cultivate (ground)’

OTurk. *tari-* (OUygh.); Karakh. *tari-* (MK, KB); MTurk. *tari-* (Abush., Sangl.); Uig. *teri-* (dial.); SUig. *tari-*; Khak. *tari-*; Oyr. *tari-*; Tv. *tari-*; Tof. *tari-* (Starostin et al. 2003: 1438).

See also **dārīg* ‘corn’. Cf. WMo. *tari-* ‘to sow, plant, plough’, pTg **tari-* ‘to cultivate, farm, plow’: Evk. *tari-* ~ *tare-*, Solon *tari-*, Manchu *tari-*, Nanai/Ulcha *tari-*.

**teke* ‘he-goat’

OUig. *teke*; Karakh. *teke* (MK, IM); Tur. *teke*; Gag. *teke*; Az. *tākā*; Turkm. *teke*; Khal. *tākā*; MTurk. *teke* (Sangl.); Uzb. *taka*; Uig. *tekā*; Krm. *teke*, *tege*; Tat. *tākā* ‘козел, баран’; Bashk. *tākā* ‘he-goat, ram’; Kirgh. *teke*; Kaz. *tekā*; KBalk. *teke*; KKalp. *teke*; Kum. *teke*; Nogh. *teke*; SUig. *teke*; Oyr. *teke*; Tv. *de’ge*, *te* ([*dhe*]); Tof. *te’he*; Chu. *taga* ‘he-goat, ram’ (Starostin et al. 2003: 1430–1431)

> Mong. *teke* ‘he-goat’.

**tirma-k* ‘harrow’

Tur. *tirmik*, Gag. *tirmik*; Tat. *tirma*; Kum. *taraq*; Yak. *taraax*; Uigh. *tarmaq*; Khak. *tarbas-ta-* ‘to harrow’ (Tenišev et al. 2001: 467–468)

A common derivative of **tirma-* ‘to scratch’.

**tōr-um* ‘camel colt’

Karakh. *torum* 1, *torpi* 2 (MK); Tur. *deve torun* 1, *torum* (dial.) 1; *torbuč* (dial.) 3, (?) *toru* (dial.) 4; Gag. (?) *tor* ‘unbroken (of a horse), untrodden (of a path)’; Turkm. *tōrum* 1; Sal. *tori* ‘foal’ (CCЯ); MTurk. *torum* 1 (Sangl., Pav. C.), *torbaq* 2 (MA 126); Uig. *topaq* 2, *topaq-torum* ‘young calves’; Tat. *tōrbaq* (KCTT) 2; Bashk. *tana-turpaq* 2; Kirgh. *torpoq* 2; Kaz. *torpaq* 2; Khak. *torbax* 2; Oyr. *torboq* 2, *torboč* (dial. Kumd.) 5; Tv. *dorum* 1; Yak. *torbos*, *torbužax* 2 (Starostin et al. 2003: 1464).

‘young camel 1, a young calf 2, a goat that has yeaned early 3, young 4, a cow that has not calved yet 5’

Probably a Proto-Altaiic root, cf. PMo **toruy* ‘young pig’: WMo. *torui* (L 827); Kh. *toroy*; Bur. *toroy*; Kalm. *torā*; Ord. *torō* ‘young donkey’; pTg **tora-kī* (~-ū) ‘boar’: Evk. *torokī*; Neg. *torokī*.

**tögi* ‘millet groats’

OUIg. *tögö* 1; MKypch. *tüwi*, *tü* 1; Tur. *dügü* 1, 2; Az. *düyü* 2; Turkm. *tüvi* 2; KBalk. *tüy* 1; Tat. *döge* 2; Nogh. *tüy* 1.

‘husked millet, millet groats 1, husked rice 2’.

A common derivative of **tög-* ‘to crush, to husk (e.g. grain)’.

**ud* (~ **od*) ‘cattle’

OTurk. *ud* (OUygh.); Karakh. *uδ* (MK); MTurk. *uy* (Бор. Бад., Abush., Pav. C.); Uig. *uy*; Kirgh. *uy*; SUig. *ut*; Oyr. *uy* (Starostin et al. 2003: 1484); ? Chu. *vil’əχ < və’yləχ* (< **od*) (Mudrak 1993).

Cf. PMo **odus* ‘wild yak, buffalo’: MMong. *odos* (HY 11); WMo. *udus* (L 862); Kh. *odos* (BAMPC) (possibly < Turkic).

**ulāla* ‘(small) horse’

Chu. *laža* 1, 3; Turkm. *alaša* 1; Tur. dial. *alaša* 2; Az. *alaša* 2; KTat. *alaša* 1; Kar. (K) *alaša* 3; Kum. *alaša* 1, 3, 4; KBalk. *alaša* 1, 3, 4; Tat. *alaša* 1, 3, 4, 5; Bashk. *alaša* 1; Kirgh. *alasa*; Kaz. *alasa*; Nogh. *alasa*; KKalp. *alasa*; Uz. dial. *lača* 2 (Sevortjan et al. 1974–2003, 1: 135–136).

‘gelding 1, bad/small horse 2, horse 3, small 4, bad, ugly 5’

Possibly a derivative of **al-* ‘to be(come) weak’. On reconstruction of the initial vowel see Tenišev et al. (2006: 181).

**ügür* ‘broomcorn millet’

OUIg. *üyür* 1, 2; Karakh. *ügür* (*yügür*; *yü:r*) 1; Chu. *vir* 1; Yak. *üöre* 1; ? Tv. *ü:rgene* 3 (Starostin et al. 2003: 1548; Tenišev et al. 2001: 458).

‘proso (broomcorn) millet 1, seeds, grains (e.g., of seasonings) 2, a k. of buckwheat 3’.

**yasmik* ‘lentils’

Chag. *yasmuq*; Tur. *yasmuq*; Turkm. *yasmiq*; Tat. *yasmiq*; Bashk. *yaθmiq*; Uzb *yasmiq* (Tenišev et al. 2001: 464–465).

A common derivative of **yas-* ‘to be(come) flat’.

**yAsna-k* ~ **yAsna-g* ‘pig’

Bulg. > Hung. *disznó*, Mari *sösna*, *sasna*; Chu. *sisna* (contamination with *sis-* ‘to defecate’) (Starostin et al. 2003: 1237; Fedotov II: 77)

The root is preserved only in the Bulgar branch but is likely to be archaic.

**yilkī* ‘herd of horses’

OTurk. *yilqī*; Karakh. *yilqī* 1; MKypch. *yilqī* 2; Chag. *ilqī*; Tur. *yilqī* 3; Turkm. *yilqī* 3, 4; Kar. (K) *yilqī* 3, 4; Kum. *yilqī* 3; Bashk. *yilqī* 2; Nogh. *yilqī* 2, 3, 4; Az. *ilxī* 3; KBalk. *jilqī* 3; Kirgh. *jilqī* 2, 4; KKalp. *žilqī* 2, 4; Kaz. *žilqī* 2; Oyr. *yilqī*, *yilyī* 3; Uzb. *yilkī*; Uigh. *žilqī*; Khak. *čilyī* 3, 4; Chul. *čilyī*; Tuv. *čilyī*; Yak. *silyī* 2 (Tenišev et al. 2001: 444–445).

‘cattle 1, horse 2, herd of horses 3, year of the horse 4’.

**yogurt* ‘curdled milk’

OTurk. *yoyrot*, *yuryut*, *yuyrut* (OUygh.); Karakh. *yuyrut*, *yoğurt* (MK); Tur. *yoğurt*, *yourt*; Gag. *yürt*; Az. *yoğurt*; Turkm. *yoğurt*; MTurk. *yağurt* (Houts., AH); Uzb. *žuryət* (dial.); Kirgh. *žūrat*; KBalk. *žuwurt*, *žuwurt*, *zuwurt*; Kum. *yuwurt*; Nogh. *yuwirt*; SUig. *yoğurt*, *yuyurt*; Yak. *suorat*.

Possibly a derivative of *yogur-* ‘to knead’ or a homonymous verb meaning ‘to thicken, condense’ (Sevortjan et al. 1974–2003, 4:207–208).

**yunt* ‘horse’

OTurk. *yunt* (Orkh., OUygh.); Karakh. *yunt* (MK); Tur. *yont*; MTurk. *yunt* (Ettuhf.), *yunad* (AH); SUig. *yut*, *yot*; ? Yak. *sono-yos* ‘young horse’ (Starostin et al. 2003: 1523).

Cf. PSam. **yuntz* ‘horse’, which may be a borrowing from pTk (Dybo 2007: 143; vice versa in Sinor 1965: 312).

**yügen* ~ **üygen* ‘bridle’

Karakh. *yügön* (MK, IM); Tur. *oyan*; Az. *yüyän*; Turkm. *üyen*, *uyan*; MTurk. *uyan* (Pav. C.); Uzb. *yugan*; Uig. *yügän*; Krm. *iygen*, *yügen*; Tat. *yögän*; Bashk. *yügän*; Kirgh. *žügön*; Kaz. *žügen*; KBalk. *žügen*; KKalp. *žüwen*; Kum. *yügen*; Nogh. *yüwen*; SUig. *yuyin* (ЯЖУ); Khak. *čügen*; Shr. *čügen*; Oyr. *üygen*; Tv. *čüyen*; Chu. *yə^wven*; Yak. *ün* (Starostin et al. 2003: 878).

Farming and the Trans-New Guinea family

A consideration

Antoinette Schapper

KITLV / University of Cologne

The island of New Guinea, located to the north of Australia, is one of the world's major centres of early agriculture and plant domestication. At the same time, a large number of the languages of New Guinea and adjacent areas share a common origin and are believed to belong to a single language family, the Trans-New Guinea family. This paper presents a first attempt to apply the farming-language dispersal hypothesis to the New Guinea case. While the archaeological literature on early agriculture in New Guinea has focused mainly on taro, there is reason to doubt that taro was associated with the Trans-New Guinea expansion. In this paper, I instead consider the role of banana and sugarcane. The occurrence in many Trans-New Guinea languages of related terms for these two crops suggests that these were part of the “farming package” which fuelled the expansion of the family and its speakers.

Keywords: New Guinea, Papuan languages, Trans-New Guinea family, vegeculture

1. Introduction

At one time, New Guinea was regarded as a “passive recipient” (Neumann 2003) of domesticated plants and animals in the first instance from Southeast and East Asia and subsequently from Central and South America (in some cases via Europe). On New Guinea and the islands across Oceania, important plants of human use such as rice (*Oryza sativa*; Fuller et al. 2009; Barker, Hunt & Carlos 2011; Huang et al. 2012; Silva et al. 2015), betel nut (*Areca catechu*; Fairburn & Swadling 2005; Zumbroich 2008), and paper mulberry (*Broussonetia papyrifera*; Chang et al. 2015; González-Lorca et al. 2015; Matisoo-Smith 2015) as well as key animals including domestic pigs (*Sus scrofa*; Larson et al. 2007, 2010; Dobney, Cucchi & Larson 2008), dogs (*Canis familiaris*; Oskarsson et al. 2011, Greig, Walter & Matisoo-Smith 2015) and chickens (*Gallus gallus*, Gongora et al. 2008; Storey et al. 2012) trace themselves

back to the Asian mainland, though the precise origins of each on the continent is hotly disputed. Three crops from the Americas are significant staples in parts of New Guinea and its surrounds: sweet potato (*Ipomoea batatas*; Roullier et al. 2013), cassava (*Manihot esculenta*; Ellen, Soselisa & Wulandari 2012), and maize (*Zea mays*; Desjardins & McCarthy 2004).

Today, the tables have turned. New Guinea is no longer considered a backwater, but rather a center for some of the earliest plant domestication events in human history (Yen 1991; Fairburn 2005; Bourke 2009; Denham 2011). Among the many plants of regional significance to be part of the early cultivation practices of the peoples of New Guinea are: giant taro (*Alocasia macrorrhiza*, Nauheimer, Boyce & Renner 2012), taro (*Colocasia esculenta*, Lebot et al. 2004), greater yam (*Dioscorea alata*, Malapa et al. 2005), pandanus nuts (*Pandanus* spp., Haberle 1995), breadfruit (*Artocarpus altilis*, Zerega, Ragone & Motley 2004), canarium nuts (*Canarium* spp., Maloney 1996), and sago palm (*Metroxylon sagu*, Kjær et al. 2004). New Guinea was also the agricultural superhighway from which banana (*Musa* spp., Perrier et al. 2009) and sugarcane (*Saccharum officinarum*, Grivet et al. 2004), two of the world's most economically important crops, were launched on a truly global journey.

Recognition of New Guinea as a cradle of agriculture has entailed a more general reframing of demographic history in the region. Increasingly, the prevailing view of the region's prehistory from the 1970s onwards involving two waves of migration is being adjusted. The pioneer migration out of Africa brought the first anatomically modern humans into the Sunda-Sahul region from 60,000 BP to 40,000 BP (Tumonggor et al. 2013; Macaulay et al. 2005). Typically characterized as hunter-gatherers and beachcombers able to make short voyages, these first peoples moved eastward from Sunda across to Sahul (O'Connell & Allen 2012). These early migrations are considered to constitute the genetic source for several modern populations in the area, the so-called "Negrito" peoples of Southeast Asia and the Australo-Melanesian peoples of Melanesia and Australia (Balme 2013; O'Connor 2007). These people spoke what are believed to be antecedents of today's Papuan and Australian languages. The second wave of migration occurred southward out of Taiwan from the mid-Holocene (4000 BP to 3000 BP) across the Philippine and Indonesian archipelagos and over the top of New Guinea into the farthest reaches of the Pacific Ocean (Hill et al. 2007; Tabbada et al. 2010). This movement is associated with the development of outrigger canoes and pottery, and the dispersal of Austronesian languages (Bellwood 2002, 2011). The technologically superior, Austronesian-language speaking newcomers are thought to have variously overwhelmed, displaced and assimilated the early populations from the first migratory wave. In between these two significant in-migration events, we now know that far from being a time of stasis, there was a zone of activity around New Guinea, likely fuelled by agricultural innovation (Donohue & Denham 2010). Recurrent

population pulses westwards from New Guinea into Island Southeast Asia from 28,000 BP until the mid-Holocene are also strongly indicated by recent lineage-specific investigations of Y-chromosomal and autosomal DNA in the region (Hill et al. 2006; Soares et al. 2008; Jinam et al. 2013; Gomes et al. 2015). What we do not yet know is who these populations pulsing out of New Guinea were and what languages they spoke.

Around the world, many language families have had their wide dispersals credited to the adoption of agriculture by their early speakers. The “Farming/Language Dispersal Hypothesis” first proposed by Renfrew (1987) sees that the development of agriculture allowed groups to build up population numbers and expand themselves and their language into wider territories. The possible applicability of this model to New Guinea is suggested on the one hand by the recent realization of its role as a plant domestication center, and on the other hand, despite an initial peopling dating back more than 45,000 years, its linguistic landscape being dominated by the Trans-New Guinea (TNG) family. Trans-New Guinea is striking both for the large number of languages that it takes in and for the wide geographic area over which they are dispersed.

In this chapter, I present a first consideration of the Trans-New Guinea expansion as an instance of Farming/Language dispersal. In particular, I use historical, ethnobotanical and linguistic data to argue that sugarcane (*Saccharum officinarum*) and bananas (*Musa* spp.) are likely to have had a central place in any proto-Trans-New Guinea agricultural package. Sugarcane and banana are well suited to fueling rapid population expansions in that they often have broad altitudinal ranges, do not require intensive gardening including the irrigation or drainage that taro requires, and can be grown in almost any soil type. The social depth of sugarcane and banana use in New Guinea also attests their historical importance in Melanesian lifestyles. I will consider reconstructions of sugarcane and bananas across the full expanse of the Trans-New Guinea family and show that similar forms are recurrent, particularly at the extremes of the family’s geographical spread. This, I will suggest, indicates that sugarcane and banana must have been part of any agricultural package possessed by early Trans-New Guinea populations.

This chapter begins with a brief introduction to the Trans-New Guinea family (Section 2) and the nature of agriculture in New Guinea (Section 3). In Section 4 I argue that, whilst taro has had the most prominent place in the archaeological literature on early agriculture in New Guinea, it is problematic to associate its domestication with the Trans-New Guinea dispersal. In Section 5 I show that, in addition to domestication in New Guinea, cognate terms for sugarcane and banana have widespread dispersals in Trans-New Guinea languages. In Section 6 I discuss the place that sugarcane and banana had in the New Guinea diet and argue that

the niche occupied by them in New Guinea food supply makes them particularly suitable for population expansions.

2. Trans-New Guinea languages

The island of New Guinea is perhaps the linguistically most diverse region of the world. It is home to 800 languages that are designated as “Papuan”. This label does not refer to a single genetically cohesive group of languages. Rather it is a negative label that encompasses languages that are not members of the Austronesian language family and occur on or around the island of New Guinea.¹ Papuan languages are among the least well described in the world. In our current state of understanding, there are between 30 and 60 families and isolates that (so far) are not demonstrably related. Figure 1 presents a conservative, or “splitter”, picture of Papuan families. It is likely, however, that in the future, as more quality data become available and careful reconstructive work proceeds, many of these will be combined into larger genealogical entities. At this point, higher groupings remain tentative proposals.

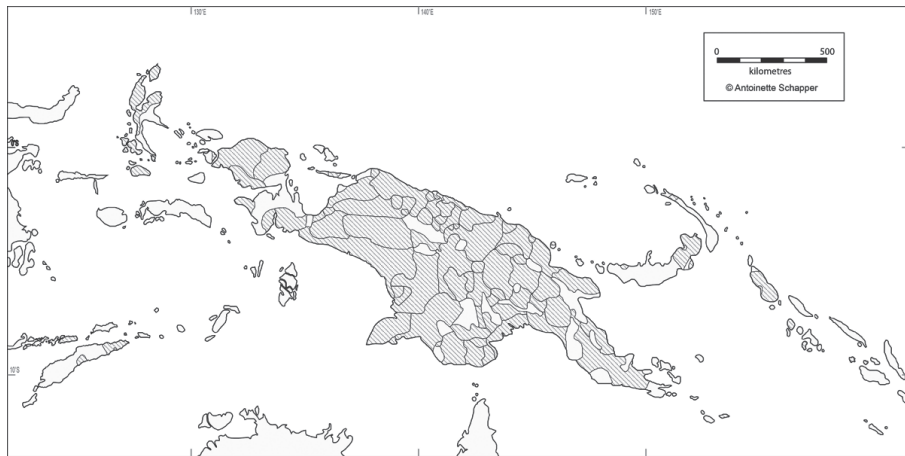


Figure 1. Papuan language families (shaded)

1. In much of the literature, emphasis is placed on “Papuan” languages not being part of the Austronesian family, and this has given rise to “non-Austronesian” as an alternative label to “Papuan.” This label is not employed here, as it does not carry with it the geographic restriction to the area of New Guinea which is so crucial to Papuanness. Austronesian languages are in fact in contact with members of multiple other (non-Papuan) language families, including Australian, Austro-Asiatic, Bantu, Tai-Kadai and Sino-Tibetan.

Of the several higher groupings of Papuan languages that have been proposed, the Trans-New Guinea family stands out for its credibility and its consequent endurance in the literature. The member families are located in the mountainous cordillera that runs for more than 2000 kilometers across New Guinea, but also extend into many lowland regions, particularly on the south coast of New Guinea, as well as to the island of Timor and its satellites several hundred kilometers to the west of New Guinea (Figure 2). While there is only partial agreement amongst linguists on the precise Trans-New Guinea membership or higher subgroupings of Trans-New Guinea languages, the family is thought to take in around 500 languages, making it potentially one of the larger families in the world.

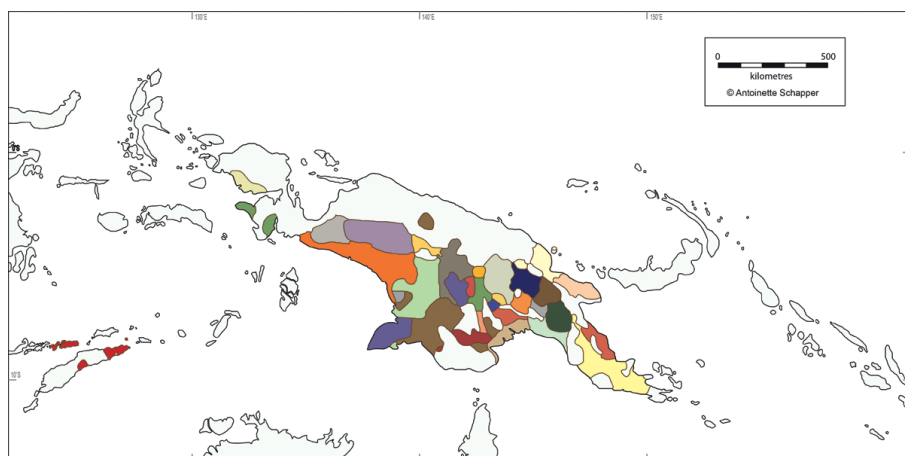


Figure 2. Posited Trans-New Guinea language families (after Ross 2005)

Trans-New Guinea is not a language family demonstrated by means of the Comparative Method in the way that the Indo-European or Austronesian families are. Trans-New Guinea is, at this stage at least, a hypothesis that seeks to account for shared lexical and grammatical phenomena in many Papuan languages by *positing*, rather than *proving*, a genealogical relationship between them. The foundations for Trans-New Guinea were laid by McElhanon & Voorhoeve's (1970) observations of lexical similarities between distant sets of languages in New Guinea. They proposed that the languages had a common origin, but did not attempt to identify regular sound correspondences. Wurm, Voorhoeve and McElhanon (1975) followed up with an expanded Trans-New Guinea "phylum" that presented the first attempt at defining Trans-New Guinea membership, but still did not apply rigorous historical methodologies. Instead, a language had to meet one or more of the following criteria to qualify as Trans-New Guinea:

1. reflexes of several forms belonging to small body of cognate sets, widely distributed among Trans-New Guinea languages;
2. reflexes of some tentatively reconstructed personal pronoun sets;
3. structural features in morphology and syntax that are common among Trans-New Guinea languages but rare/absent in non-Trans-New Guinea languages, such as switch reference morphology on medial verbs and body-tally counting.

On this basis, Trans-New Guinea was said to take in 491 “primary” languages and an additional 256 Papuan languages which were classified as “secondary” languages containing significant non-Trans-New Guinea substrates. Secondary Trans-New Guinea languages were typically located outside the central cordillera of New Guinea and lacked many of the Trans-New Guinea structural features and/or did not fully reflect the Trans-New Guinea pronominal paradigm.

Since then, much of the work on the Trans-New Guinea hypothesis has focused on the value of evidence from pronominal paradigms in determining Trans-New Guinea membership. Ross (2005) and Suter (2012) reconstruct pronouns and pronominal morphemes for proto-Trans-New Guinea. See Table 1 and Table 2 respectively for their reconstructed paradigms.

Table 1. Proto-Trans-New Guinea free pronouns

	Singular	Plural
1st person	*na	*ni ~ *nu
2nd person	*ŋga	*ŋgi
3rd person	*ya	*i

Table 2. Proto-Trans-New Guinea object prefixes

	Singular	Plural
1st person	*na-	–
2nd person	*ga-	–
3rd person	*wa-, Ø	*ya-

However, we still lack credible reconstructions of proto-Trans-New Guinea lexicon and phonology based on systematic agreements between languages belonging to distantly related subgroups. Pawley (1995, 1998, 2001, 2005, 2012) represent attempts at lexical and phonological reconstruction of proto-Trans-New Guinea. The results of this work are limited in their value; the top-down approach and the small sample of Trans-New Guinea families taken into account mean that reconstruction is heavily skewed towards the languages of the Eastern Highlands and is based to a significant extent on impressionistic observations of similarity rather

than completely systematic identification of correspondences. Accordingly, sorting lookalikes from related lexemes is near to impossible based on our current state of knowledge. Bottom-up reconstructive work for the many proposed Trans-New Guinea families is essential for achieving a rigorous proof of Trans-New Guinea and a proper understanding of its history. Whether there is even sufficient shared lexicon across the families for the application of the Comparative Method is unclear. The inadequate understanding of sound correspondences across Trans-New Guinea means that this paper will be limited to pointing out lexical resemblances in the agricultural domain. These are potential candidates for reconstruction to proto-Trans-New Guinea that will require fuller demonstration in the future.

Based on our current understanding of Trans-New Guinea, it appears that the family began somewhere in the Eastern Highlands and burst out in all directions across the expanse of New Guinea (Figure 3). The Eastern Highlands is the best candidate for the Trans-New Guinea homeland because it has the highest concentration of primary subgroups. As we move westward, Trans-New Guinea families become less diverse internally and cover larger territories than in the east, suggesting that those families are the result of more recent spread. This view is further supported by recent and ongoing bottom-up reconstructive work using the Comparative Method which has shown that numerous western families that were previously considered discrete, primary subgroups of Trans-New Guinea can in fact be grouped together into larger genealogical units: the Anim family which joins the Warkay-Bipim, Marind-Yaqay, the Lake Murray, Lower Fly River, and Inland Gulf families together (Usher & Suter 2015), the Ok-Awyu family in which the large central New Guinean Greater Awyu and Greater Ok families are grouped

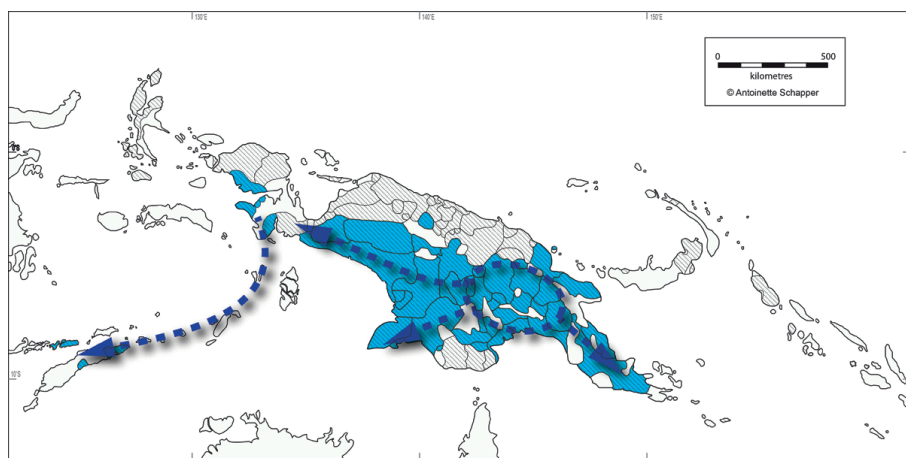


Figure 3. Posited homeland and direction of Trans-New Guinea expansions

together (based on reinterpreting the results of van den Heuvel & Fedden 2014), and the Greater West Bomberai family in which the Timor-Alor-Pantar languages are grouped with the languages of Bomberai peninsula (Usher and Schapper in preparation).

The timing of the Trans-New Guinea dispersal remains an open question. The widely divergent lexicons of Trans-New Guinea families mean that historical linguistics struggles to find the cognate vocabulary across the whole sweep of Trans-New Guinea families that is required to establish regular sound correspondences and therefore relatedness. This may be due to the Trans-New Guinea dispersal being at the temporal limits of the Comparative Method between 12,000 BP–8,000 BP (Rankin 2008: 207–208). Alternatively, it may be due to sociolinguistic factors in New Guinea that drive divergence in lexicon and, at the same time, convergence in typological features (see different descriptions of this setting in, e.g., Foley 1986: 283; Ross 1996, 2001; Thurston 1987, 1989). There is little to no archaeology that touches on the Trans-New Guinea dispersal on the New Guinea mainland. However, at the western fringe of the family some circumstantial dating of the Trans-New Guinea dispersal can be hazarded. The cuscus species *Phalanger orientalis* originates in New Guinea and is known to have been transported by humans to islands in eastern Indonesia (Heinsohn 2010). Direct dating of *P. orientalis* bones from recent archaeological excavations on Timor shows the presence of the marsupial on Timor starting at around 3000 BP (O'Connor 2015: 27). We may speculate that it was speakers of proto-Timor-Alor-Pantar, the common ancestor of the Timor-Alor-Pantar languages and a Trans-New Guinea family, who brought the cuscus with them when they departed from Bomberai peninsula. Moving back from the 3000 BP date for proto-Timor-Alor-Pantar, we may speculate on a mid-Holocene date (6000–10,000 BP) for the family as a whole, but little more.

3. Agriculture and its emergence in New Guinea

Traditionally, food production and supply systems in New Guinea were highly diverse (Bourke 2009). Most groups in New Guinea depended on combinations of starchy staples and tree crops. The starchy staples of New Guinea included taro, several species of yam, bananas, sago, and, more recently, sweet potato. Tree crops such as canarium nuts, okari, pandanus, candlenuts and breadfruit, added energy and oil to the diet. Other nutrients were obtained from vegetables, such as the leaves of various fig species (*Ficus dammaropsis*, *Ficus wassa*), of *Rungia klossii*, and of *Dicliptera papuana*, as well as edible grasses such as sugarcane, vegetable cane (*S. edule*), and nastus bamboo (*Nastus elastus*). Protein came primarily from hunting, fishing, and gathering of fungi and grubs. The introduction of pigs and other

domesticated animals from the mid- to late Holocene meant that access to protein would have become more reliable in the last millennia. Variations in reliance on these different food supplies depended on environmental differences in the areas inhabited by individual groups; the relative contributions of starch-rich plant and tree crops reflected altitude, while the significance of tree crops as well as hunting and gathering related to forest access.

Despite considerable variation in subsistence practices, vegetative, or asexual, propagation – rather than sexual reproduction from seed – characterizes plant cultivation in New Guinea in general (Denham 2011). The traditional starchy staples, sugarcane, as well as tree crops such as pandanus and banana are all vegetatively propagated. The predominance of vegetative propagation makes it, however, difficult to pinpoint the advent of agriculture in New Guinea. The best evidence we have for early plant exploitation in New Guinea comes from the Kuk Swamp site in the Waghi valley of the Eastern Highlands of Papua New Guinea. Finds made here indicate that primitive forms of plant cultivation, but not necessarily plant domestication, were practiced from 10,000 BP (Denham et al. 2003). Archaeobotanical evidence attests to the cultivation of bananas and sugarcane beginning between 6900–6400 BP (Denham et al. 2004). However, it is difficult on the basis of the archaeological record to differentiate plant management as practiced by hunter-gatherers from vegetative (agriculture by vegetative rather than seed propagation) of domesticated plants as practiced by farmers, let alone any of the many intermediate types that might exist. The swampy soils of much of the New Guinea highlands do not preserve plant remains well and, where they are preserved, the complex domestication histories of many New Guinean crops make it difficult to tell wild and cultivated forms apart.

So whilst New Guinea is increasingly accepted as a center of early plant exploitation and cultivation, many questions remain as to the process of domestication for many plants and the timing of the emergence of agriculture there. In the next section I discuss taro, a crop with a contested history, but one that is conventionally linked to early New Guinea agriculture.

4. Beyond taro

There is yet to be an attempt, systematic or otherwise, to connect the advent of agriculture in New Guinea to the dispersal of the Trans-New Guinea languages. The only remarks on the topic thus far are short and come from Pawley and Hammarström (forthcoming, echoing earlier brief remarks in Pawley 1998:200). They note the existence of widespread reflexes of a tentatively reconstructed proto-Trans-New Guinea form *ma ‘taro’ (or *mV in Pawley 2005). These authors remark

on the tenuous and circumstantial character of the link between agriculture and Trans-New Guinea, summing up the evidence as they see it as follows: “We know of no other widely distributed cognate sets for names of plants and their parts and for implements and processes associated with their cultivation.”

Associating early Trans-New Guinea peoples with taro cultivation is, however, problematic. Domesticated taro, *Colocasia esculenta*, is cultivated widely in New Guinea and Southeast Asia for its large edible corms. Wild taros are also found throughout this area. They have small corms of low starch content and are toxic due to the presence of high levels of calcium oxalate crystals. Domestication is thought to have taken place by prehistoric peoples selecting less acrid varieties with corms of higher starch content for vegetative propagation.

Where the domestication of taro took place remains unresolved. For the greater part of the 20th century, eastern India and Southeast Asia were held to be the most likely origin and domestication center of taro (Spier 1951; Yen & Wheeler 1968; Kuruvilla & Singh 1981). In the 1970s, Golson (1976, 1977) made archaeological finds at Kuk Swamp that agricultural infrastructure suitable for taro cultivation were present from 9000 BP. This gave rise to the alternative hypothesis that taro was domesticated in New Guinea (Yen 1980; Coates et al. 1988). Observations of wild taro populations in New Guinea also made New Guinea into a stronger candidate for the center of the plant’s domestication (Matthews 1990, 1991). The notion became further entrenched with fossil evidence of apparently taro-derived starch residues (Loy, Spriggs & Wickler 1992; Denham et al. 2003; Fullagar et al. 2006) and of taro pollen at Kuk Swamp (Haberle 1995).

There is little empirical support for New Guinea as a domestication center for taro. The most recent molecular and genetic studies are united in the view that the data is consistent with domestication occurring west of the Wallace Line in the Indo-Malayan region, and not in New Guinea. On the basis of an analysis of chloroplast and nuclear DNA diversity, Ahmed (2014) concludes that domesticated taro most likely originated in South to Southeast Asia. In his analysis, the haplotype grouping of all the taro cultivars and some of the wild taros in New Guinea is inconsistent with New Guinea being an independent primary center of taro domestication. Similarly, the analysis of molecular parameters in Chair et al. (2016) revealed that taro cultivars in Asia had the highest number of private alleles and Shannon index. On this basis, they also concluded that taro was most likely domesticated in South or Southeast Asia. Both studies explicitly dispute the results of Lebot & Aradhya (1991), the most widely cited molecular paper claiming New Guinea as the domestication center of taro, pointing out problems with the interpretation of the data. Both of these studies put forward Northeast India as the most likely domestication center.

If taro was not in fact domesticated in New Guinea, how do we then interpret the apparent taro finds at Kuk Swamp from 9000 BP? The taro found there may have been wild rather than domesticated. The natural distribution of wild taro species extends from South and Southeast Asia to New Guinea, northern Australia, the Solomon Islands and even New Caledonia (Matthews 1991). Neumann (2003) suggests that Kuk Swamp could represent a primitive form of cultivation of wild taro rather than domestication proper. This fits with the fact that even today much New Guinean cultivation is more accurately characterized as “plant management” rather than agriculture in the strict sense of the word (see Section 3). It is also consistent with the suggested early uses of wild taro in the Solomon Islands from 28,000 BP (Loy, Spriggs & Wickler 1992). The fossil starch residues and pollen spores found at Kuk Swamp also do not confirm the presence of domesticated taro as cultivated and wild forms cannot be distinguished from one another in the analyses.

The second problem with associating taro with the earliest speakers of Trans-New Guinea languages is linguistic. Numerous writers on New Guinea have observed the lexical stability/instability of nomenclatures for tuberous crops, not only diachronically but also synchronically. Scholars working in New Guinea frequently note difficulties in obtaining terminologies for tubers. For instance, Sillitoe (1980), who made a detailed study of sweet potato nomenclature among the Wola of Papua New Guinea (Angal Heneng [akh] TNG, Engan-Kewa-Huli family),² found a “startling lack of agreement” over the identification of cultivars. His study of taro nomenclature gave similar results, finding up to 50% discrepancy in the naming of plants. Milliken (nd) and Heider (1970) similarly remark on the variable nomenclatures for cultivars of sweet potato and taro among the Yali (Yali [yli] TNG, Dani family) and Dani (Mid Grand Valley Dani [dnt] TNG, Dani family) respectively. Over time, such variable naming practices will result in radical divergence in tuber terminologies even at the lowest levels of relatedness (high lexical replacement is a feature of New Guinea languages generally, see references in Section 2).

Simple cover labels like ‘taro’ and ‘yam’ (or their Indonesian/Tok Pisin equivalents), which dominate the literature and are often used in eliciting local names, typically subsume a number of species with distinct Linnaean names and thus serve to further complicate the identification of cognates. The result is very little comparability in tuber vocabularies over the New Guinea expanse. In his wide-ranging study of the dispersal of cultural vocabulary in eastern New Guinea, Dutton (1973, 1977) observes that the generic terms for the tuberous crops, sweet potato,

2. The names for people and their languages often differ considerably in linguistic and anthropological works. As such, throughout this chapter, for each group referred to I provide the Glottolog.org name for the group along with the ISO-639-3 codes, the Trans-New Guinea affiliation and the family affiliation.

taro and yam, are hugely variable and the etyma sets so interwoven with one another that it is largely impossible to determine the original referent. In the western half of New Guinea Hays (2005) finds similar variability in the tuber referent of members of the different etyma sets. Because of the cross-familial appearance and lack of reconstructability to even low-level families of most tuber terms, his study also highlights that diffusion rather than inheritance is the best explanation for the identified sets. Diffusion is also strongly indicated for the form *ma ~ *mV ‘taro’ which Pawley has attributed to proto-Trans-New Guinea. Hays (2005: 642–643) identifies a widespread etymon set #mao³ (including forms such as *ma*, *mao*, *mau* etc.) which primarily refers to ‘taro’. Members are found in numerous Trans-New Guinea and non-Trans-New Guinea families in New Guinea in a pattern that is inconsistent with inheritance.

In sum, whilst taro is widely cultivated in New Guinea, there are problems making an association between it and the Trans-New Guinea dispersal. The molecular and genetic evidence such as it currently exists does not support the view that taro was domesticated in New Guinea. This does not preclude domesticated taro from being part of early agriculture in New Guinea, as the archaeology suggests. It does, however, mean that the innovation of taro agriculture is unlikely to have been the sole driver of a Trans-New Guinea expansion out of the Eastern Highlands; peoples to the west would presumably have already been in possession of domesticated taro dispersed eastwards from the South(east) Asian mainland. Finally, the instability of tuber names and the apparent diffusion of tuber vocabulary across family boundaries means that reconstruction of a taro term to proto-Trans-New Guinea is fraught with difficulties.

In the following sections, I consider the evidence for sugarcane and banana as part of proto-Trans-New Guinea agricultural package. I suggest that there is *prima facie* a stronger case for associating these two crops with the Trans-New Guinea dispersal than for that of taro.

5. Proto-Trans-New Guinea sugarcane and banana reconstructed

Whilst taro nomenclature in New Guinea is problematic, the reconstruction of terms for sugarcane and banana is more of a prospect to a family of considerable time-depth such as Trans-New Guinea. Dutton (1973, 1977) noticed that terms for each of these crops do not display the frequent semantic changes and substitutions

3. Here I use # to mark a word that is not a reconstruction, but rather a generalization across forms in an etyma set that crosses language family/subgroup boundaries. * is reserved for words that are (thought to be) truly reconstructable to a proto-language on the basis of the Comparative Method.

that tuberous crops do. In Dutton's studies, generic names for sugarcane and banana rather show a high degree of stability within families, likely because they are not readily substituted by introduced foods that encroach upon their role in the diet. He observes that sugarcane especially stands out, with reflexes of protoforms never appearing as anything other than sugarcane.

Table 3. Posited Proto-Trans-New Guinea *jaBu 'sugarcane'⁴

Proto-Dagan *japu: Daga *jaup*; Mapena *jaup*; Gwedena *jopu*; Jimajima *jabu*; Maiwa *jup*.

Proto-Yareban *jawau (Usher nd): Abia, Doriri *java*; Yareba *jawau*.

Proto-Manubaran *(ar)epa (Usher nd): Doromu *afa* ~ *areha*; Gebi *areha*; Maranomou *arehæ*; Maria *areha*; Oiso *araxa*; Maiagolo *ara*; Uduri *araha*.

Kwalean: Mulaha *eva*.

Proto-Goilalan *japu: Biangi *jabi*, Weri *jap*; Kunimaipa *japu*.

Proto-Binandere *jowu (Smallhorn 2011): Guhu-Samane *japu*; Suena *jou*; Zia *jou*; Binandere *dou*; Aeka *jo*; Orokaiva *jobu*; Hunjara-Kaina Ke *jovu*; Ewage-Notu *jou*; Yegha *jou*; Gaina *joβu*; Baruga *joβu*; Korafe *joβu*.

Proto-Kainantu-Goroka *ja:pi (Usher pers. comm.): Proto-Gorokan **jap(i)* (after Scott 1978 for Proto-Eastern Central): Fore *ja:bú*; Gimi *zabi*; Alekano *zahiʔ*; Benabena *jáfi*; Dano *ávoso*; Tokano *ábosa*; Inoke-Yate *jofe*; Kamano *jáfóʔ*; Yagaria *éve*; Yaweyuha *yahu*; Siane *áfó*. Proto-Kainantu **ja:pi* (after Usher nd.): Proto-East Kainantu (Usher nd) **ja:pi*: Waffa *ja:ki*; Vaantura *ka:ʔv*; Afaqina *sa:k.v*. Proto-North Kainantu (Usher nd) **ja:ʔ*: Agardabi *ja:ʔ-i*; Akuna *ja:ʔ-i*.

Proto-Chimbu-Waghi *bo: Kuman *bo*; Maring *bo*; Salt-Yui *bo*; Melpa *po*; Ku Waru *po*.

Proto-Finisterre-Huon *ba: Gwahatike *bi*. Proto-Huon **ba* (Edgar Suter pers. comm.): Sialum *be*; Ono *ba*; Mape *be*; Naga *bo*; Kâte *bo*; Wamorâ *be*; Mâgobineng *be*; Sene *bac*; Momare *ba*; Migabac *ba*; Burum *bi*.

Proto-Turama-Kikorian *jou (after Usher nd): Kaser *jou*; Dugeme *io*; Barikiwa *jóù*; Mouwase *rou*.

Madang: Proto-Mabusu *ja (Ross 2014 cited in Greenhill nd); Proto-Rai Coast **juwa*: Biyom *juu*; Tauya *juwa*.

Dem (family level isolate) *mbè*

Teberan: Dadibi *gabo*, Folopa *hð*.

Proto-Bomberai *umbas: Mor *muas(a)*, Iha *mbes*, Mbaham *mbe:s*

Proto-Timor-Alor-Pantar *huba (after Schapper et al. 2014): Bunaq *up*; Nedebang *huφa*, West Pantar *habua*, Kaera *ub*, Klon *əba*, Abui *fā*, Sawila *ipua*, Wersing *upa*; Makasae, Makalero *ufa*, Fataluku *upa*, Oirata *uha*.

4. Alternatively, *juBa. *B is used here because it is uncertain at this stage whether this should be reconstructed as prenasalized or not, i.e., /b/ or /mb/.

In a survey of generic sugarcane terms across Trans-New Guinea families, I also observed striking similarities in terms across widely dispersed groups. Table 3 sets out reflexes of the proposed proto-Trans-New Guinea sugarcane term *jaBu. Reflexes extend from the eastern ‘tail’ of New Guinea through the Eastern Highlands, with a gap in the central New Guinea, before cognates resume in western New Guinea and spread all the way out to the distant far flung Timor-Alor-Pantar languages of Island South East Asia.

Generic banana terms across Trans-New Guinea families present a similar picture, with a single widely dispersed term identifiable. Table 4 sets out reflexes of the proposed proto-Trans-New Guinea banana term *mungo[l]. Whilst there are fewer reflexes of *mungo[l] than for *jaBu, they nonetheless have a strikingly wide distribution that is indicative of antiquity. Reflexes are found in the easternmost families of Trans-New Guinea, in pockets of the central highlands, and finally at the western fringe of the family in the Trans-New Guinea languages of Bomberai and insular Timor-Alor-Pantar languages.

Table 4. Posited Proto-Trans-New Guinea *mungo[l] ‘banana’

Proto-Yareban *moʔo (Usher nd): Moikodi *moʔo*; Aneme Wake *moʔo*; Nawaru *mo*; Yareba *mo*.

Proto-Mailuan *mogo: Binahari *um*; Bauwaki *moʔo*; Mailu *magari* ‘ripe banana’.

Proto-Madang *muŋgol (Ross 2014 cited in Greenhill nd): Aisi *maŋ*; Sirva, Mum, Apali *man*; Nend *aniŋ*; Sam (Wongbe Dialect) *mungol*; Anjam-Lalok *munŋe*; Kare *meaŋga*; Baimak *mu:ɡ*; Wagi, Mawan, Nake, Utu, Silopi Bagupi, Yoidik, Wamas, Rapping *mug*; Rempi *mu:k*; Garus *muk*; Mosimo, Murupi *mugu*; Saruga *mu:gu*; Samosa *mogu*; Watiwa *mɛ*.

Proto-Bosavi *magu: Bosavi *magu*; Kaluli *magu*; Dibiyaso *mase*; Onabasulu *mabu*;

Proto Duna-Bogaia *maga(C): Bogaya *mʌɡʌn*; Duna *makapo*.

Turama-Kikori: Rumu *kamiki*.

Proto-Bomberai *muŋga: Mor *moga*, Tanahmerah *moga*; Iha *nunguo*, Mbaham *munguo*.

Proto-Timor-Alor-Pantar *magol (after Schapper et al. 2014): Bunaq *mok*; Alor-Pantar: Nedebang *mai*, West Pantar *maggi*, Teiwa *muhui*, Kaera *mogoi*, Klolon *mgol*, Kamang *mori*, Wersing *mlul*; Eastern Timor: Makalero, Makasae, Fataluku *muʔu*, Oirata *mu*.

Donohue and Denham (2009) and Denham and Donohue (2009) identify a wide-spread banana term #muku in Austronesian languages in Island Melanesia, in particular in the arc of islands that runs between New Guinea and Flores (see the map and discussion in Schapper 2015 which presents more Austronesian terms than the original 2009 papers). Whilst they conclude that #muku is ultimately of Papuan origin, Donohue and Denham do not make the link between Trans-New Guinea and the dispersal of this term beyond New Guinea. Their survey of Papuan banana terms was superficial, only picking up “reflexes” of their #muku maverick form in

two Papuan families, Timor-Alor-Pantar languages and Yareban languages.⁵ A fuller survey made possible by the construction of the database described in Greenhill (2015), the observations of Blench (2016) and a burst of reconstructive work on Trans-New Guinea families means that Proto-Trans-New Guinea can be identified as the originator of the #muku set.

Bearing in mind the limitations on Trans-New Guinea reconstructions discussed in Section 2, the cognate sets for sugarcane and banana set out here are promising candidates for reconstruction to proto-Trans-New Guinea and, in turn, for a place for those plants in a proto-Trans-New Guinea agriculture package.

6. History of sugarcane and banana and their exploitation in New Guinea

The wide-ranging suite of domesticates from numerous plant taxa and varied manner of their exploitation found in New Guinea necessitates looking beyond the conventional cereal and tuberous crops typically discussed in the literature on early agriculture and frequently associated with language family expansions in Southeast Asia. We have already seen in Section 5 that there are promising etyma sets for sugarcane and banana that appear to go back to proto-Trans-New Guinea. The place of sugarcane and bananas in the diets of New Guinea people is therefore an issue that deserves attention. A proper appreciation of how and how much sugarcane and banana are, or at least were in the past, exploited by the people of New Guinea enhances the picture of a Trans-New Guinea expansion fueled by them. In this section, I suggest that our understanding of the history of sugarcane and banana means that both are good candidates for domestication and early cultivation in New Guinea.

Although their domestication histories are not entirely resolved, molecular and genetic data make clear that sugarcane and, at least, some species of bananas are likely to have undergone initial domestication on New Guinea. *Saccharum robustum*, the wild precursor of the domesticated *Saccharum officinarum*, is found only on New Guinea and nearby islands, placing sugarcane's domestication clearly east

5. The one non-Trans New Guinea Papuan language which Donohue and Denham (2009) claim to have a #muku “reflex” is Tehit, with *oga* ‘banana’. However, it is by no means certain that the Tehit form belongs to the #muku set. Tehit *oga* ‘banana’ is the only member of the whole of Donohue & Denham’s (2009) #muku set that has lost the initial nasal, a feature which makes it suspicious. It could just as easily be seen as a member of the #loka set (< #kalo < #qaRutay) which is found on nearby Halmahera, including in the related West Makian language (which Donohue and Denham 2009 erroneously listed as Austronesian). The choice by Donohue and Denham (2009) to give the Timor-Alor-Pantar languages a “West Papuan” classification, although that theory has not been given any credence by Papuanists for decades, seems calculated to exaggerate the link between the Timor-Alor-Pantar and Tehit banana terms.

of the Wallace Line (Grivet et al. 2004). With many more cultivars and identified (sub-)species, bananas have more complex domestication and dispersal pathways than sugarcane. New Guinea, however, is the center for some cultivars including *Musa acuminata* ssp. *banksii* and *Musa troglodytarum*, and represents one of the most likely locations of earliest banana domestication (De Langhe et al. 2009; Perrier et al. 2011).



Photo 1. Sugarcane garden in the highlands inland from Rigo in Southeast Papua New Guinea in the area where TNG languages are spoken. Taken in 1928 by the Sugar Expedition to the Territories of Papua and New Guinea, organised by the United States Department of Agriculture. © Smithsonian Institution.

In New Guinea and surrounds, sugarcane is consumed by sucking the juice from the chewed cane, while the banana fruit is eaten both cooked and raw, depending on the type. The relative importance of sugarcane and bananas has declined in the last century. Nonetheless, still today bananas are grown by 96% of the rural Papua New Guinean population and it remains the most important food crop for 9% of them and an important food for a further 32% (Bourke & Allen 2009: 195). Sugarcane is

grown by 99% of rural Papuan New Guineans and it is estimated that as much as 35 kilograms are chewed per person each year (Bourke & Allen 2009: 200).

Sugarcane and bananas are more versatile than other crops such as taro and sago in terms of where they will grow and the conditions they require. Both sugarcane and bananas have broad altitudinal ranges, being grown from sea level up into the highlands above 2000 m. They can be grown in almost any soil types so long as there is adequate drainage. In New Guinea, they are often planted in spent gardens. For example, amongst the Telefolmin (Telefol [tif], TNG, Ok family) “[w]hen the soil is nearly depleted of its fertility the garden is given over to banana or sugar cane” (Schuurkamp 1995: 107). Needing between eight or nine months, they are the last crops in a garden to become ripe (Clarke 1971: 163). At the same time, both sugarcane and banana will yield food even in times of drought and do not require the same kind of intensive tending as taro or sweet potato.

In descriptions of gardens in New Guinea, sugarcane and bananas are near ubiquitous, typically appearing alongside a tuberous staple. For instance, amongst speakers of Siane ([snp] TNG, Goroka family), Jentsch and Doetsch (1986: 285–286) observed: “Vorwiegend werden Süßkartoffeln, Bananen und Zuckerrohr angebaut ... An zweiter Stelle der Häufigkeiten der Anbauarten [nach Süßkartoffelanbau] flogt Zuckerrohr, an dritter Banane. Andere Anbauarten folgen in weitem Abstand.” In his intensive agricultural study of the Raiapu Enga ([enq] TNG, Engan family), Waddell (1969) found that, with the exception of gardens given over to sweet potato and taro monoculture, all mixed gardens contained banana plants and three quarters sugarcane. In many parts of New Guinea we also have records of banana and sugarcane monoculture gardens. Amongst the Nokopo (Yopno [yut] TNG, Finisterre-Huon family), Kocher Schmid (1991: 83–86) records that “[bananas] figure prominently in the Nokopo diet ... Their cultivation is given especially careful attention and they occupy a privileged position in the cropping system ... Bananas are cultivated in gardens on their own, ...”. In 1928 Brandes also observed large gardens given over entirely to sugarcane with large scaffolds supporting the cane (Photo 1). Daniels and Daniels (1993) bring together many more observations of sugarcane monoculture in New Guinea.

Understanding sugarcane’s central role in traditional subsistence in New Guinea is particularly important because, unlike bananas, it is not known to constitute a staple of the diet in other parts of the world. It is, however, often named as a main contributor to the New Guinean diet. Petermann (1915: 25) notes of the Weri speakers [wer] TNG, Goilalan family: “Wie die Waria, von denen sie sich auch sonst nicht weiter unterscheiden, leben auch die Wate der Hauptsache nach von Jams und Zuckerrohr.” Similarly, amongst the Marind ([mrz] TNG, Anim family) “[d]ie Hauptnahrung ... bestand aus Süßkartoffeln und Zuckerrohr, ...” (Luyken & Jansen 1960: 145). Yet, the importance of sugarcane to the diet has typically been

underestimated by causal observers. It is repeatedly discounted as only a snack food, ignoring the fact that sugarcane is one of the few plants that stores its carbohydrate as sucrose. However, its significant contribution to the overall diet is recognised by scholars particularly studying subsistence in New Guinea. Some examples follow.



Photo 2. Wiru ([wiu], TNG, family level isolate) people with lengths of sugarcane. Each guest receives lengths of sugarcane indicating a promise to give a portion of pork (Strathern and Stewart 1999). © The Pamela J. Stewart and Andrew J. Strathern Archive.

Eipomek [eip] TNG, Mek family:

Das Zuckerrohr wurde bei einer Stärke von ca. 2,5–3,5 cm geerntet und nur roh verzehrt. Obgleich es zu den wichtigsten Nahrungspflanzen im Malingdam-Bereich gehörte, war es kaum ein Bestandteil der Mahlzeiten. Man kaute allein die mehr oder weniger süßlichen, hartfaserigen oder weicheren Stengelstücke zwischenzeitlich, besonders im Gartenland und auf längeren Wegen.

(Koch 1984: 96)

Angal Heneng [akh] TNG, Engan-Kewa-Huli family:

The Wola think of sugar cane as a source of refreshment rather than as a food, which they may enjoy at any time of day, not only at meal times. It figures significantly in their diet through such snacks.

(Sillitoe 1983: 88)

Yopno [yut] TNG, Finisterre-Huon family:

Sugar cane is consumed in considerable amounts during the day while people work in their gardens. It is valued not only because of its easily-available contents of carbohydrate but also as a source of liquid. In the main garden area there is no other source of liquid available to people as there are no permanent creeks or springs.

(Kocher Schmid 1991:95)

It is clear from these quotes that sugarcane is an important source of calories in indigenous diets as well as water, making it an ideal food for away from home.

Sugarcane and banana are also frequently noted as a food of social import in New Guinea. Across the highlands, special plots of sugarcane and bananas are grown near houses for use in ceremonies and rituals (Brown 1978: 151). Sugarcane is of particular significance as the first thing offered to guests on their arrival (see Photo 2). Baruya ([byr], TNG, Angan family) law makes special provision for the cutting of sugarcane when it comes to guests: “Do not cut down the sugarcanes [your husband] has planted without his permission, unless guests pay a visit. Then you should rush to cut them down and give them to drink” (Godelier 1986: 43). Neglecting to grow sugarcane for guests is a gross social transgression and a slight on the manhood of the gardener amongst the Edolo ([etr], TNG, Bosavi family; Herdt 1999: 43 & 111). The social depth of sugarcane use across New Guinea in particular speaks to its long importance in New Guinean societies.

In short, both sugarcane and banana constitute highly adaptable crops that can be planted almost in any environment to produce consumables without intensive gardening. In the past, they were not just supplementary foods, but high-calorie staples, and in the case of sugarcane, a source of moisture while mobile. These characteristics, I suggest, make sugarcane and bananas ideal crops for highly mobile peoples moving across highlands and valleys, such as the early speakers of proto-Trans-New Guinea must have been, expanding in all directions out of the Eastern Highlands.

7. An agricultural package for Trans-New Guinea

There is no doubt that one of the world’s major centers of early agriculture and plant domestication lies in New Guinea. There is also a growing consensus that a large number of the languages of New Guinea and adjacent areas share a common origin and constitute a single language family, the Trans-New Guinea Phylum. Yet no substantial attempt has up to now been made to link these two observations using the Farming/Language Dispersal Hypothesis.

In this paper I have proposed the elements of an application of that hypothesis to the New Guinea case. The occurrence in many Trans-New Guinea languages of related terms for two crops, sugarcane and banana, suggests that these were part of the “farming package” which fuelled the expansion of the family and its speakers. Both are major crops of great contemporary cultural and economic importance throughout the Papuan language area. The archaeological literature on early agriculture in New Guinea has focused mainly on a different crop, taro. At this stage, however, the linguistic evidence that taro was associated with the Trans-New Guinea expansion appears much weaker than in the cases of banana and sugarcane.

The study of the Trans-New Guinea family is in its infancy. We know little about the constituency of Trans-New Guinea and the Comparative Method has yet to be applied across the board in defining its putative higher-order subgroups. Given the wide-ranging suite of domesticates from numerous plant taxa found in New Guinea, future work will likely add more elements to the proto-Trans-New Guinea agriculture package.

Acknowledgements

Funding from the Netherlands Organisation for Scientific Research VENI project “The evolution of the lexicon. Explorations in lexical stability, semantic shift and borrowing in a Papuan language family” and from the Volkswagen Stiftung DoBeS project “Aru languages documentation” is gratefully acknowledged. Thanks to Timothy Usher and Edgar Suter for providing me with the relevant reconstructions from the Trans-New Guinea families they work on. The paper was significantly improved by the patient comments of Andy Pawley, Malcolm Ross, the volume’s editors and Laurent Sagart. Extra special thanks goes to David Henley who spent many hours over many years discussing aspects of agriculture in Southeast Asia with me. All errors are my own.

References

- Ahmed, Ibrar. 2014. Evolutionary dynamics in taro (*Colocasia esculenta* L.). PhD dissertation, Massey University, New Zealand.
- Balme, Jane. 2013. Of boats and string: The maritime colonisation of Australia. *Quaternary International* 285: 68–75. doi:10.1016/j.quaint.2011.02.029
- Barker, Graeme, Hunt, Christopher & Carlos, Jane. 2011. Transitions to farming in Island Southeast Asia: Archaeological, biomolecular and palaeoecological perspectives. In *Why Cultivate? Anthropological and Archaeological Approaches to Foraging-Farming Transitions in Southeast Asia*, Graeme Barker & Monica Janowski (eds), 61–74. Cambridge: McDonald Institute for Archaeological Research.

- Bellwood, Peter. 2002. Farmers, foragers, languages, genes: The genesis of agricultural societies. In *Examining the Farming/Language Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 17–28. Cambridge: McDonald Institute for Archaeological Research, Cambridge University.
- Bellwood, Peter. 2011. Holocene population history in the Pacific Region as a model for world-wide food producer dispersals. *Current Anthropology* 52: 363–378. doi:10.1086/658181
- Blench, Roger. 2016. Things your classics master never told you: a borrowing from Trans New Guinea languages into Latin. Ms.
- Bourke, R. Michael. 2009. History of agriculture in Papua New Guinea. In *Food and Agriculture in Papua New Guinea*, R. Michael Bourke & Tracy Harwood (eds), 10–26. Canberra: ANU Press.
- Bourke, R. Michael & Allen, Bryant. 2009. Village food production systems. In *Food and Agriculture in Papua New Guinea*, R. Michael Bourke & Tracy Harwood (eds), 193–269. Canberra: ANU Press.
- Brown, Paula. 1978. *Highland Peoples of New Guinea*. Cambridge: CUP.
- Chair, Hana *et al.* 2016. Genetic diversification and dispersal of taro (*Colocasia esculenta* (L.) Schott). *PLoS One* 11 (6) :e0157712. doi:10.1371/journal.pone.0157712
- Chi-Shan Chang, Hsiao-Lei Liu, Ximena Moncada, Andrea Seelenfreund, Daniela Seelenfreund, & Kuo-Fang Chung. 2015. A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry. *PNAS* 112 (44): 13537–13542; published ahead of print October 5, 2015. doi:10.1073/pnas.1503205112
- Clarke, William C. 1971. *Place and People: An Ecology of a New Guinean Community*. Canberra: Australian National University Press.
- Coates, David J., Yen, Douglas E. & Gaffey, P. M. 1988. Chromosome variation in taro, *Colocasia esculenta*: Implications for origin in the Pacific. *Cytologia* 53: 551–560. doi:10.1508/cytologia.53.551
- Daniels, John & Daniels, Christian. 1993. Sugarcane in prehistory. *Archaeologia Oceania* 28: 1–7. doi:10.1002/j.1834-4453.1993.tb00309.x
- De Langhe, Edmond, Vrydaghs, Luc, de Maret, Pierre, Perrier, Xavier & Denham, Tim. 2009. Why bananas matter: An introduction to the history of banana domestication. *Ethnobotany Research and Applications* 7: 165–177. doi:10.17348/era.7.0.165-177
- Denham, Timothy & Donohue, Mark. 2009. Pre-Austronesian dispersal of banana cultivars West from New Guinea: linguistic relics from Eastern Indonesia. *Archaeology in Oceania* 44: 18–28. doi:10.1002/j.1834-4453.2009.tb00041.x
- Denham, Timothy. 2011. Early agriculture and plant domestication in New Guinea and Island Southeast Asia. *Current Anthropology* 52 (4): 379–395. doi:10.1086/658682
- Denham, Timothy *et al.* 2003. Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. *Science* 301: 189–193. doi:10.1126/science.1085255
- Denham, Timothy, Haberle, Simon & Lentfer, Carol. 2004. New evidence and interpretations for early agriculture in highland New Guinea. *Antiquity* 78: 839–857. doi:10.1017/S0003598X00113481
- Desjardins, Anne E. & McCarthy, Susan A. 2004. Maize in southeast Asian archipelago and Australia. In *Milho, Makka and Yu Mai: Early Journeys of Zea Mays to Asia*, Anne E. Desjardins & Susan A. McCarthy (eds). Online publication, National Agricultural Library. <https://www.nal.usda.gov/research/maize/chapter6.shtml>

- Dobney, Keith, Cucchi, Thomas & Larson, Greger. 2008. The pigs of Island Southeast Asia and the Pacific: New evidence for taxonomic status and human-mediated dispersal. *Asian Perspectives* 47 (1): 59–74. doi:10.1353/asi.2008.0009
- Donohue, Mark & Denham, Tim. 2009. Banana (*Musa* spp.) domestication in the Asia-Pacific region: Linguistic and archaeobotanical perspectives. *Ethnobotany Research & Applications* 7: 293–332. doi:10.17348/era.7.0.293-332
- Donohue, Mark & Denham, Tim. 2010. Farming and language in Island Southeast Asia. *Current Anthropology* 51: 223–256. doi:10.1086/650991
- Dutton, Tom. 1973. “Cultural” items of basic vocabulary in the Gulf and other districts of Papua (Part 1: Foodstuffs and Associated Agricultural Terms). In *The Linguistic Situation in the Gulf District and Adjacent Areas, Papua New Guinea*, Karl J. Franklin (ed.), 411–607. Canberra: Pacific Linguistics.
- Dutton, Tom. 1977. The distribution of cultural vocabulary in Papua. In *New Guinea Area Languages and Language Study, Vol. 3: Language, Culture, Society, and the Modern World*, Stephen A. Wurm (ed.), 51–100. Canberra: Pacific Linguistics.
- Ellen, Roy, Soselisa, Hermien L. & Wulandari, Asri Peni. 2012. The biocultural history of *Manihot esculenta* in the Moluccan Islands of Eastern Indonesia: Assessing evidence for the movement and selection of cassava germplasm. *Journal of Ethnobiology* 32 (2): 157–184. doi:10.2993/0278-0771-32.2.157
- Fairbairn, Andrew. 2005. An archaeobotanical perspective on plant-use practices in lowland northern New Guinea. *World Archaeology* 37: 487–502. doi:10.1080/00438240500395748
- Fairbairn, Andrew & Swadling, Pamela. 2005. Re-dating mid-Holocene betelnut (*Areca catechu* L.) and other plant use at Dongan, Papua New Guinea. *Radiocarbon* 47(3): 377–382. doi:10.1017/S0033822200035153
- Foley, William A. 1986. *The Papuan Languages of New Guinea*. Cambridge: CUP.
- Fullagar, Richard, Field, Judith, Denham, Tim & Lentfer, Carol. 2006. Early and mid-Holocene processing of taro (*Colocasia esculenta*) and yam (*Dioscorea* sp.) at Kuk Swamp in the highlands of Papua New Guinea. *Journal of Archaeological Science* 33: 595–614. doi:10.1016/j.jas.2005.07.020
- Fuller, Dorian Q. *et al.* 2009. The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323: 1607–1610. doi:10.1126/science.1166605
- Godelier, Maurice. 1986. *The Making of Great Men. Male Domination and Power among the New Guinea Baruya*. Cambridge: CUP.
- Golson, Jack. 1976. Archaeology and agricultural history in the New Guinea Highlands. In *Problems in Economic and Social Archaeology*, G. de G. Sieveking, I. A. Longworth & K. E. Wilson (eds), 201–220. London: Duckworth.
- Golson, Jack. 1977. No room at the top: Agricultural intensification in the New Guinea Highlands. In *Sunda and Sahul*, J. Allen, J. Golson & R. Jones (eds), 601–638. London: Academic Press.
- Gomes, Sibylle M. *et al.* 2015. Human settlement history between Sunda and Sahul: A focus on East Timor (Timor-Leste) and the Pleistocenic mtDNA diversity. *BMC Genomics* 16: 70. doi:10.1186/s12864-014-1201-x
- Gongora, Jaime *et al.* 2008. Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences, USA* 105 (30): 10308–13. doi:10.1073/pnas.0801991105
- Gonzalez-Lorca, J. *et al.* 2015. Ancient and modern introduction of *Broussonetia papyrifera* ([L.] Vent.; Moraceae) into the Pacific: Genetic, geographical and historical evidence. *New Zealand Journal of Botany* 53: 75–89. doi:10.1080/0028825X.2015.1010546

- Greenhill, Simon. J. nd. TransNewGuinea.org website. <<http://transnewguinea.org/>>
- Greenhill, Simon J. 2015. TransNewGuinea.org: An online database of New Guinea Languages. *PLoS ONE* 10 (10): e0141563. doi:10.1371/journal.pone.0141563
- Greig, Karen, Walter, Richard & Matisoo-Smith, Elizabeth A. 2015. Dogs and people in Southeast Asia and the Pacific. In *The Routledge Handbook of Bioarchaeology in Southeast Asia and the Pacific Islands*, M. Oxenham & H. Buckley (eds), 462–482. Abingdon: Routledge.
- Grivet, Laurent, Daniels, Christian, Glaszman, Jean-Christophe & D'Hont, Angelique. 2004. A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research and Applications* 2: 9–17. doi:10.17348/era.2.0.9-17
- Haberle, Simon G. 1995. Identification of cultivated Pandanus and Colocasia in pollen records and the implications for the study of early agriculture in New Guinea. *Vegetation History and Archaeobotany* 4: 195–210. doi:10.1007/BF00235751
- Hays, Terence E. 2005. Vernacular names for tubers in Irian Jaya: implications for prehistory. In *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, A. Pawley, R. Attenborough, J. Golson & R. Hide (eds), 625–670. Canberra: Pacific Linguistics.
- Heider, Karl G. 1970. *The Dugum Dani. Papuan Culture in the Highlands of West New Guinea*. Chicago IL: Aldine.
- Heinsohn, Thomas E. 2010. Marsupials as Introduced Species. In *Altered Ecologies: Fire, Climate and Human Influence on Terrestrial Landscapes*, Simon Haberle, Janelle Stevenson & Matthew Prebble (eds.), 133–176. Canberra: Terra Australis, ANU Epress.
- Herd, Gilbert. 1999. *The Sambia: Ritual, Sexuality, and Change in Papua New Guinea*. Chicago IL: University of Chicago Press.
- Hill, Catherine *et al.* 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Molecular Biology and Evolution* 23: 2480–2491. doi:10.1093/molbev/msl124
- Hill, Catherine *et al.* 2007. A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80: 29–43. doi:10.1086/510412
- Huang, Xuehui *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501. doi:10.1038/nature11532
- Jentsch, Thunar & Doetsch, Rainer. 1986. *Keman, eine Siedlung im Hochland von Papua-Neuguinea: Bestandsaufnahme und Untersuchung des kulturellen Wandels*. Berlin: Dietrich Reimer.
- Jinam, Timothy A., Phipps, Maude E., Saitou, Naruya & The Hugo Pan-Asian SNP Consortium. 2013. Admixture patterns and genetic differentiation in negrito groups from West Malaysia estimated from genome-wide SNP Data. *Human Biology* 85: 173–188. doi:10.3378/027.085.0308
- Kjær, Anders, Barfod, Anders S., Asmussen, Conny B. & Seberg, Ole. 2004. Investigation of genetic and morphological variation in the sago palm (*Metroxylon sagu*; *Arecaceae*) in Papua New Guinea. *Annals of Botany* 94: 109–117. doi:10.1093/aob/mch112
- Koch, Gerd. 1984. *Malingdam: Ethnographische Notizen über einen Siedlungsbereich im oberen Eipomek-Tal des zentralen Bergland von Irian Jaya (West-Neuguinea), Indonesien*. Berlin: Dietrich Reimer.
- Kocher Schmid, Christin. 1991. *Of People and Plants: a Botanical Ethnography of Nokopo Village, Madang and Morobe Provinces, Papua New Guinea* [Ethnologisches Seminar der Universität und Museum für Volkerkunde]. Basel: Wepf & Co.
- Kuruvilla, K. M. & Singh, Avtar. 1981. Karotypic and electrophoretic studies on taro and its origin. *Euphytica* 30: 405–413. doi:10.1007/BF0034004

- Larson, Greger *et al.* 2007. Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Science, USA* 104(12): 4834–9. doi:10.1073/pnas.0607753104
- Larson, Greger *et al.* 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences USA* 107(17): 7686–7691. doi:10.1073/pnas.0912264107
- Lebot, Vincent & Aradhya, Mallikarjuna K. 1991. Isozyme variation in taro (*Colocasia esculenta* (L.) Schott) from Asia and Oceania. *Euphytica* 56: 55–66.
- Lebot, Vincent *et al.* 2004. Characterisation of taro (*Colocasia esculenta* (L.) Schott) genetic resources in Southeast Asia and Oceania. *Genetic Resources and Crop Evolution* 51: 381–392. doi:10.1023/B:GRES.0000023453.30948.4d
- Loy, Thomas H., Spriggs, Matthew & Wickler, Stephen. 1992. Direct evidence for human use of plants 28,000 years ago: starch residues on stone artefacts from the northern Solomon Islands. *Antiquity* 66: 898–912. doi:10.1017/S0003598X00044811
- Luyken, R. & Jansen, A. A. 1960. Nutritional state of the Marind-Anim tribe in South New Guinea. *Tropical Geographical Medicine* June 12: 145–148.
- Macaulay, Vincent *et al.* 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034–1036. doi:10.1126/science.1109792
- Malapa, Roger, Arnau, Gemma, Noyer, Jean-Louis & Lebot, Vincent. 2005. Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution* 52: 919–929. doi:10.1007/s10722-003-6122-5
- Maloney, Bernard K. 1996. Canarium in the Southeast Asian and Oceanic archaeobotanical and pollen records. *Antiquity* 70 (270): 926–933. doi:10.1017/S0003598X00084180
- Matthews, Peter Joseph. 1990. The Origins, Dispersal and Domestication of Taro. PhD dissertation, The Australian National University, Canberra.
- Matthews, Peter Joseph. 1991. A possible tropical wildtype taro: *Colocasia esculenta* var. *aquatilis*. *Indo-Pacific Prehistory Association Bulletin* 11: 69–81. doi:10.7152/bippa.v11i10.11374
- Matisoo-Smith, Elizabeth A. 2015. Tracking Austronesian expansion into the Pacific via the paper mulberry plant. *Proceedings of the National Academy of Sciences USA* 112(44): 13432–13433. doi:10.1073/pnas.1518576112
- McElhanon, Kenneth A. & Voorhoeve, Clemens L. 1970. *The Trans-New Guinea Phylum: Explorations in Deep-level Genetic Relationships*. Canberra: Pacific Linguistics.
- Milliken, William. nd. An account of the ethnobotany of the Yali people in the highlands of New Guinea (West Papua). Royal Botanic Garden, Edinburgh. <<http://rbg-web2.rbge.org.uk/ethnobotany/Yali.pdf>>
- Nauheimer, Lars, Boyce, Peter C. & Renner, Susanne S. 2012. Giant taro and its relatives: A phylogeny of the large genus *Alocasia* (Araceae) sheds light on Miocene floristic exchange in the Malesian region. *Molecular Phylogenetics and Evolution* 63(1): 43–51. doi:10.1016/j.ympev.2011.12.011
- Neumann, Katharina. 2003. New Guinea: A cradle of agriculture. *Science* 301(5630): 180–181. doi:10.1126/science.1086677
- O'Connell, James F. & Jim Allen. 2012. The restaurant at the end of the universe: Modelling the Colonisation of Sahul. *Australian Archaeology* 74: 5–17.
- O'Connor, Sue. 2007. New evidence from East Timor contributes to our understanding of earliest modern human colonisation east of the Sunda Shelf. *Antiquity* 81: 523–35. doi:10.1017/S0003598X00095569

- O'Connor, Sue. 2015. Rethinking the Neolithic in Island Southeast Asia, with particular reference to the archaeology of Timor Leste and Sulawesi. *Archipel* 90: 15–47. doi:10.4000/archipel.362
- Oskarsson, Mattias C. R. *et al.* 2011. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proceedings of the Royal Society B* 2011. doi:10.1098/rspb.2011.1395
- Pawley, Andrew. 1995. C.L. Voorhoeve and the Trans New Guinea Phylum hypothesis. In *Tales from a Concave World. Liber Americum Bert Voorhoeve*, Connie Baak, Mary Bakker & Dick van der Meij (eds), 83–122. Leiden University, Projects Division, Department of Languages and Cultures of South-East Asia and Oceania.
- Pawley, Andrew. 1998. The Trans New Guinea hypothesis: a reassessment. In *Perspectives on the Bird's Head*, Jelle Miedema, C. Ode & R.A.C. Dam (eds), 655–690. Projects Division, Dept. of Languages and Cultures of Southeast Asia and Oceania, Leiden University. Editions Rodopi.
- Pawley, Andrew. 2001. The Proto Trans New Guinea obstruents. In *The Boy from Bundaberg: Essays in Melanesian Linguistics in Honour of Tom Dutton*, A. Pawley, M. Ross & D. Tryon (eds), 261–300. Canberra: Pacific Linguistics.
- Pawley, Andrew. 2005. Trans New Guinea Hypothesis: Recent research and its implications. In *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide (eds), 67–107. Canberra: Pacific Linguistics.
- Pawley, Andrew. 2012. How reconstructable is Proto-Trans New Guinea? Problems, progress, prospects. In *History, Contact and Classification of Papuan Languages*, Part One, H. Hammarström & W. van den Heuvel (eds), 88–164. Port Moresby: Linguistic Society of Papua New Guinea.
- Pawley, Andrew & Hammarström, Harald. Forthcoming. The Trans New Guinea. In *The Languages and Linguistics of the New Guinea Area. A Comprehensive Guide*, B. Palmer (ed.). Berlin: De Gruyter Mouton.
- Perrier, Xavier *et al.* 2009. Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobotany Research and Applications* 7: 199–216. doi:10.17348/era.7.0.199-216
- Perrier, Xavier *et al.* 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences USA* 108(28): 11311–11318. doi:10.1073/pnas.1102001108
- Petermann, August. 1915. *Dr. A. Petermann's Mitteilungen aus Justus Perthes' geographischer Anstalt*, Vol. 61. Gotha: Justus Perthes.
- Rankin, R. L. 2008. The Comparative Method. In *The Handbook of Historical Linguistics*, B. D. Joseph & R. D. Janda (eds.), 183–212. Blackwell Publishing Ltd, Oxford, UK. doi:10.1002/9780470756393.ch1
- Renfrew, Colin. 1987. *Archaeology and Language: the Puzzle of Indo-European Origin*. Cambridge: Jonathan Cape.
- Ross, Malcolm. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, Mark Durie & Malcolm Ross (eds), 180–217. Oxford: OUP.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic languages in North-West Melanesia. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds), 134–166. Oxford: OUP.
- Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In *Papuan Pasts. Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, A. Pawley, R. Attenborough, J. Golson & R. Hide (eds), 15–65. Canberra: Pacific Linguistics.

- Roullier, Caroline, Kambouo, Rosa, Paofa, Janet, McKey, Doyle & Lebot, Vincent. 2013. On the origin of sweet potato (*Ipomoea batatas* (L.) Lam.) genetic diversity in New Guinea, a secondary center of diversity. *Heredity* 110: 594–604. doi:10.1038/hdy.2013.14
- Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel* 90: 99–151. doi:10.4000/archipel.371
- Schapper, Antoinette, Huber, Juliette & van Engelenhoven, Aone. 2014. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In *Alor-Pantar Languages: History and Typology*, Marian Klamer (ed.), 99–154. Berlin: Language Science Press.
- Schuurkamp, Gerrit J. T. 1995. *The Min of the Papua New Guinea Star Mountains: A Look at Their Traditional Culture and Heritage*. Kempt Head Point, Nova Scoti: Ok Tedi Mining.
- Scott, Graham. 1978. *The Fore Language of Papua New Guinea*. Canberra: Pacific Linguistics.
- Sillitoe, Paul. 1980. Confusions in the classifications. How the Wola name their plants. *Ethnos* 45: 133–156. doi:10.1080/00141844.1980.9981196
- Sillitoe, Paul. 1983. *Roots of the Earth: Crops in the Highlands of Papua New Guinea*. Manchester: Manchester University Press.
- Silva, Fabio *et al.* 2015. Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. *PLoS ONE* 10 (9): e0137024. doi:10.1371/journal.pone.0137024
- Smallhorn, Jacinta. 2011. *The Binanderean Languages of Papua New Guinea*. Canberra: Pacific Linguistics.
- Soares, Pedro *et al.* 2008. Climate change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution* 25: 1209–1218. doi:10.1093/molbev/msn068
- Spier, Robert F. G. 1951. Some notes on the origin of taro. *Southwestern Journal of Anthropology* 7: 69–76. doi:10.1086/soutjanth.7.1.3628650
- Storey, Alice A. *et al.* 2012. Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures. *PLoS ONE* 7 (7): e39171. doi:10.1371/journal.pone.0039171
- Strathern, Andrew J. & Stewart, Pamela J. 1999. On the theory of stability and change in ritual: the legacy of Roy Rappaport. Review essay on Roy Rappaport's (1999) *Ritual and Religion in the Making of Humanity*. *Social Analysis* 43 (3): 116–121.
- Suter, Edgar. 2012. Verbs with pronominal object prefixes in Finisterre-Huon languages. In *History, Contact and Classification of Papuan Languages*, Part one, Harald Hammarström & Wilco van den Heuvel (eds), 23–58. Port Moresby: Linguistic Society of Papua New Guinea.
- Tabbada, Kristina A. *et al.* 2010. Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution* 27: 21–31. doi:10.1093/molbev/msp215
- Thurston, William R. 1987. *Processes of Change in the Languages of North-Western New Britain*. Canberra: Pacific Linguistics.
- Thurston, William R. 1989. How exoteric languages build a lexicon: Esoterogeny in West New Britain. In *VICAL 1, Oceanic Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, Ray Harlow & Robin Hooper (eds), 555–579. Auckland: Auckland: Linguistic Society of New Zealand.
- Tumonggor, Meryanne K. *et al.* 2013. The Indonesian archipelago: An ancient genetic highway linking Asia and the Pacific. *Journal of Human Genetics* 58: 165–73. doi:10.1038/jhg.2012.154
- Usher, Timothy. nd. New Guinea World Website. <<https://sites.google.com/site/newguineaworld/>>
- Usher, Timothy & Schapper, Antoinette. In preparation. The Greater West Bomberai family. *Oceanic Linguistics*.

- Usher, Timothy & Suter, Edgar. 2015. The Anim languages of Southern New Guinea. *Oceanic Linguistics* 54: 110–142. doi:10.1353/ol.2015.0003
- van den Heuvel, Wilco & Fedden, Sebastian. 2014. Greater Awyu and Greater Ok: Inheritance or contact? *Oceanic Linguistics* 53: 1–35. doi:10.1353/ol.2014.0008
- Waddell, Eric W. 1969. The Dynamics of a New Guinea Highlands Agricultural System. PhD dissertation, Australian National University.
- Wurm, Stephen A., Voorhoeve, Clemens L. & McElhanon, Kenneth. 1975. The Trans New Guinea phylum in general. In *New Guinea Area Languages, Vol. 1: Papuan Languages and the New Guinea Linguistic Scene*, Stephen A. Wurm (ed.), 299–322. Canberra: Pacific Linguistics.
- Yen, Douglas E. & Wheeler, Jocelyn M. 1968. Introduction of taro into the Pacific: The indications of the chromosome numbers. *Ethnology* 7: 259–267. doi:10.2307/3772891
- Yen, Douglas E. 1980. The Southeast Asian foundations of Oceanic agriculture: A reassessment. *Journal de la Société des Océanistes* 36: 140–147. doi:10.3406/jso.1980.3031
- Yen, Douglas E. 1991. Domestication: the lessons from New Guinea. In *Man and a Half: Essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*, Andrew Pawley (ed.), 558–569. Auckland: Polynesian Society.
- Zerega, Nyree J. C., Ragone, Diane & Motley, Timothy J. 2004. Complex origins of breadfruit (*Artocarpus altilis*, Moraceae): implications for human migrations in the Pacific. *American Journal of Botany* 91: 760–766. doi:10.3732/ajb.91.5.760
- Zumbroich, Thomas J. 2008. The origin and diffusion of betel chewing: A synthesis of evidence from South Asia, Southeast Asia and beyond. *E-journal of Indian Medicine* 1: 87–140.

The domestications and the domesticators of Asian rice

George van Driem

Institut für Sprachwissenschaft, Universität Bern, Switzerland

Rice genetics has now provided molecular evidence for three distinct domestications of Asian rice, giving rise to *ahu*, *indica* and *japonica* rice and subsequently involving the multidirectional introgression of favoured alleles between these three families of *Oryza sativa* cultivars. The phylogeography of Asian wild and cultivated rice species also permits inferences with regard to the likely geographical range within which these three domestication processes involving Asian cultivated rice unfolded. Evidence from linguistic palaeontology permits the identification of two language families whose linguistic ancestors pose the likeliest candidates for the earliest rice domesticators, Austroasiatic and Hmong-Mien. The linguistic arguments and population genetic evidence on Asian rice are assessed. Recent advances in palaeobotany as well as a number of currently prevalent misunderstandings in rice archaeology are discussed. Another set of evidence from linguistic palaeontology involving reconstructible etyma denoting megafauna in light of the early Holocene distribution of these megafaunal species provides a geographical indication for the location of the early Austroasiatic homeland. Furthermore, the molecular genetics of human populations are discussed in order to shed light on the prehistory and geography of the Austroasiatic, Hmong-Mien and other language families. Finally, a synthesis of the disparate sets of evidence is presented.

Keywords: rice (*Oryza sativa*), Hmong-Mien, Austroasiatic, phylogeography, preservation bias

1. Rice genetics and rice domestications

In 1883, the director of the botanical garden in Geneva, Alphonse-Louis-Pierre Pyrame de Candolle, argued that the origin of cultivated rice lay in China and that rice was introduced to India from China (1883: 285, 309–311). Later, Nikolaj Ivanovič Vavilov (1926) argued against a Chinese origin for rice and contended

instead that the origin of Asian rice lay in India, whence the crop had spread to China and Japan. The old controversy about the original homeland of cultivated rice persisted well into the early years of the new millennium. In the Himalayan handbook, I have recounted how this controversy has influenced historical linguistic discourse over the years (van Driem 2001: 324–327 et passim). One might like to think that the old polarisation of arguments had been rendered obsolete ever since the evidence of molecular genetics has been brought to bear on the resolution of the question.

Three principal populations of cultivated rice *Oryza sativa* are distinguished, comprising the families of cultivars known as *ahu*, *indica* and *japonica* rice. Whereas the latter two varieties are characterised by wet cultivation, *ahu* rice is cultivated on dry fields and terraces and is sometimes referred to imprecisely as “upland rice”. This dry land cultivar is known in Assamese as আহু *ahu*, in Nepali घैया as *ghaiyā* and in Bengali as আউশ *āuš*. The Assamese name *ahu* arguably provides the most apt candidate for an English name for this cultivar, both because this family of cultivars is most widespread throughout Assam and because the Assamese name *ahu* lends itself readily to being pronounced well in English. Neither the Nepali nor the Bengali name remain quite intact once uttered by someone who subjects the words to an English phonology. The Bengali name *āuš*, in particular, has the tendency to get unrecognisably transmogrified in the mouths of English speakers.

In the older literature before the turn of the millennium, *japonica* rice was often held to come from a wild precursor *Oryza rufipogon*, whereas *indica* rice was thought to derive from a wild precursor *Oryza nivara*. New research has not rendered this view entirely obsolete, but has instead refined our understanding of wild *Oryza rufipogon* as a highly diverse species that has long been undergoing a prolonged process of speciation. Rather, wild *nivara* rice can most accurately be considered to be an annual self-pollinating ecotype or subspecies of *rufipogon*, since these wild rice populations interbreed to a limited extent and therefore constitute a single internally diverse species complex. In the noughties, population genetic research based on the genome of wild and cultivated varieties of rice supported the novel hypothesis that Asian rice had been domesticated twice (Kovach et al. 2007; Sweeney & McCouch 2007; Kovach et al. 2009).

At one point, the mutation coding for a whiter grain pericarp (*rc*) changed the reddish seed of wild rice into the white seeds of modern rice. This gene is shared by the majority of rice cultivars, and the trait was held to have introgressed from *japonica* into both *indica* and *ahu* rice (Sweeney et al. 2007). Soon other parts of the tangled tale of rice domestication were unravelled. Although the *japonica* and *indica* cultivar families essentially derive from a single species of wild rice, the time of divergence of about 100,000 years calculated for the two distinct ancestral *rufipogon* subspecies from which the two cultivars had derived indicated

independent domestications. At the same time, although *ahu* rice was found to be genetically more closely affiliated to *indica* than to *japonica* rice, *ahu* rice appeared to have resulted from yet a third distinct domestication process (Londo et al. 2006). Subsequent genetic studies on Asian rice have corroborated these findings and identified the molecular footprints of three independent domestication events in different parts of Asia. Moreover, molecular evidence has demonstrated that the introgression of domesticated traits had occurred not just unidirectionally from *japonica* into *ahu* and *indica* rice, but multidirectionally from *ahu* and *indica* into *japonica* as well (McNally et al. 2009; Civián et al. 2015).

The prehistory of rice cultivation and rice domestication is convoluted as a direct consequence of the role played by human rice cultivators. The prehistory of rice cultivation involved three distinct domestications as well as the sustained cultural exchange of rice cultivar knowledge over time between the populations of early rice cultivators. The cultivation and domestication of the annual self-pollinating *nivara* ecotype of *Oryza rufipogon* led to the development of the *indica* cultivar of *Oryza sativa*, and for various reasons it is likely that this process may have transpired in the Brahmaputra river basin. In this area, Asian rice was long cultivated before it was domesticated through selective breeding by humans, and grain shattering cultivars are still cultivated to this day. Various rice species other than *Oryza sativa* that have generally been deemed to be wild likewise continue to be cultivated in Assam. An eastern domestication of a perennial swamp subspecies of *Oryza rufipogon* gave rise to the *japonica* cultivar of *Oryza sativa*. The mutation *sh4* led to the partial development of the abscission zone where the mature grain detaches from the pedicle, and the reduced brittleness of the rachides reduced grain shattering. Subsequently, human domestication also favoured genes coding for a whiter grain pericarp (*rc*) and erecter stalks (*Prog1*).

Several stages in the domestication of *indica* rice entailed the introduction of the traits *sh4*, *rc* and *Prog1* into the *nivara* gene pool through introgressive hybridisation, involving backcrossing with the *japonica* cultivar. The hill tracts surrounding the Brahmaputra river basin may have been where the domestication of *ahu* rice took place. The three domestication events which gave rise to modern rice cultivars took place long ago, and the relative popularity of many *japonica* strains today is likely to represent a secondary development on the grander time scale. Even subsequent to early cultivation, the exchange of rice cultivar knowledge between rice cultivating peoples persisted over time. The *javanica* cultivar has been demonstrated to represent a tropical variety of *japonica*, whereas a number of famous long-grained aromatic varieties, such as Indian *bāsmatī* rice, have likewise been shown to derive from *japonica* (Parsons et al. 1999; Garris et al. 2005). By contrast, Thai jasmine rice, for instance, has been shown to represent an *indica* variety, with the

fragrant allele of the betaine aldehyde dehydrogenase gene *BADH2* introduced by introgression (Kovach et al. 2009).

Once, a team of geneticists ventured to conjecture that the introgression of the white grain pericarp (*rc*) allele might be construed as possible evidence for a single domestication of rice between 13,500 and 8,200 years ago, which they ventured to situate in the Yangtze river basin (Molina et al. 2011). Remarkably, this conjecture was not supported by the team's own phylogenetic data. Rather, the geneticists in question explicitly deferred to arguments advanced by archaeologists anxious to see the lower Yangtze recognised as the unique home of rice domestication (Fuller & Qin 2010; Fuller et al. 2010). On the basis of their own molecular findings, the geneticists were unable to exclude that *indica* and *japonica* had been independently domesticated, highlighting the possibility "that both *indica* and *japonica* originated from highly differentiated *Oryza rufipogon* gene pools that were not sampled" (Molina et al. 2011: 5). In fact, their evidence suggested that the wild *rufipogon* populations of the Indian subcontinent and mainland Southeast Asia, or some now extinct *rufipogon* population, may have been ancestral to all domesticated rice.

When not prejudiced *a priori* by an adamantly articulated archaeological opinion, rice geneticists have explained instead that the widespread transfer of the whiter grain pericarp (*rc*) gene more immediately "implies contact among the people who cultivated the different subspecies" (Sweeney et al. 2007: 1419). Evidence from both linguistic palaeontology and human population genetics inspired a reconstruction that involved precisely such an intense interaction between the early Yangtzeans, who were ancestral to the Hmong-Mien, and the ancient Austroasiatics (van Driem 2011, 2012). We shall recapitulate the evidence for this reconstruction and examine several of the principal implications of this model below.

By contrast, the simplistic model of a single rice domestication in the lower Yangtze advocated by some archaeologists who happen to work in that particular region not only flies in the face of the molecular genetic findings on Asian rice, this single domestication model also overlooks the human cultivators, who served not as unwitting mediators, but acted as knowledgeable agents during the three prolonged rice domestications. In their enthusiasm for the lower Yangtze basin, the archaeologists in question once allowed their reasoning to be clouded by denial of the preservation bias and consequently strayed beyond what I have called "the epistemological event horizon in archaeology" (van Driem 2017).

2. Linguistic palaeontology and the early rice cultivators

In 1830, Julius von Klaproth (1830: 112–113) became the first to discuss the pre-historical implications of the occurrence of phonologically regular reflexes in the

languages of a particular family for reconstructible etyma denoting particular plant and animal species with well-defined geographical ranges. Inspired by von Klaproth's pioneering work, Adolphe Pictet (1859) introduced the term "linguistic palaeontology" to denote an attempt to understand the ancient material culture of a language family or geographically to locate its *Urheimat* on the basis of the lexical items which can be reliably reconstructed for the common ancestral language. The reflexes of reconstructed roots should be attested across the various branches of the family, and their phonological development should be historically regular. With regard to rice, the two language families which reflect rice agriculture terminology most robustly are Austroasiatic and Hmong-Mien.

The Austroasiatic language family boasts the most impressive reconstructible repertoire of rice agriculture terms. Gérard Diffloth (2005) has adduced the following eleven reconstructible Austroasiatic roots: *(kə)ḃaːʔ 'rice plant', *rəŋkoːʔ 'rice grain', *cəŋkaːm 'rice outer husk', *kəndək 'rice inner husk', *pʰeːʔ 'rice bran', *təmpal 'mortar', *jənreːʔ 'pestle', *jəmpɪəʔ 'winnowing tray', *guːm 'to winnow', *jərmuəl 'dibbling stick' and *kəntuːʔ 'rice complement', i.e. accompanying cooked food other than rice. Diffloth has long been the most knowledgeable authority on the comparative study of Austroasiatic. The historical phonology and grammar of Austroasiatic are not as tractable as the comparative study of Hmong-Mien, since Austroasiatic exhibits far greater internal diversity than does Hmong-Mien. Strecker's (1987) Hmong-Mien phylogeny recognised the three branches Hmongic (*Miáo*), Mienic (*Yáo*) and Ho Nte (*Shē*), and treated the precise classification of the Na-e language as problematic. More recently, Ratliff (2010) presented an improved Hmong-Mien family tree. In terms of its internal diversity, the Hmong-Mien language family looks like a vestigial branch of what once may have been a greater linguistic phylum, which Starosta (2005) called "Yangtzean".

Martha Ratliff (2004, 2010) identified ten rice cultivation etyma as reconstructible to the Proto-Hmong-Mien level: *hnrəŋH 'cooked rice', *hɲɔn 'rice head, head of grain', *mbləu 'rice plant, paddy', *mphɛk 'chaff', *mblut 'glutinous', *ljɪŋ 'paddy field', *ljim 'sickle', *ŋkjuəX 'rice cake', *tuX 'husk/pound rice' and *tsɛŋH 'rice steamer'. Five rice agriculture terms are reconstructible to the Proto-Hmongic level: *S-phjæ^C 'chaff', *mbljæ^C 'have food with rice', *ʔrin^A 'dry (rice) in sun', *ntsuw^C 'husked rice' and *tshɛŋ^B 'husked rice or millet'. The two roots *hmei^B 'husked rice' and the rice measure etymon *hrau^A are reconstructible to the Proto-Mienic level. Six of the ten reconstructible Proto-Hmong-Mien etyma are also found in Old Chinese, where, however, they are more likely to represent ancient loans into Sinitic from Hmong-Mien rather than the other way around (*pace* Ratliff 2004: 158–159).

First of all, the Old Chinese forms 秫 *məlut (*shú*) 'glutinous millet' (i.e. not rice), 田 *tʰiŋ (*tián*) 'field', 鎌 *[r]em (*lián*) 'sickle', 粃 *[g](r)a(k)-s (*jù*) 'cakes', 擣 *tʰuʔ (*dǎo*) 'pound, thresh' and 甑 *s-təŋ-s (*zèng*) 'steamer' are not reconstructible

to Trans-Himalayan, of which Sinitic is but a subgroup (van Driem 2005, 2007, 2014a; Old Chinese forms as given by Baxter & Sagart 2014a, 2014b; cf. Ho 2016). Similarly, Ratliff relates Proto-Hmongic *ʔjeŋ^A ‘seedling’ and Proto-Mienic *ʔjaŋ^A ‘seedling’ to Middle Chinese 秧 *ʔjang (*yāng*), evidently due to a discrepancy in vocalism between the Hmongic and Mienic forms, and relates Proto-Hmongic *ljeŋ^A ‘rice measure’ to Old Chinese 量 *[r]aŋ (*liàng, liáng*), but again neither etymon is well reflected in Trans-Himalayan outside of Sinitic. Moreover, not only are the earliest textual attestations of the Chinese forms 田 *lʰiŋ ‘field’, 糕 *[g](r)a(k)-s ‘cakes’, 甑 *s-təŋ-s ‘steamer’ and the measure word 量 *[r]aŋ absolutely ambiguous as to what kind of grain they refer to (though 田 *lʰiŋ ‘field’ may reflect a Hmong-Mien loan into Sino-Bodic rather than just into Sinitic), furthermore the form 糕 *[g](r)a(k)-s ‘cakes’ is not actually an Old Chinese form, for its first known attestation occurs in the poetry anthology of the feudal state of Chǔ, entitled 楚辭 *Chǔcí*, dating from the Hàn period, whereas 鎌 *[r]em ‘sickle’ likewise first occurs in the Hàn period as a western dialect word (Wolfgang Behr, p.c., 19 April 2011).

The Proto-Mienic *hmei^B ‘husked rice’ appears to correspond to Old Chinese 米 *[m]ʰijʔ (*mǐ*), and rice terms containing a bilabial nasal initial appear in other languages in the east of the Trans-Himalayan area, e.g. Bái *me*³³ ‘husked rice’, Jinuo *a*⁴⁴*mε*⁴⁴ ‘rice’, Black Lahu *mi*³³ ‘paddy’, Nusu *me*³³*me*³¹ ‘rice’, Garo *mi*, Dimasa *mai* ‘rice’, Tangkhul *ma* ‘paddy’, Sgaw Karen *me* ‘boiled rice’. Yet the meanings of these forms are disparate, viz. paddy, hulled rice, boiled rice, and the forms may represent mere look-alikes, since no regular phonological correspondence is yet known to obtain between them. Paul Benedict “set up” a Bodo-Koch proto-form *mey or *may ‘rice, paddy’ (1972: 149), which Matisoff later inflated to “*ma ⇔ *may or *mey” (2003: 216, 231) by adding a “monophthongal allofam” and stressing the uncertainty of the rhyme. In fact, no rice agricultural terminology can be confidently reconstructed for the Trans-Himalayan phylum as a whole, an issue noted by Blench (2009).

Rice cultivation terminology is likely to have been borrowed into Sinitic from ancient Hmong-Mien rice cultivators at a time when Proto-Sinitic millet growers intensified their cultural exchange with their southern neighbours. The main split in the Hmong-Mien family is between Hmongic and Mienic. The scattered distributions of the modern language communities belonging to each of these two branches exhibit approximately the same geographical range, which is roughly bisected by the Pearl River. On the basis of the historical sources, it has long been mooted that the geographical centre of gravity of the family would originally have lain further north along the middle Yangtze (Cushman 1970). The historically attested distribution of the early Hmong-Mien tribes during the Eastern Zhōu (770–256 BC) is shown in Figure 1. There is currently no palaeobotanical evidence for the co-cultivation

of rice and foxtail millet along the middle Yangtze until around 3800 BC (Nasu et al. 2006).

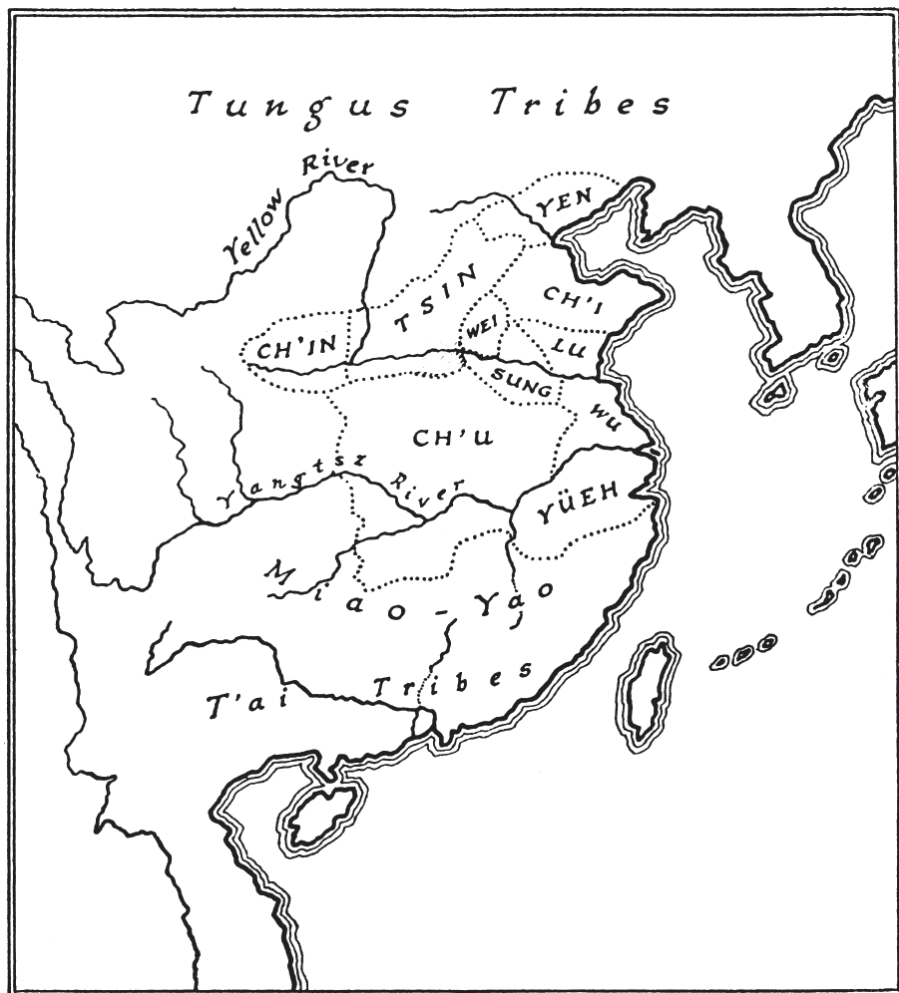


Figure 1. The relative position of early Hmong-Mien (Miáo-Yáo) tribes and early Kradai (T'ai) tribes with respect to the Yangtze and to Old Chinese territory in late Zhōu times, with the hatched portion representing the imperial domain (reproduced from Forrest 1948: 129)

Population genetic findings indicate three distinct domestications of Asian rice. Linguistic palaeontology provides evidence that enables us to ascertain the likely ethnolinguistic identity of two of the three Asian rice domesticators, i.e. the ancient Austroasiatics and the ancient Hmong-Mien. It might appear parsimonious

to ascribe the domestication of the *japonica* cultivar putatively to the Hmong-Mien and the domestication of *indica* and perhaps also *ahu* rice to the ancient Austroasiatics, but the prehistorical reality may have been more intricate. A more interesting proposal emerging from a synthesis of the disparate sets of evidence is presented below. First, however, we shall address problems with the archaeology of rice agriculture and with the argumentation used by archaeobotanists advocating a single original domestication of Asian rice in the lower Yangtze basin.

3. Challenges to the archaeology of rice agriculture

The archaeology of rice agriculture is plagued by an empirical quandary commonly known in the field as a preservation bias. This empirical issue pertains to the archaeological recoverability of rice agriculture sites. The traces of ancient farming communities tend to have been better preserved in the hill tracts surrounding the Brahmaputra flood plains than on the fertile fields themselves. Likewise, in the Yangtze river basin, most salvageable rice agriculture sites are in the foothills or at the base of the foothills (Nakamura 2010). Yet the earliest rice-based cultures may first have developed on those very flood plains. Perhaps the remains of the first rice cultivating cultural assemblages lie buried forever deep beneath the silty sediments of the sinuous lower Brahmaputra basin. Maybe the palaeobotanical evidence for the earliest domestications of rice was washed out by the Brahmaputra long ago and now lies submerged in the depths of the Bay of Bengal.

Archaeologists have looked for the remains of early rice agriculture and indeed found them at some sites and not at some others. The recovered remains of early cultivated rice are of differing antiquity and reflect distinct stages of domestication. Unsurprisingly, archaeologists have not found the remains of early rice agriculture in those places where they have not yet bothered to look. Vast swathes of Asia covering the areas identified by rice geneticists (Londo et al. 2006; Molina et al. 2011; Civián et al. 2015) as harbouring likely sites for the domestication of Asian rice have not been subjected to systematic archaeological and palaeobotanical investigation. The archaeology of northeastern India, the Indo-Burmese borderlands, Burma and the northern Bay of Bengal littoral is virtually unresearched. Political, cultural, geographical and logistic factors have conspired to impede intensive archaeological research in a vast area extending from the lower Brahmaputra basin to the Tenasserim.

Despite the molecular genetic evidence for three independent rice domestications and multidirectional introgression of alleles between the three families of cultivars *ahu*, *indica* and *japonica*, Fuller argued in several publications for a single domestication of Asian rice near the mouth of the Yangtze, where circumstances

and substrate conditions happen strongly to have favoured the preservation of the palaeobotanical remains of early agriculture (Fuller & Qin 2009, 2010; Fuller 2012). His team then resorted to modelling in an attempt to buttress their claim with their archaeological assumptions built into the model (Silva et al. 2015). The model yielded the intuitively satisfying result that the rate of exchange of alleles accelerates over time as domestication progresses, but the trouble with the simulation was that the data fed into the model were largely fortuitous in terms of their geography.

The epistemological problem here is fundamental in nature and, as the old saw has it, the absence of evidence does not constitute the evidence of absence. Fuller (2012), though cursorily acknowledging this problem, initially continued to stress the absence of palaeobotanical evidence in areas where archaeologists had not sought such evidence. The argument for a single domestication in the Lower Yangtze relied on a tacit denial of the ramifications of the preservation bias and on the conceit that the absence of evidence somehow represented the evidence of absence. Continued reliance on this conceit became untenable in face of the utter dearth of archaeobotanical research on rice agriculture in most of the relevant areas (van Driem 2011). The advice was evidently taken to heart, and the population genetic findings on rice were also heeded, inspiring an intended programme of archaeobotanical research that now fortunately envisages the targeting of these regions (Stevens et al. 2016; Fuller et al. 2016).

In consonance with previous rice genetic findings, Choi et al. (2017) conceded the molecular evidence for “significant gene flow in both directions” between the three families of cultivars *ahu*, *indica* and *japonica*. Yet once again on the basis of the entrenched archaeological argumentation, Choi et al. (2017) attempted to mitigate the observed introgression of alleles from *ahu* and *indica* into the *japonica* family of cultivars by speculating that the “introgression from *aus/indica* to *japonica*, however, may have occurred during the diversification phase of rice”. Trying to reinterpret inconvenient and possibly contradictory molecular genetic findings for Asian rice in order to fit them into the mould of a single domestication in the lower Yangtze leads further afield from an interdisciplinary consilience on rice and has brought Choi et al. (2017) to what they have rather optimistically qualified as “a paradox”. Similarly, several incongruous conclusions drawn by Huang et al. (2012) are debunked by Civián et al. (2015).

Despite the archaeological work conducted in the Ganges and Yangtze basins, much of the archaeology of ancient rice agriculture simply remains unknown because little substantive work has been done in the most relevant areas, e.g. northeastern India, Bangladesh, the Indo-Burmese borderlands and Burma. The gargantuan lacunae in archaeological research highlight the impotence of argumentation in favour of a single domestication around the mouth of the Yangtze that denies the epistemological consequences of preservation bias, and even palliates

those molecular genetic findings that are inconvenient to the lower Yangtze unique rice cradle narrative. Future archaeological research will have to come to terms with both the reality and the ramifications of the strong preservation bias in rice agriculture archaeology. Many parts of northeastern India and the Indo-Burmese borderlands have maintained highly diverse rice cultures to the present day. One archaeologist of cereal cultivation in China has cogently argued the need for expanding the scope of archaeological research beyond the Yangtze river basin into these areas, i.e. Lu (2006, 2009).

At the same time, the absence of evidence for rice agriculture of great antiquity in mainland Southeast Asia, despite the relatively more well researched archaeology of the region, presently embarrasses those who have lately taken to espousing Robert von Heine-Geldern's (1917) homeland theory for Austroasiatic around the lower course of the Mekong, without acknowledging the original author of this hypothesis (Sidwell & Blench 2011). However, the fact that the archaeology of northeastern India, the Indo-Burmese borderlands, Burma and the northern Bay of Bengal littoral is virtually unresearched does not similarly compromise homeland proposals in this region. Moreover, the various rice cultivation methods practised in the Brahmaputra basin to this day and the nature of the substrate render it unlikely that palaeobotanical remains would ever be found, notwithstanding the long-term practice of rice agriculture in the region, as meticulously documented by Hazarika (2014, 2017). This incontrovertible given presents an additional epistemological challenge to archaeologists who propound that rice was domesticated around the mouth of the Yangtze.

Furthermore, the argumentation in favour of a single original rice domestication in the lower Yangtze basin also relies heavily on an exaggerated importance attributed to domestication in a highly restricted sense and on grain shattering. This undue emphasis stems inevitably from the archaeological focus on the micromorphological study of rice remains. Domestication in the restricted semantic sense of genetic modification by human agency was perhaps not in all places and at all times as pivotal as Fuller has made it out to be in his writings. It has been claimed that foxtail millet *Setaria italica* and broomcorn millet *Panicum miliaceum* were already collected in the middle Yellow River valley 23,000 ago and already cultivated 19,500 years ago, a full ten millennia anterior to domestication (Li 2015). Li's early dates are certainly questionable, however, and Hu et al. (2008) have argued that millet does not appear to have been a very important source of dietary protein until some time after domestication. Yet the fact remains that grain cultigens were gathered in the wild and subsequently cultivated for long stretches of time before the process of domestication began (Larson et al. 2014). Moreover, some cultigens never or hardly undergo much domestication in the restricted sense of measurable microanatomical modifications by artificial genetic selection.

In a similar vein, the human domestication of Asian rice favoured the mutation *sh4*, which codes for the partial development of the abscission zone where the mature grain detaches from the pedicle so that the diminished brittleness of the rachides reduced grain shattering. It was human agency that facilitated the introgression of genes coding for a whiter grain pericarp (*rc*) and erecter stalks (*Prog1*) from one family of rice cultivars into another. However, domestication that can be measured in terms of morphological differences in microanatomical structure is not necessary for sustained cultivation over long spans of time.

A number of species of wild rice do not just commonly occur, but are also reportedly still cultivated in northeastern India, e.g. *Oryza rufipogon*, *Oryza nivara*, but especially *Oryza officinalis*, *Oryza meyeriana*, *Oryza perennis* and *Oryza granulata*. The shattering of the rice grains onto the field surface does not in practice impede the harvesting of such rice, which continues to be gathered both for human consumption and for use as animal feed (Hazarika 2005, 2006, 2013, 2017). In addition to such cultivated “wild” rice species, many hundreds of indigenous *Oryza sativa* cultivars are grown in this region. Cultivated Asian rice is harvested three times a year in most areas throughout the Brahmaputra basin, using different seasonal cultivation regimes.

The *ahu* family of cultivars is most usually sown directly onto rain-fed upland fields, mainly for swidden or জুম *jhūm* cultivation, but this group also exhibits considerable diversity. The usual growing season in lower areas extends from late March to early July, in the mid hills from late April to early October, and in the upper hills from late June to late December. An early harvest is also practised in some areas, with a growing season from February to May, in which case the rice seedlings are transplanted and irrigated. Some other *ahu* cultivars with a growing season from May to August may likewise employ transplanted seedlings, which may or may not be irrigated.

Another family of rice cultivars is known as শালি *śāli* [xali]. The growing season for these lowland rice cultivars usually stretches from late July to early December, and for some varieties a late growing season from late August to early January is observed. The rice seedlings are transplanted, and the rice is irrigated. Another family of rice cultivars is known as বড়ো *baḍo* [bɔːɔ]. These wetland cultivars are sown in stagnant wetlands or in irrigated fields. The growing season is from late November to early May. It may be significant that the name of this set of rice cultivars in Assamese happens to be homophonous with the Assamese name for the indigenous Trans-Himalayan ethnic group dispersed throughout the Brahmaputra basin. Another family of rice cultivars is known as আচৰা *ācrā* [asːɾa]. These shallow-water cultivars grow in water that is one to two feet deep. The growing season stretches from late March to early December. Yet another family of rice cultivars is বাও *bāo* [bao]. These deep-water cultivars grow in water that is two to five feet deep,

and can thrive in water that is more than twice that deep, and the growing season stretches from late March to early December (Hazarika 2014, 2017).

Despite weaknesses in the reasoning employed by archaeologists in their eagerness to gain recognition for the lower Yangtze basin as the unique cradle of rice domestication, the archaeology of rice agriculture has nonetheless produced important results. The domestication of *japonica* rice through genetic modification by selective breeding was possibly effectuated along the Yangtze by people, who previously relied far more heavily on the collecting of acorns, water chestnuts and foxnuts before becoming reliant on rice cultivation. In terms of measurable modifications to microanatomical morphology, the process of domestication appears to have begun in the middle of the sixth millennium and to have been largely completed by the end of the fifth millennium BC (Fuller et al. 2009; Nakamura 2010; Zhao 2010; Fuller & Qin 2009; Ruddiman et al. 2008; Fuller, Harvey & Qin 2007). Currently the oldest datable domesticated rice remains from the Pearl River delta date from ca. 3000 BC (Yang et al. 2016).

Rice cultivation reached the Yellow River basin during the third millennium BC (Crawford & Shen 1998) and Formosa and Vietnam between 2500 and 2000 BC (Higham & Lu 1998), but only spread throughout the Indochinese peninsula between 1500 and 500 BC (Weber et al. 2010; Oxenham et al. 2015). It has been claimed that rice may have been cultivated in the Gangetic basin as early as 7000 BC (Sharma et al. 1980; Pal 1990; Agrawal 2002), but the current datable evidence for the actual domestication of rice in the middle Ganges dates from no earlier than the second half of the third millennium BC. In line with the molecular genetics, archaeogenetic data from Asian rice remains found in sites in India and Thailand show hybridisation between *indica* and *japonica* cultivars of domesticated rice after their initial domestications (Castillo et al. 2016), even though the sterility of hybrids sometimes acts as a barrier that helps to keep the two cultivars distinct (Chen et al. 2008).

Both broomcorn and foxtail millet agriculture were practised in the high and arid hills of what today is Sichuān province from ca. 4000 to 2500 BC. By 2700 BC, both rice and foxtail millet were cultivated by the inhabitants of the Bāodūn culture (ca. 2700–1700 BC) in the Chéngdū plain in what today is west-central Sichuān (d'Alpoim Guedes 2011; d'Alpoim Guedes et al. 2013). Based on the dating of the few known sites, such as མཁཱ་རོ་ mKhar-ro near ཇམ་མོ་ལྷོ་ Chab-mdo (van Driem 2001: 430–431), it has been conjectured that the spread of agriculture to the Tibetan plateau was posterior to this date by archaeologists who envisage the agricultural colonisation of Sichuān and eastern Tibet as proceeding from the middle Yangtze (d'Alpoim Guedes et al. 2014; d'Alpoim Guedes 2015). Although it appears likely that agriculture facilitated human habitation of the Tibetan plateau at around this time (Chen et al. 2015), various types of evidence indicate that the Tibetan plateau

was permanently occupied long beforehand (Xiang et al. 2013; Huerta-Sánchez et al. 2014; Lorenzo et al. 2014; van Driem 2015a; Lou et al. 2015; Hackinger et al. 2016; Lu et al. 2016). Indeed, eastern Tibet and modern Sichuān lay beyond the periphery of the ancient rice corridor, which extended from the Brahmaputra basin to the Yangtze basin by way of Burma and Yúnnán.

4. Zooming in on the Austroasiatic and Hmong-Mien homelands

Scholars have sought to situate the Austroasiatic *Urheimat* as far west as the Indus valley and as far east as the Yangtze delta or insular Southeast Asia. Purely from the point of view of the current geographical distribution of Austroasiatic language communities, more logical contenders for the Austroasiatic homeland are the Indian subcontinent, the Bay of Bengal littoral, mainland Southeast Asia and the middle Yangtze. The gaping lacunae in palaeobotanical research are convenient to the argument in favour of the middle and lower Yangtze basin, where conditions happen to have favoured the preservation of archaeologically recoverable remains. Linguistically, the old hypothesis that proposed Old Chinese 江 *k'ron (jiāng) 'Yangtze' to be a loan from Austroasiatic emboldened Pulleyblank (1983) to envision a major Austroasiatic presence all along the eastern seaboard from Vietnam to Shāndōng, and to impute an Austroasiatic ethnolinguistic identity to the Lóngshān horizon. This interpretation of the linguistic data has notably been challenged by Zhāng (1998).

Four types of evidence help us to zoom in on the possible geographical location of the Austroasiatic homeland. The first type of evidence, already mentioned, is linguistic and involves the current geographical distribution of Austroasiatic language communities, which is shown in Figure 2. Both the centre of gravity of the phylum on the basis on the geographical distribution of modern Austroasiatic language communities as well as the deepest phylogenetic divisions in the family tree point to the northern Bay of Bengal littoral. The deepest historical division in the family's phylogeny lies between Munda in the west and Khasi-Aslian in the east, which would put the homeland on either side of the Ganges and Brahmaputra delta. Even the deepest division within the Khasi-Aslian trunk, i.e. the split into Khasi-Pakanic and Mon-Khmer, would suggest a point of dispersal for Khasi-Aslian between South Asia proper and mainland Southeast Asia proper. The family tree of Austroasiatic, showing the correct phylogenetic position for Pearic, presented by Diffloth for the first time at Agay in 2012, is shown in Figure 3. The internal phylogeny of the Munda branch has not, however, been established.

The second and third type of evidence involve linguistic palaeontology. The Proto-Austroasiatic rice terms adduced above, reconstructed by Gérard Diffloth,



Figure 2. The geographical distribution of Austroasiatic language communities

constitute the second set of evidence. The suspected geographical ranges for the three rice domestications identified by Londo et al. (2006) on the basis of the geographical distribution of genetic markers in the wild precursor *Oryza rufipogon* are shown in Figure 4. The third set of evidence involves reconstructed roots denoting megafauna in the Proto-Austroasiatic lexicon in light of the attested geographical distribution of these species in the Holocene. This set of evidence formed the topic of an earlier study (van Driem 2012), for which Anne-Marie Bacon and Danièle Fouchier of the research unit *Dynamique de l'Évolution Humaine* at the *Centre National de la Recherche Scientifique* in Paris generously furnished the Holocene distribution maps. The Proto-Austroasiatic etyma reconstructed by Gérard Diffloth (2005: 78) evoke the fauna and ecology of a tropical humid homeland environment:

*mra:k 'Indian peafowl *Pavo cristatus*' or 'Javan peafowl *Pavo muticus*'

*tərkuət 'tree monitor *Varanus nebulosus* or *bengalensis*'

*tənyu:ʔ 'binturong *Arctitis binturong*'

*(bən)jo:l ~ *(j(ə)rm)o:l 'Sunda pangolin *Manis javanica*' or 'Chinese pangolin *Manis pentadactyla*'

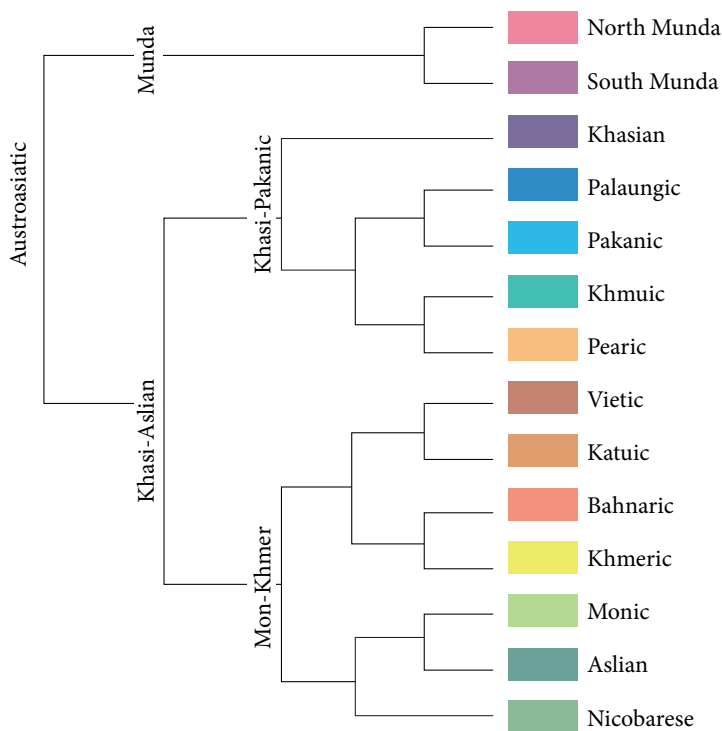


Figure 3. The family tree of Austroasiatic (Diffloth 2012). Unlike the Khasi-Aslian branch, the internal phylogeny of the Munda branch has not been established

*dəkan ‘Sumatran bamboo rat *Rhizomys sumatrensis*’, ‘Chinese bamboo rat *Rhizomys sinensis*’, ‘hoary bamboo rat *Rhizomys pruinosus*’

*kaciaŋ ‘the Asian elephant *Elephas maximus*’

*kiaç ‘mountain goat *Capricornis sumatrensis*’

*rəma:s ‘Indian rhinoceros *Rhinoceros unicornis*’, ‘Javan rhinoceros *Rhinoceros sondaicus*’ or ‘Sumatran rhinoceros *Dicerorhinus sumatrensis*’

The Holocene distribution maps included in the 2012 study are not reproduced here. Instead, Figure 5 offers a synthesis of the mapped data by depicting the area where the ranges of the species for which the Proto-Austroasiatic lexicon has reconstructible etyma overlap in northeastern India, the Indo-Burmese borderlands and Burma. A comparison of Figures 4 and 5 shows that the areas suggested for an Austroasiatic homeland by the two sets of linguistic palaeontological evidence correspond to a large degree. The fourth and last set of evidence pertains to human populations genetics.

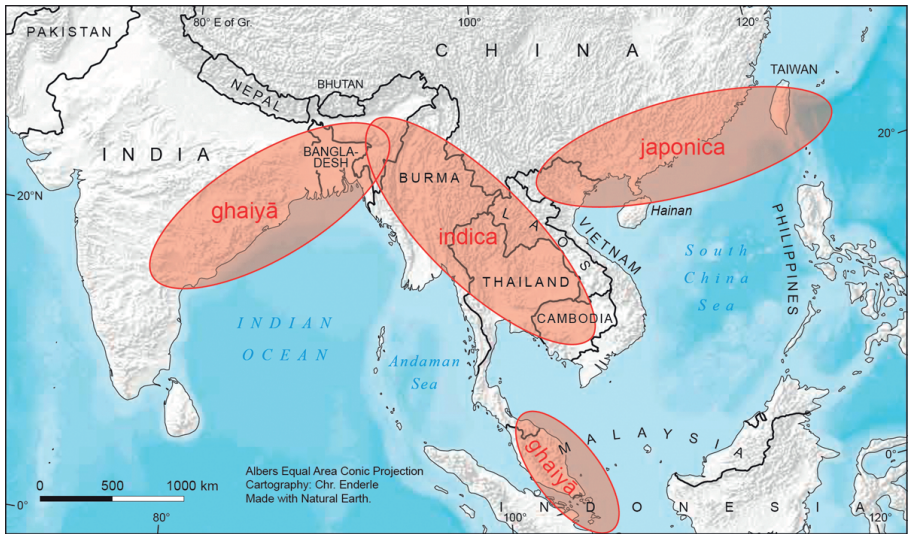


Figure 4. The geographical ranges for the possible domestication of (A) *ghaiyā* or upland rice, (B) wet *indica* rice and (C) the *japonica* cultivar, based on the geographical distribution of genetic markers in the wild precursor *Oryza rufipogon* (adapted from Londo et al. 2006)



Figure 5. The region of overlap of the geographical ranges of megafaunal species for which Proto-Austroasiatic etyma are reconstructible

5. The Father Tongue correlation and the East Asian linguistic phylum

Evidently, it cannot be repeated too often that a proto-language can only be reconstructed on the basis of linguistic evidence and that the linguistic ancestors of any modern language community were not necessarily the same people as the community's biological forebears. Although these points have long been reiterated from the time of Julius von Klaproth (1823) and Max Müller (1872), these lessons are often lost on some audiences. By the same token, each of us has countless ancestors via numerous lineages. There is no such thing as a pure race. In fact, in molecular genetic terms there is no such thing as race (Cavalli-Sforza, Menozzi and Piazza 1994). We are all members of one large human family. Moreover, even when languages and genes happen to exhibit a correlation, such a marker relationship should not be confused with identity. The correlation of a particular chromosomal marker with the distribution of a certain language family must not be simplistically equated with populations speaking languages of a particular linguistic phylum. Rather, molecular markers on the Y chromosome serve as proxies or tracers for the movements of paternal ancestors.

When studying the distribution of maternally inherited markers in the mitochondrial DNA and paternally inherited markers on the Y chromosome, a Swiss-Italian team of population geneticists soon found that it was easier to find statistically relevant correlations between the language of a particular community and the paternally inherited markers prevalent in that community than between the language and the most salient maternally inherited markers found in that speech community. This Father Tongue correlation was first described by Poloni et al. (1997, 2000). On the basis of this finding, it was inferred that paternally inherited polymorphisms may serve as markers for linguistic dispersals in the past, and that a correlation of Y chromosomal markers with language may point towards male-biased linguistic intrusions. The Father Tongue correlation is ubiquitous but not universal. Its preponderance allows us to deduce that a mother teaching her children their father's tongue must have been a prevalent and recurrent pattern in linguistic prehistory.

There are a number of reasons why we might expect this outcome. The Y chromosome underwent a global bottleneck towards the end of the last ice age, when certain paternal clades started eradicating or out-competing other clades (Karmin et al. 2015). The founding dispersals of many major language families appear to be related to the robust spread and reproductive success of the bearers of a subset of Y chromosomal haplogroups that survived this bottleneck. As a consequence, the global phylogeography of Y chromosomal haplogroups is shallower in terms of time depth than the worldwide mitochondrial landscape. The initial human colonisation of any virgin part of the planet must have involved both sexes in order for

a population of progeny to establish itself. Once a population is in place, however, subsequent migrations could have been heavily gender-biased. Subsequently, male intruders could impose their language whilst availing themselves of the womenfolk already in place. In this regard, population geneticist Toomas Kivisild (2014) has wryly characterised warfare as a sex-specific pathology linked to the Y chromosome. Whereas the landscape of paternal lineages often appears to correlate with language at the comparatively shallower time depth of the linguistically reconstructible past, correlations between maternal lineages and linguistic phylogeography discerned to date have been underwhelming. The Father Tongue hypothesis suggests that linguistic dispersals were, at least in most parts of the world, posterior to initial human colonisation and that many linguistic dispersals were predominantly later male-biased intrusions. Such patterns are observed worldwide.

In two previous studies, I have shown that the geographical distribution and phylogeography of subclades of the Y chromosomal haplogroup O appear to be correlated with the dissemination of four recognised language families, viz. Austroasiatic, Trans-Himalayan, Hmong-Mien and Austro-Tai (van Driem 2014b, 2015b). These four language families were united into a single East Asian linguistic phylum in a hypothesis proposed by Starosta (2005). In presenting my own tweaked recension of Starosta's East Asian family tree in 2012 in Benares (van Driem 2014b), shown in Figure 6, I pointed out that Starosta was the most recent exponent of a long tradition of linguists who had attempted to unite one or more of these language families into a grander linguistic phylum and, in so doing, ventured beyond the epistemological constraints of what I call the "linguistic event horizon". This horizon is the maximal time depth accessible through methodologically sound linguistic reconstruction and the boundary beyond which any reconstructions are at one point reduced to sheer speculation. Scholars who have proposed earlier renditions of the East Asian linguistic phylum have ranged from methodologically rigorous historical linguists such as Blust (1996) to megalocomparativists such as Benedict (1942), and from those offering just unsupported conjecture, e.g. Schlegel (1901, 1902), to those providing sound evidence in the form of phonologically regular correspondences, e.g. Ostapirat (2005, 2013).

The shared morphological vestiges adduced by Starosta in support of his East Asian linguistic phylum comprised the agentive prefix *<m->, the patient suffix *<-n>, what he called the instrumental prefix *<s-> and what he termed the perfective prefix *<n->. A discussion of the merits of the evidence advanced by Starosta for this linguistic phylum strikes me as being of little utility, since I consider the phylum to lie at the linguistic event horizon and therefore doubt whether this issue can ever be conclusively resolved on the basis of firmly reconstructible linguistic evidence. Rather, Starosta himself proposed that the "potential utility" of his hypothesis lay "in helping to focus scholars' efforts on particular specific questions,

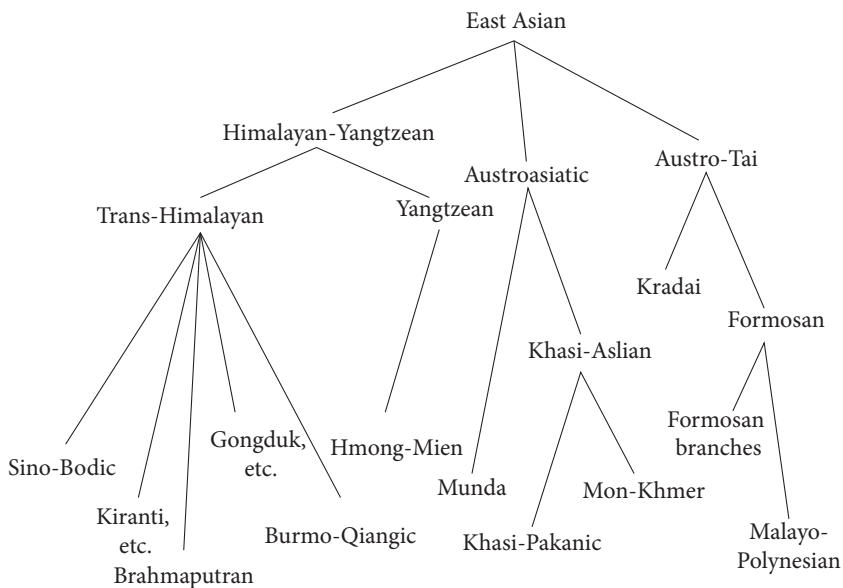


Figure 6. The 2012 Benares recension of Stanley Starosta’s 2001 Périgueux East Asian linguistic phylum (Starosta 2005; van Driem 2014b)

resulting in the replacement of parts of this hypothesis with better supported arguments” (2005: 194).

The resolution of the Y chromosomal tree is constantly being enhanced. Haplogroup labels are updated to reflect our improved understanding of the phylogeny. Mutations numbers tend to remain unchanged, provided that the markers in question prove to be reliable in defining haplogroups. Conventional haplogroup labels of the Y Chromosome Consortium are still widely in use, but have been replaced here with the newer labels of the International Society of Genetic Genealogy, reflecting refinements incorporated up to the 12th of May 2017. In my two previous studies, I noted that the paternal haplogroup O1b1a1a (M95) was correlated with populations speaking languages belonging to the Austroasiatic language family, the haplogroup O2a2b1 (M134) with the Trans-Himalayan language family, the haplogroup O2a2a1a2 (M7) with Hmong-Mien and the haplogroup O1 (F265, M1354) with the Austro-Tai language family.

The complex history of Sinitic populations featured successive constellations of dynastic empires governed from geographically ever shifting capitals, whereby subjugated and neighbouring populations as well as immigrants were absorbed. Not surprisingly therefore, HàN Chinese populations tend to represent an amalgam of East Asian paternal lineages. Yet even in HàN Chinese populations, the molecular marker associated with the spread of a Trans-Himalayan father tongue, i.e.

haplogroup O2a2b1 (M134), taken together with its subclade O2a2b1a1 (M117), occurs in a much higher frequency than any other O subclade, and approximately twice as frequently as the next most frequent fraternal subclade O2a1c (002611) (Yan et al. 2011; Wang et al. 2013; Yao et al. 2017).

In observing the non-random correlation of these four recognised language families with subclades of the paternal haplogroup O, I speculated that the four major East Asian language families were the result of prehistoric bottlenecks. Palaeolithic populations were small, and the effective founder population sizes of the major modern paternal subclades must have been quite small, whilst new populations arise from the small surviving subsets that have passed through bottlenecks. The four language families Austroasiatic, Trans-Himalayan, Hmong-Mien and Austro-Tai appear to have arisen in this way in correlation with specific paternal lineages.

In another study, we showed that the Munda branch of Austroasiatic had arisen as the result of a sexually biased linguistic intrusion into the Indian subcontinent from the region to the north of the Bay of Bengal (Chaubey et al. 2010). As a consequence of the comparatively younger date and the nearly absolute gender asymmetry of this linguistic intrusion, it appears that the deepest division within the Khasi-Aslian trunk of Austroasiatic, i.e. the split between Khasi-Pakanic and Mon-Khmer, might perhaps be more indicative of the geographical location of the Austroasiatic homeland than the split between Munda and Khasi-Aslian. If we accept this line of reasoning, then the point of dispersal for Khasi-Aslian would appear to have lain in the area between South Asia proper and mainland Southeast Asia proper.

6. Rice and the East Asian dispersal

Long before the linguistically reconstructible past, at a time that lay well beyond the linguistic event horizon, the paternal haplogroup K (M9) was centred in the area between South Asia and Southeast Asia, where the ancestral K* appears to have been situated. This clade spawned many successful paternal lineages, some of which moved into insular Southeast Asia, i.e. the haplogroups S (M69) and M (M304), whereas other clades moved back westward into South Asia and beyond, viz. the haplogroups Q (M242), R (M201), T (M89) and L (M429) (Karafet et al. 2015). The geographical locus of yet another descendant subclade lay in the Eastern Himalaya, i.e. the ancestral haplogroup NO (M214). Millennia after the two paternal lineages N and O had split up, the bearers of haplogroup N set out for East Asia just after the last glacial maximum, braving ice and tundra, and – in a grand counterclockwise sweep – migrated across northern Eurasia as far as west as Lapland, whilst the

ancestral form *N appears to have been situated in northern Burma (Rootsi et al. 2007; Derenko et al. 2007; Mirabal et al. 2009; Ilumäe et al. 2016).

The paternal clade O is a marker that was overwhelmingly shared by the linguistic ancestors of what Starosta (2005) called the East Asian linguistic phylum. The non-random correlation of the subclades of this particular Y chromosomal haplogroup with the four recognised language families enables us to infer the following sequence of events. Millennia before the end of the last glacial maximum, the paternal lineage O (M175) split into the subclades O2 (M122) and O1 (F265, M1354), as shown in Figure 7. The two subclades can be putatively assigned to two geographical loci, with the haplogroup O1 (F265, M1354) moving eastward into East Asia south of the Yangtze, whilst bearers of the O2 (M122) haplogroup settled in the general region of the Eastern Himalaya.

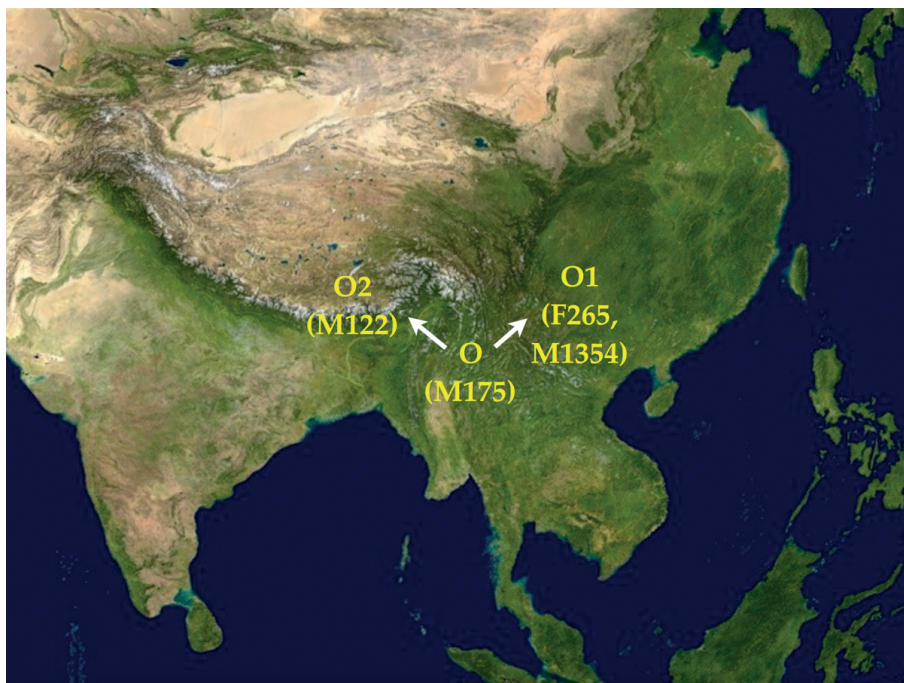


Figure 7. After the last glacial maximum, the Y chromosomal haplogroup O (M175) split into the subclades O1 (F265, M1354) and O2 (M122)

Subsequently, as temperature and humidity increased after the last glacial maximum, haplogroup O split further into the paternal lineages that serve as tracers for the spread of Trans-Himalayan, Hmong-Mien, Austroasiatic and Austro-Tai. The O1 (F265, M1354) lineage south of the Yangtze split into the subclades O1b (M268) and O1a (M119), with the latter moving eastward to the Fújiàn hill tracts and across

the strait to settle on Formosa, which so became the *Urheimat* of the Austronesians (cf. Abdulla et al. 2009). Subsequently, the subclade O1b (M268) gave rise to the filial subclades O1b2 (M176) and O1b1a1a (M95). The bearers of haplogroup O1b1a1a (M95) became the progenitors of the Austroasiatics (van Driem 2007; Chaubey et al. 2010). The Austroasiatics spread throughout the Salween drainage and thence to southern Yúnnán, northern Thailand and western Laos. In time, the Austroasiatics would spread as far as the Mekong delta, the Malay peninsula and the Nicobars. Secondly, bands of male Austroasiatics would introduce both their language and their paternal lineage, O1b1a1a (M95), to the indigenous peoples of the Choṭā Nāgpur, as shown in Figure 8.



Figure 8. A male-biased linguistic intrusion introduced both Austroasiatic language and a paternal lineage, haplogroup O1b1a1a (M95), into the indigenous population of the Choṭā Nāgpur

The linguistic palaeontological evidence adduced above shows that the ancestral Austroasiatics practised rice agriculture, whilst the geographical distribution of haplogroup O1b1a1a (M95) correlates neatly with populations speaking Austroasiatic languages. The inference can therefore be made that Asian rice was cultivated by the ancestral bearers of haplogroup O1b1a1a (M95). The fraternal clade O1b2 (M176), which we may call “para-Austroasiatic”, spread eastward, where they disseminated

rice agriculture to the lower Yangtze. Although the genetic legacy of the eastward migration of the bearers of the O1b2 (M176) persists residually today in mainland East Asia, these ancestral fathers left no linguistic trace of the father tongue that they once spoke, except for perhaps an old name for the Yangtze river that was ultimately borrowed by Old Chinese as 江 *k'ron (*jiāng*), as proposed by Pulleyblank (1983).

This para-Austroasiatic paternal lineage O1b2 (M176) advanced as far as the Korean peninsula and also represents a major wave of immigration recorded in the Japanese genome. We can identify the O1b2 (M176) lineage with the Yayoi people, who introduced rice agriculture to Japan, as early as the second millennium BC, during the final phase of the Jōmon period. In addition to rice, the Yayoi also introduced other crops of continental origin to Japan such as millet, wheat and melons. The gracile Yayoi immigrants soon outnumbered the more robust and less populous Jōmon people, who were Palaeolithic hunters and foragers and the descendants of earlier waves of peopling, including the first anatomically modern humans to populate the Japanese archipelago.

About twelve thousand years ago, at the dawn of the Holocene, in the southeastern Himalayas and eastern slopes of the Tibetan Plateau, haplogroup O2 (M122) gave rise to the ancestral Trans-Himalayan paternal lineage O2a2b1 (M134) and the “Yangtzean” or Hmong-Mien paternal lineage O2a2a1a2 (M7), as shown in Figure 9. It is a reasonable conjecture that the bearers of the polymorphism O2a2b1 (M134) at first remained in the Eastern Himalaya, which today continues to represent the centre of phylogenetic diversity of the Trans-Himalayan language family based on the geographical distribution of primary linguistic subgroups. Only later would early Trans-Himalayan language communities spread into northeastern India, southeastern Tibet and northern Burma, but first the bearers of the O2a2a1a2 (M7) lineage migrated eastward to settle in the areas south of the Yangtze. On their way, the early Hmong-Mien encountered the ancient Austroasiatics, from whom they adopted rice agriculture. The intimate interaction between ancient Austroasiatics and the ancestral Hmong-Mien not only involved the sharing of knowledge about rice agriculture technology, but also left a genetic trace in the high frequencies of haplogroup O1b1a1a (M95) in today’s Hmong-Mien and of haplogroup O2a2a1a2 (M7) in today’s Austroasiatic populations.

On the basis of these Y chromosomal haplogroup frequencies, Cai et al. (2011:8) observed that Austroasiatics and Hmong-Mien are “closely related genetically” and ventured to speculate about “a Mon-Khmer origin of Hmong-Mien populations”. It would be more accurate to infer that the incidence of haplogroup O2a2a1a2 (M7) in Austroasiatic language communities of Southeast Asia indicates a significant Hmong-Mien paternal contribution to the early Austroasiatic populations whose descendants settled in Southeast Asia, whereas the incidence of haplogroup O2a2a1a2 (M7) in Austroasiatic communities of the Indian subcontinent

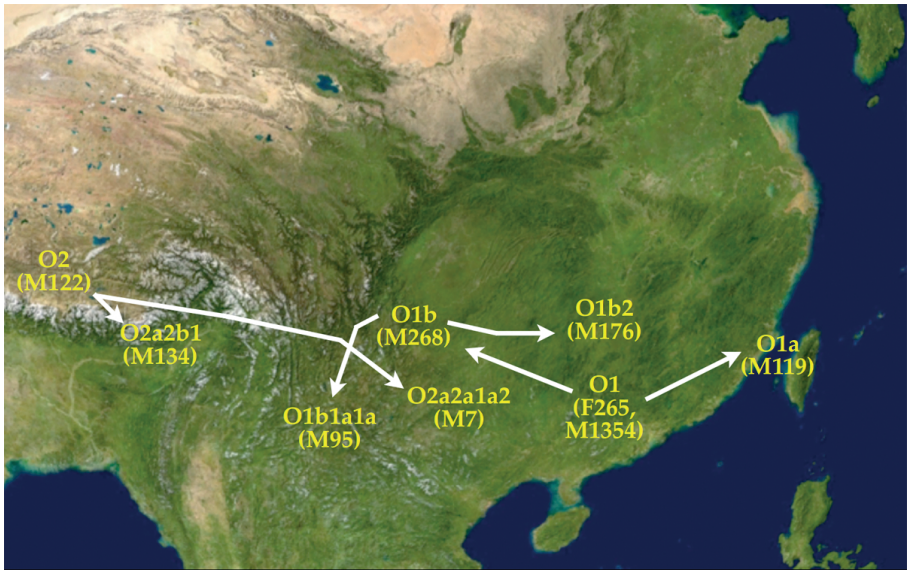


Figure 9. At a more recent time depth, paternal lineages branched into new subclades, and each event involved a linguistic bottleneck leading to language families that today are reconstructible as distinct linguistic phyla. The O1 (F265, M1354) lineage gave rise to the O1a (M119) and O1b (M268) subclades. The former moved eastward to the Fújiàn hill tracts and across the strait to Formosa, which so became the *Urheimat* of the Austronesians. Bearers of the paternal lineage O1b (M268) domesticated Asian rice and spawned the paternal subclades O1b1a1a (M95) and O1b2 (M176). Haplogroup O1b1a1a (M95) is the Proto-Austroasiatic paternal lineage, whereas the para-Austroasiatic fraternal clade O1b2 (M176) spread eastward, sowing seed along the way. The haplogroup O2 (M122) gave rise to the paternal subclades O2a2b1 (M134) and O2a2a1a2 (M7). The spread of the molecular marker O2a2b1 (M134) from the Eastern Himalaya serves as a tracer for the dissemination of people speaking languages of the Trans-Himalayan family, whereas the paternal lineage O2a2a1a2 (M7) serves as a tracer for the spread of people speaking languages of the Hmong-Mien family.

is undetectably low. The incidence of the Y chromosomal haplogroup O1b1a1a (M95) amongst the Hmong-Mien appears to indicate a slightly lower Austroasiatic paternal contribution to Hmong-Mien populations than vice versa. As the Hmong-Mien moved eastward, the bearers of para-Austroasiatic haplogroup O1b2 (M176) likewise continued to move east.

Three domestications of Asian rice *Oryza sativa*, involving the cultivar families *ahu*, *indica* and *japonica*, took place through the agency of ancient rice cultivators who bore three distinct paternal lineages, i.e. the Austroasiatic paternal subclade

O1b1a1a (M95), the para-Austroasiatic paternal lineage O1b2 (M176) and the “Yangtzean” or Hmong-Mien paternal lineage O2a2a1a2 (M7). The region between the Brahmaputra river basin and the Yangtze river basin runs through Burma and southern Yúnnán and harbours numerous ecotypes and topographies. In this area, the domestication of three different families of Asian rice cultivars took place, each suited to a different ecology.

The three populations involved not only exchanged paternal lineages but also rice knowledge which enabled the introgression of favoured traits between the three families of cultivars *ahu*, *indica* and *japonica*. I propose that the cultivar families *ahu* and *indica* were first cultivated by the ancient Austroasiatics and by the ancient Hmong-Mien or Yangtzeans, whereas the domestication of *japonica* rice was conducted by the bearers of the para-Austroasiatic paternal lineage O1b2 (M176), who left no linguistic trace other than perhaps an old para-Austroasiatic toponym for the Yangtze, but whose descendants surfaced in the archaeological record of the Japanese archipelago as the people behind the Yayoi culture.

Meanwhile, the bearers of Y chromosomal haplogroup O2a2b1 (M134) in the eastern Himalayan region expanded further eastward throughout Sìchuān and Yúnnán, north and northwest across the Tibetan plateau as well as further westward across the Himalayas and southward into the Indo-Burmese borderlands. On the Brahmaputra plain, the early Trans-Himalayans encountered the Austroasiatics, who had preceded them. The relative frequencies of the Y chromosomal haplogroup O1b1a1a (M95) in Trans-Himalayan speaking populations of the Indian subcontinent (Sahoo et al. 2006; Reddy et al. 2007) suggest that a subset of the paternal ancestors of some Trans-Himalayan populations in northeastern India, e.g. certain Bodo-Koch communities, may originally have been Austroasiatic speakers who were linguistically assimilated by Trans-Himalayans.

Finally, the ancestral Trans-Himalayan paternal lineage O2a2b1 (M134) spread from the Eastern Himalaya in a northeasterly direction to the North China plain. At a much later and shallower time depth, the Trans-Himalayan paternal lineage O2a2b1 (M134) spread in tandem with early Sinitic speaking populations southward expansion from the Yellow River basin into southern China during the Qín dynasty in the third century BC. The ancestral Trans-Himalayan paternal lineage O2a2b1 (M134) is intrusively present in the Korean peninsula and beyond.

References

- Abdulla, Mahmood Ameen *et al.* 2009. Mapping human genetic diversity in Asia. *Science* 326: 1541–1545. doi:10.1126/science.1177074
- Agrawal, Dharma P. 2002. The earliest pottery and agriculture in South Asia. In *The Origins of Pottery and Agriculture*, Yoshinori Yasuda (ed.), 81–88. New Delhi: Lusre Press and Roli Books, for the International Center for Japanese Studies.
- d'Alpoim Guedes, Jade. 2011. Millets, rice, social complexity and the spread of agriculture to the Chengdu plain and southwest China. *Rice* 4(3): 104–113. doi:10.1007/s12284-011-9071-1
- d'Alpoim Guedes, Jade. 2015. Rethinking the spread of agriculture to the Tibetan plateau. *Holocene* 25(9): 1–13.
- d'Alpoim Guedes, Jade *et al.* 2013. Site of Baodun yields earliest evidence for the spread of rice and foxtail millet agriculture to south-west China. *Antiquity* 87: 758–771. doi:10.1017/S0003598X00049449
- d'Alpoim Guedes, Jade *et al.* 2014. Moving agriculture onto the Tibetan plateau: the archaeobotanical evidence. *Archaeological and Anthropological Sciences* 6(3): 255–269. doi:10.1007/s12520-013-0153-4
- Baxter, William Hubbard & Sagart, Laurent. 2014a. Baxter-Sagart Old Chinese reconstruction, version 1.1. <<http://ocbaxtersagart.lsa.umich.edu>> (20 September 2014). doi:10.1093/acprof:oso/9780199945375.001.0001
- Baxter, William Hubbard & Sagart, Laurent. 2014b. *Old Chinese: A New Reconstruction*. Oxford: OUP. doi:10.1093/acprof:oso/9780199945375.001.0001
- Benedict, Paul King. 1942. Thai, Kadai, and Indonesia: A new alignment in southeastern Asia. *American Anthropologist* 44: 576–601. doi:10.1525/aa.1942.44.4.02a00040
- Benedict, Paul King. 1972. *Sino-Tibetan: A Conspectus*. Cambridge: CUP. doi:10.1017/CBO9780511753541
- Blench, Roger Marsh. 2009. If agriculture cannot be reconstructed for Proto-Sino-Tibetan, what are the consequences? Paper presented at the 42nd Conference on Sino-Tibetan Languages and Linguistics at Chiang Mai, 3 November.
- Blust, Robert. 1996. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. In *Prehistoric Settlement of the Pacific*, Ward H. Goodenough (ed.), 117–160. Philadelphia PA: American Philosophical Society.
- Cai Xiaoyun *et al.* 2011. Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *Public Library of Science* 6(8): e24282.
- de Candolle, Alphonse-Louis-Pierre Pyrame. 1883. *Origine des plantes cultivées*. Paris: Librairie Germer Ballière et Compagnie.
- Castillo, Cristina Cobo *et al.* 2016. Archaeogenetic study of prehistoric rice remains from Thailand and India: evidence of early japonica in South and Southeast Asia. *Archaeological and Anthropological Sciences* 8(3): 523–543. doi:10.1007/s12520-015-0236-5
- Cavalli-Sforza, Luigi Luca, Menozzi, Paolo & Piazza, Alberto. 1994. *The History and Geography of Human Genes*. Princeton NJ: Princeton University Press.
- Chaubey, Gyaneshwer *et al.* 2010. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution* 28(2): 1013–1024. doi:10.1093/molbev/msq288

- Chen, Jiongjiong *et al.* 2008. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences* 105(32): 11436–11441. doi:10.1073/pnas.0804761105
- Chen, Fahu *et al.* 2015. Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science* 347 (6219): 248–250. doi:10.1126/science.1259172
- Choi, Jae Young *et al.* 2017. The rice paradox: Multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution* <<https://doi.org/10.1093/molbev/msx049>>. doi:10.1093/molbev/msx049
- Civáň, Peter, Craig, Hayley, Cox, Cymon J. & Brown, Terence A. 2015. Three geographically separate domestications of Asian rice. *Nature Plants*, 1, article No. 15164 <doi:10.1038/nplants.2015.164>. See van Driem (2011: 131–132) for additional references on rice and the population genetics of rice.
- Crawford, Gary W. & Shen, Chen. 1998. The origins of rice agriculture: Recent progress in East Asia. *Antiquity* 72: 858–866. doi:10.1017/S0003598X00087494
- Cushman, Richard D. 1970. Rebel Haunts and Lotus Huts: Problems in the Ethnohistory of the Yao. PhD dissertation, Cornell University.
- Derenko, Miroslava *et al.* 2007. Y-chromosome haplogroup N dispersal from south Siberia to Europe. *Journal of Human Genetics* 52: 763–770. doi:10.1007/s10038-007-0179-5
- Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 77–80. London: Routledge Curzon. doi:10.4324/9780203343685_chapter_5
- Diffloth, Gérard. 2012. The four registers of Pearic. Paper presented at the 22nd Meeting of the Southeast Asian Linguistics Society, Agay, 2 June 2012.
- van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language*, 2 Vols. Leiden: Brill.
- van Driem, George. 2005. Tibeto-Burman vs. Indo-Chinese: Implications for population genetics, archaeologists and prehistorians. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 81–106. London: Routledge Curzon. doi:10.4324/9780203343685_chapter_6
- van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1(2): 211–270. doi:10.1163/2405478X-90000023
- van Driem, George. 2011. Rice and the Austroasiatic and Hmong-Mien homelands. In *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, Nick J. Enfield (ed.), 361–389. Canberra: Pacific Linguistics.
- van Driem, George. 2012. The ethnolinguistic identity of the domesticators of Asian rice. *Comptes Rendus Palévol* 11(2): 117–132. doi:10.1016/j.crpv.2011.07.004
- van Driem, George. 2014a. Trans-Himalayan. In *Trans-Himalayan Linguistics*, Nathan Hill & Thomas Owen-Smith (eds), 11–40. Berlin: Mouton de Gruyter.
- van Driem, George. 2014b. A prehistoric thoroughfare between the Ganges and the Himalayas. In *50 Years after Daojali-Hading: Emerging Perspectives in the Archaeology of Northeast India*, Tiatoshi Jamir & Manjil Hazarika (eds), 60–98. New Delhi: Research India Press.
- van Driem, George. 2015a. Health in the Himalayas and the Himalayan homelands. In *Tibetan and Himalayan Healing*, Charles Ramble & Ulrike Roesler (eds), 161–194. Kathmandu: Vajra Publications.

- van Driem, George. 2015b. The Himalayas as a prehistoric corridor for the peopling of East and Southeast Asia. In *Nepal: An Introduction to the Natural History, Ecology and Human Environment in the Himalayas*, Georg Miehe & Colin Pendry (eds), 318–325. Edinburgh: Royal Botanic Garden Edinburgh.
- van Driem, George. 2017. The epistemological event horizon in archaeology. In *Prehistory and Archaeology of Northeast India*, Manjil Hazarika (ed.), xv–xvii. New Delhi: OUP.
- Forrest, Robert Andrew Dermod. 1948. *The Chinese Language*. London: Faber and Faber.
- Fuller, Dorian Q., Harvey, Emma & Qin, Ling. 2007. Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity* 81: 316–331. doi:10.1017/S0003598X0009520X
- Fuller, Dorian Q. & Qin, Ling. 2009. Water management and labour in the origins and dispersal of Asian rice. *World Archaeology* 41 (1): 88–111. doi:10.1080/00438240802668321
- Fuller, Dorian Q. et al. 2009. The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323: 1607–1610. doi:10.1126/science.1166605
- Fuller, Dorian Q. & Qin, Ling. 2010. Declining oaks, increasing artistry, and cultivating rice: The environmental and social context of the emergence of farming in the Lower Yangtze Region. *Environmental Archaeology* 15 (2): 139–159. doi:10.1179/146141010X12640787648531
- Fuller, Dorian Q. et al. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* 2(2): 115–131. doi:10.1007/s12520-010-0035-y
- Fuller, Dorian Q. 2012. Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. *Rice* 4 (3–4): 78–92.
- Fuller, Dorian Q., Weisskopf, Alison R. & Castillo, Cristina Cobo. 2016. Pathways of rice diversification across Asia. *Archaeology International* 19: 84–96. doi:10.5334/ai.1915
- Garris, Amanda J., Tai, Thomas H., Coburn, Jason, Kresovich, Steve & McCouch, Susan. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631–1638. doi:10.1534/genetics.104.035642
- Hackinger, Sophie et al. 2016. Wide distribution and altitude correlation of an archaic high-altitude adaptive EPAS1 haplotype in the Himalayas. *Human Genetics* 135 (4): 393–402. doi:10.1007/s00439-016-1641-2
- Hazarika, Manjil. 2005. Neolithic Culture of Northeast India with Special Reference to the Origins of Agriculture and Pottery. MA thesis, Deccan College, Pune.
- Hazarika, Manjil. 2006. Neolithic culture of northeast India: A recent perspective on the origins of pottery and agriculture. *Ancient Asia* 1: 25–43. doi:10.5334/aa.06104
- Hazarika, Manjil. 2013. Prehistoric cultural affinities between Southeast Asia, East Asia and Northeast India: An exploration. In *Unearthing Southeast Asia's Past*, Vol. 1, Marijke J. Klokke & Véronique Degroot (eds), 16–25. Singapore: National University.
- Hazarika, Manjil. 2014. Austroasiatic and Tibeto-Burman Linguistic Dispersals in Northeast India: An Archaeological Perspective. PhD dissertation, University of Bern.
- Hazarika, Manjil. 2017. *Prehistory and Archaeology of Northeast India*. New Delhi: OUP. doi:10.1093/oso/9780199474660.001.0001
- von Heine-Geldern, Robert. 1917. Kopffjagd und Menschenopfer in Assam und Birma und ihre Ausstrahlungen nach Vorderindien. *Mitteilungen der Anthropologischen Gesellschaft in Wien* XXXVII: 1–65.
- Higham, Charles & Lu, Tracey. 1998. The origins and dispersal of rice cultivation. *Antiquity* 72: 867–877. doi:10.1017/S0003598X00087500

- Ho Dah-an (何大安). 2016. Such errors could have been avoided: Review of *Old Chinese: A New Reconstruction* by William H. Baxter and Laurent Sagart. *Journal of Chinese Linguistics* 44(1): 175–230. doi:10.1353/jcl.2016.0004
- Hu, Yaowu, Wang, Shougong, Luan, Fengshi, Wang, Changsui & Richards, Michael P. 2008. Stable isotope analysis of humans from Xiaojingshan site: Implications for understanding the origin of millet agriculture in China. *Journal of Archaeological Science* 35: 2960–2965. doi:10.1016/j.jas.2008.06.002
- Huang, Xuehui *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–503. doi:10.1038/nature11532
- Huerta-Sánchez, Emilia *et al.* 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512: 194–197. doi:10.1038/nature13408
- Illumäe, Anne-Mai *et al.* 2016. Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *American Journal of Human Genetics* 99: 163–173. doi:10.1016/j.ajhg.2016.05.025
- Karafet, Tatiana M., Fernando L. Mendez, Herawati Sudoyo, J. Stephen Lansing & Michael F. Hammer. 2015. Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *European Journal of Human Genetics* 23: 369–373.
- Karmin, Monika *et al.* 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research* 25 (4): 459–466. doi:10.1101/gr.186684.114
- Kivisild, Toomas. 2014. Mapping chromosome diversity in Eurasia. Paper presented at the conference Migrations and Transfers in Prehistory Asian and Oceanic Ethnolinguistic Phylogeography, University of Bern, 28 to 30 July.
- von Klaproth, Julius Heinrich. 1823. *Asia Polyglotta*. Paris: A. Schubart.
- von Klaproth, Julius Heinrich. 1830. Réponse à quelques passages de la préface du roman chinois intitulé: Hao khieou tchhouan, traduit par M. J. F. Davis. *Journal Asiatique* V: 97–122.
- Kovach, Michael J., Sweeney, Megan T. & McCouch, Susan. 2007. New insights into the history of rice domestication. *Trends in Genetics* 23(11): 578–587. doi:10.1016/j.tig.2007.08.012
- Kovach, Michael J., Calingacion, Mariafe N., Fitzgerald, Melissa A. & McCough, Susan R. 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences of the United States of America* 106 (34): 14444–14449. doi:10.1073/pnas.0904077106
- Larsona, Greger *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* 111(17): 6139–6146. doi:10.1073/pnas.1323964111
- Li Liu. 2015. A long process towards agriculture in the middle Yellow River valley, China: Evidence from macro- and micro-botanical remains. *Journal of Indo-Pacific Archaeology* 35: 3–14.
- Londo, Jason P., Chiang, Yu-chung, Hung, Kuo-Hsiang, Chiang, Tzen-Yuh & Schaal, Barbara A. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences of the United States of America* 103(25): 9578–9583. doi:10.1073/pnas.0603152103
- Lorenzo, Felipe R. *et al.* 2014. A genetic mechanism for Tibetan high-altitude adaptation. *Nature Genetics* 46(9): 951–958. doi:10.1038/ng.3067
- Lou, Haiyi *et al.* 2015. A 3.4-kb copy-number deletion near *EPAS1* is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. *American Journal of Human Genetics* 97 (1): 54–66. doi:10.1016/j.ajhg.2015.05.005
- Lu, Dongsheng *et al.* 2016. Ancestral origins and genetic history of Tibetan highlanders. *American Journal of Human Genetics* 99 (3): 580–594. doi:10.1016/j.ajhg.2016.07.002

- Lu, Tracey (= Lǚ Lièdān). 2006. The occurrence of cereal cultivation in China. *Asian Perspectives* 45 (2): 129–158. doi:10.1353/asi.2006.0022
- Lu, Tracey (= Lǚ Lièdān). 2009. Prehistoric coexistence: The expansion of farming society from the Yangtze River Valley to western South China. *Asian Perspectives (Senri Ethnological Studies)* 74: 47–52.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman*. Berkeley CA: University of California Press.
- McNally, Kenneth L. et al. 2009. Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America* 106(30): 12273–12278. doi:10.1073/pnas.0900992106
- Mirabal, Sheyla et al. 2009. Y chromosome distribution within the geolinguistic landscape of northwestern Russia. *European Journal of Human Genetics* 2009: 1–14.
- Molina, Jeanmaire et al. 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America* 108(20): 8351–8356. doi:10.1073/pnas.1104686108
- Müller, Friedrich Max. 1872. *Über die Resultate der Sprachwissenschaft: Vorlesung gehalten in der kaiserlichen Universität zu Strassburg am XXIII. Mai MCCCCLXXII*. Strassburg: Karl J. Trübner & London: Trübner & Co.
- Nakamura Shin-ichi. 2010. The origin of rice cultivation in the Lower Yangtze region. *China. Archaeological and Anthropological Sciences* 2(2): 107–113. doi:10.1007/s12520-010-0033-0
- Nasu, Hirō, Momohara, Arata, Yasuda, Yoshinori & He, Jiejun. 2006. The occurrence and identification of *Setaria italica* (L.) P. Beauv. (foxtail millet) grains from the Chengtoushan site (ca. 5800 cal B.P.) in central China, with reference to the domestication centre in Asia. *Vegetation History and Archaeobotany* 16(6): 481–494. doi:10.1007/s00334-006-0068-4
- Ostapirat, Weera. 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 107–131. London: Routledge Curzon. doi:10.4324/9780203343685_chapter_7
- Ostapirat, Weera. 2013. Austro-Tai revisited. Paper presented at the 23rd Annual Meeting of the Southeast Asian Linguistic Society, Chulalongkorn University, 29 May.
- Oxenham, Marc F. et al. 2015. Emergence and diversification of the Neolithic in southern Vietnam: Insights from coastal Rach Nui. *Journal of Island and Coastal Archaeology* 10(3): 309–338. doi:10.1080/15564894.2014.980473
- Pal, J. N. 1990. The early farming culture of northern India. *Bulletin of Deccan College Post-Graduate and Research Institute* 49: 297–304.
- Parsons, B. J., Newbury, H. J. & Jackson, M. T. 1999. The genetic structure and conservation of aus, aman and boro rices from Bangladesh. *Genetic Resources and Crop Evolution* 46(6): 587–598. doi:10.1023/A:1008749532171
- Pictet, Adolphe. 1859, 1863. *Les origines indo-européennes ou les aryas primitifs: Essai de paléontologie linguistique (première partie, seconde partie)*. Paris: Joël Cherbuliez.
- Poloni, Estella Simone et al. 1997. Human genetic affinities for Y chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *American Journal of Human Genetics* 61: 1015–1035 (cf. the erratum published in 1998 in the *American Journal of Human Genetics* 62: 1267). doi:10.1086/301602

- Poloni, Estella Simone, Ray, Nicolas, Schneider, Stefan & Langaney, André. 2000. Languages and genes: Modes of transmission observed through the analysis of male-specific and female-specific genes. In *Proceedings: Evolution of Language, 3rd International Conference 3–6 April 2000*, Jean-Louis Dessalles & Laleh Ghadakpour (eds), 185–186. Paris: École Nationale Supérieure des Télécommunications.
- Pulleyblank, Edwin George. 1983. The Chinese and their neighbours in prehistoric and early historic times. In *The Origins of Chinese Civilization*, David N. Keightley (ed.), 411–466. Berkeley CA: University of California Press.
- Ratliff, Martha. 2004. Vocabulary of environment and subsistence in the Hmong-Mien protolanguage. In *Hmong/Miao in Asia*, Nicholas Tapp, Jean Michaud, Christian Culas & Gary Yia Lee (eds), 147–165. Chiang Mai: Silkworm Books.
- Ratliff, Martha. 2010. *Hmong-Mien Language History*. Canberra: Pacific Linguistics.
- Reddy, B. Mohan *et al.* 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *Public Library of Science One* 2 (11): e1141. doi:10.1371/journal.pone.0001141.
- Rootsi, Siiri *et al.* 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15: 204–211. doi:10.1038/sj.ejhg.5201748
- Ruddiman, William F., Guo, Zhengtang, Zhou, Xin, Wu, Hanbin & Yu, Yanyan. 2008. Early rice farming and anomalous methane trends. *Quaternary Science Reviews* 27: 1291–1295. doi:10.1016/j.quascirev.2008.03.007
- Sahoo, Sanghamitra *et al.* 2006. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences of the United States of America* 103(4): 843–848. doi:10.1073/pnas.0507714103
- Schlegel, Gustave. 1901. Review: *Elements of Siamese Grammar* by O. Frankfurter, PhD. Bangkok: Printed at the American Presbyterian Mission Press, Leipzig, Karl W. Hiersemann, 1900, *T'oung Pao* (Série II) II: 76–87.
- Schlegel, Gustave. 1902. *Siamese Studies* [T'oung Pao, New Series II, Volume II, Supplement]. Leiden.
- Sharma, Govardhan R., Misra, V. D., Mandal, D., Misra, B. B. & Pal, J. N. 1980. *Beginnings of Agriculture: From Hunting and Food Gathering to Domestication of Plants and Animals*. Allahabad: Abhinav Prakashan.
- Sidwell, Paul & Blench, Roger. 2011. The Austroasiatic *Urheimat*: The southeastern riverine hypothesis. In *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, Nick J. Enfield (ed.), 317–345. Canberra: Pacific Linguistics.
- Silva, Fabio *et al.* 2015. Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. *PLOS One* 21. <doi:10.1371/journal.pone.0137024>.
- Starosta, Stanley. 2005 (posthumous). Proto-East-Asian and the origin and dispersal of languages of East and Southeast Asia and the Pacific. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 182–197. London: Routledge Curzon (Editorial misrepresentations that had crept into Starosta's posthumously published East Asian phylogeny were rectified that same year (cf. van Driem 2005)). doi:10.4324/9780203343685_chapter_11
- Stevens, Chris J. *et al.* 2016. Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *The Holocene* 26(10): 1541–1555. doi:10.1177/0959683616650268

- Strecker, David. 1987. The Hmong-Mien languages. *Linguistics of the Tibeto-Burman Area* 10(2): 1–11.
- Sweeney, Megan T. & McCouch, Susan R. 2007. The complex history of the domestication of rice. *Annals of Botany* 100: 951–957. doi:10.1093/aob/mcm128
- Sweeney, Megan T. *et al.* 2007. Global dissemination of a single mutation conferring white pericarp in rice. *Public Library of Science – Genetics* 3(8): 1418–1424.
- Vavilov, Nikolaj Ivanovič. 1926. Centry proisxoždenija kulturnyx rastenij (Studies on the origin of cultivated plants). *Trudy po Prikladnoj Botanike, Genetike i Selekcii (Bulletin of Applied Botany, Genetics and Plant Breeding)* 16(2): 1–248.
- Wang, Chuan-Chao, Yan, Shi, Qin, Zhen-Dong, Lu, Yan & Ding, Qi-Liang. 2013. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *Journal of Systematics and Evolution* 51(3): 280–286. doi:10.1111/j.1759-6831.2012.00244.x
- Weber, Steve, Lehman, Heather, Barela, Timothy, Hawks, Sean & Harriman, David. 2010. Rice or millets: Early farming strategies in prehistoric central Thailand. *Archaeological and Anthropological Sciences* 2(2): 79–88. doi:10.1007/s12520-010-0030-3
- Xiang, Kun *et al.* 2013. Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Molecular Biology and Evolution* 30(8): 1889–1898. doi:10.1093/molbev/mst090
- Yan, Shi *et al.* 2011. An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics* 19(9): 1013–1015. doi:10.1038/ejhg.2011.64
- Yang, Xiaoyan *et al.* 2016. New radiocarbon evidence on early rice consumption and farming in South China. *The Holocene* 7.<doi:10.1177/0959683616678465>.
- Yao, Xiaotian *et al.* 2017. Improved phylogenetic resolution for Y-chromosome haplogroup O2a1c-002611. *Scientific Reports* 7: 1146. doi:10.1038/s41598-017-01340-z
- Zhao Zhijun. 2010. New data and new issues for the study of the origin of rice agriculture in China. *Archaeological and Anthropological Sciences* 2(2): 99–105. doi:10.1007/s12520-010-0028-x
- Zhāng Hóngmíng. 1998. Chinese etyma for river. *Journal of Chinese Linguistics* 26(1): 1–47.

Macrofamilies and agricultural lexicon

Problems and perspectives

George Starostin

Russian State University for the Humanities / Russian Presidential Academy,
Moscow

It is more or less self-evident that the origins of agriculture cannot be directly associated with the ancestral speakers of any of the commonly accepted, non-controversial language families such as Indo-European, Semitic, Dravidian, etc., since these origins go much deeper back in time than any of these ancestral languages. Consequently, in this paper I present a brief overview of some of the most promising, if controversial, hypotheses on deep-level language relationship between various linguistic stocks of Western and Central Eurasia in terms of whether or not there is a chance of reconstructing at least a small amount of agricultural terminology for such hypothetical entities as Proto-Nostratic, Proto-Sino-Caucasian, and Proto-Afroasiatic. The overview leads to the conclusion that some of the most archaic agricultural terminology in the Near East may be associated with the North Caucasian linguistic family and, possibly, also with Basque as its nearest genetic relative; at the same time, evidence of ancient agricultural lexicon in the Afroasiatic stock remains at best circumstantial, whereas evidence from various lineages of “Nostratic” is practically non-existent.

Keywords: origins of agriculture, long-range comparison, macrofamilies, linguistic paleontology

Introduction

The task of associating the earliest known archaeological cultures that must have belonged to agricultural societies with specific cultural and linguistic lineages is an important interdisciplinary challenge that, nevertheless, runs into obstacles some would consider virtually impassable. If, according to the general consensus (see, e.g., Larson et al. 2014), we agree to broadly associate the origins of Eurasian agriculture with the Levant and an approximate age of 12–10 thousand years BP, this implies that the first agricultural lexicon, comprising names for cultivated crops,

agricultural tools, and various processes involved in cultivation, must have arisen around the same time – presumably out of multiple semantic shifts, as names for wild plants and common physical processes were extrapolated onto the new “cultural” meanings. Unfortunately, if we consider all known linguistic families whose protolanguages satisfy the following three conditions:

1. they have been reconstructed to more or less general satisfaction, so that their historical reality is not a point of contention between the majority of specialists;
2. they are generally agreed to have contained at least a certain amount of semantically unambiguous agricultural terms;
3. they are generally agreed to have been spoken either in the Levant area or in regions not too far removed from it;

– then none of these families, including such linguistic taxa as Indo-European, Semitic, Dravidian, Kartvelian, North Caucasian (Nakh-Daghestanian and Abkhaz-Adyghe), can be reasonably claimed to have had their ancestral languages spoken as early as the required date.

Available evidence, ranging from glottochronological dating to linguistic/archaeological correlations, indirectly indicates that all these protolanguages must have disintegrated not earlier than 5–6 millennia ago (see the brief overview given in Gell-Mann, Peiros & Starostin 2009), by which time agriculture had already had an established presence of at least several thousand years in the Near East. Attempts to significantly extend the chronological range of some of these families based on circumstantial evidence – the most notable of these is Colin Renfrew’s famous hypothesis on the spread of farming driven by Indo-European migrations (Renfrew 1987), allegedly supported by certain modern quantitative studies (Gray & Atkinson 2003; Bouckaert et al. 2012), but hardly by any proper linguistic or archaeological evidence (Anthony 2013; Pereltsvaig & Lewis 2015).

Unfortunately, what this means is that the issue of identifying the cultural lineages of Eurasia’s first farmers is inextricably linked to the controversial issue of deep-level linguistic reconstruction, where commonly accepted and generally reconstructible linguistic stocks are linked together in “macrofamilies”, or “phyla”, in a set of usually controversial hypotheses. It is theoretically possible, of course, that the earliest farmers, such as the ones who tilled the virgin soil at Tell Abu Hureyra about 12,000 years ago, spoke a language or several languages that became a linguistic dead end, leaving no traceable lineage other than a substrate layer in other linguistic stocks – a situation that is routinely encountered all over the Old World; cf., for instance, the old pre-Indo-European substrates in Europe that “donated” a large part of their cultural lexicon to the Indo-Europeans before disappearing off the radars of history; see, e.g., Kroonen 2012 on how such a substrate could have affected the agricultural lexicon in Germanic and other European branches

of Indo-European. However, this only implies the futility of asking ourselves the question, “which of the known linguistic stocks goes directly back to the language of the inventors of agriculture?” – a question that hardly makes even speculative sense, considering that the transition to a sedentary agricultural lifestyle must have taken place in several stages, including a lengthy “pre-domestication” phase (Harris 1989), and could have involved the speakers of multiple communities with different cultural and linguistic backgrounds.

A question that makes far more sense is, “to which earliest linguistic stocks, as hypothetical as they currently seem to be, can we ascribe a potentially reconstructible layer of agricultural lexicon?” Naturally, this would hardly make sense to a rigorously skeptical comparative linguist, who would rightfully object that talking about potential agricultural lexicon in a protolanguage like Nostratic, whose very existence, according to the skepticist’s opinion, has not been demonstrated beyond reasonable doubt by conventional means (grammar, basic lexicon, etc.), would be the equivalent of having a serious talk about the dietary habits of the Loch Ness monster. But for those who take a somewhat more flexible position on the issue of macrofamily hypotheses – in particular, those who are willing to evaluate the evidence for macrofamilies on a graduated basis, distinguishing between more and less probable connections based on the quantity and quality of presented argumentation (see discussion in Starostin 2014) – an investigation of possible links on the level of cultural lexicon, including agriculture, were it to lead to positive results, could add to the seriousness of the overall argument. In fact, even an *areal* interpretation of any such links, conducted in the “diffusionist” paradigm rather than arguing for inheritance of the terms in question from a common ancestor, could shed some light on the cultural and linguistic backgrounds of the earliest farmers.

With this particular goal in mind, in this paper I would like to briefly survey some of the better elaborated hypotheses on Eurasian macrofamilies in terms of whether they have anything to say on the potential agricultural leanings of their original speakers, and to point out which of them seem to offer more promise for future research on this issue and which ones would probably represent a dead end in this respect. Although the survey will inevitably suffer from being superficial and omitting lots of important details (some of which may be, however, easily looked up by means of provided references), I hope that it will help to paint a relatively concise comparative picture that may be useful for linguists and non-linguists alike. Note that, due to considerations of volume and in order to maintain a tighter focus, I will limit myself exclusively to discussions of families with possible ties to the Near Eastern agricultural centers, leaving aside the almost equally important issue of – probably independent – agricultural origins in South Asia and the Far East, and, consequently, the agricultural terminology in such families as Austronesian or Austroasiatic, and exclusively to the sphere of farming, leaving aside the also equally

important issue of pastoral lexicon (naturally, it should not be taken as a given that languages without agricultural lexicon should also be expected to lack words having to do with domesticated animals – for instance, Proto-Altaiic may well have been a “pastoral”, but not an “agricultural” language, see below).

A. Nostratic

The borders of the Nostratic macrofamily, originally proposed by Holger Pedersen in 1903 and carefully reworked into a proper comparative-historical linguistic theory by Vladislav Illich-Svitych (Illich-Svitych 1971–1984) and Aharon Dolgopolsky (2008), have never been established to the complete satisfaction of all scholars who support the hypothesis in general. All versions of it, however, include at least a “core” constituency of Indo-European, Uralic, and Altaic (the latter limited to at least Turkic, Mongolic, and Tungusic, but possibly also including Korean and Japonic), while two other families, Kartvelian and Dravidian, are sometimes seen as more remotely connected with the former three. Afroasiatic, a huge macrofamily in its own rights, has always figured prominently as a sub-member of Nostratic in the theories of Illich-Svitych and Dolgopolsky, as well as several other prominent Nostraticists, e.g. Allan Bomhard, but has been excluded from the macrofamily by Sergei Starostin and other members of the Moscow school of comparative linguistics largely on the basis of its internal and external lexicostatistical data, suggesting that it should rather be treated as a “sister” than “daughter” family of Nostratic proper (Gell-Mann, Peiros & Starostin 2009: 23); in any case, the distance between various branches of Afroasiatic is so enormous that Afroasiatic data deserve their own special treatment (see below). Inclusion into Nostratic of other small families and language isolates of Eurasia, such as Chukchee-Kamchatkan (an old idea of A. Dolgopolsky), Eskimo-Aleut (Oleg Mudrak), and Sumerian (Allan Bomhard), has not found widespread acceptance among the general community of Nostraticists so far.

Of the three “core” and two “peripheral” hypothetical branches of Nostratic, three are most definitely traced back to protolanguages whose speakers practiced at least some form of agriculture, namely, Indo-European; Kartvelian (with self-evident terms such as **qan-* ‘to plough’, etc., see Fähnrich 2007: 699); and Dravidian (Krishnamurti 2003: 8–9, with a list of uncontroversially reconstructed Proto-Dravidian farming terms). For Uralic, comparison between its two primary branches, Finno-Ugric and Samoyed, yields no signs of agriculture (Hajdu 1975); however, this by no means counts as sufficient proof that speakers of Proto-Nostratic could not have agricultural practices, either: situations in which former agriculturalists switch back to a hunting-gathering or pastoral lifestyle are well known, and if the

Proto-Uralic homeland, as is sometimes suspected, was not conducive to agriculture (e.g. if we place it in West Siberia; see Janhunen 2009: 71), any traces of an earlier agricultural lexicon may have been lost in the language together with the practice itself. As for Altaic, the situation here is as controversial as the Altaic hypothesis itself and will be separately discussed below: for now, it suffices to say that most individual members of Altaic, with the notable exception of the Tungusic hunter-gatherers, show at least some familiarity with agriculture at the ancestral level, but these ancestral levels by themselves are quite shallow and hardly diagnostic of much earlier times.

Regarding Nostratic itself, to the best of our knowledge, no attempts have ever been made to prove that such an ancient protolanguage might have possessed an agricultural lexicon. No agricultural terms were presented in Illich-Svitych's original dictionary, and Dolgopolsky (1998: 26–28) explicitly states that Proto-Nostratic had “no words for specifically agricultural activities (sowing, ploughing, harrowing, etc.)”, although he presents evidence for several terms with ambiguous semantics, such as ‘to harvest (cereal)’, ‘edible cereals’, and ‘kernel, grain’, all of which, even assuming that the etymologies represent historical reality, could be interpreted in terms of a gatherer lifestyle. Not a single unambiguous agricultural term is reconstructed in any of the editions of Allan Bomhard's Nostratic dictionary, including the latest (Bomhard 2014), where any such terms in daughter branches of Nostratic are usually traced back to verbs with more general semantics, e.g. ‘to sow’ ← ‘to throw, cast, scatter, etc.’ The latter circumstance is particularly important, since a successful demonstration of the secondary origins (through semantic shifts) of agricultural terminology in particular families qualifies as a near-conclusive argument that their remote ancestors were not familiar with farming practice at all.

The impossibility of reconstructing agriculture for Proto-Nostratic will probably be perceived with satisfaction by both opponents of the Nostratic hypothesis, for whom the absence of Proto-Nostratic agricultural terminology is easily explicable by the phantom nature of Nostratic as such, and its proponents, who would rather date the existence of Proto-Nostratic to the pre-agricultural stage of the Neolithic period or even earlier – based on linguistic paleontology, as is done by Dolgopolsky (1998), or on lexicostatistics, as practiced by the Moscow school of comparative linguistics (Gell-Mann, Peiros, & Starostin 2009). There are, however, at least two widely discussed intermediate nodes between the non-controversial groupings and the highly hypothetical Proto-Nostratic that also deserve consideration and could theoretically turn out to be more revealing, since their protolanguages are naturally younger than Proto-Nostratic: namely, Indo-Uralic and Altaic (or “Transeurasian”, a new term for the same taxonomic entity used by such contemporary scholars as Martine Robbeets).

Indo-Uralic, a somewhat less ambitious hypothesis than Nostratic, is based on a series of very significant isoglosses between Proto-Indo-European and Proto-Uralic in the subsystems of grammatical morphemes and basic lexicon, unlikely to have arisen by chance and, even though this is not a consensus opinion, better correlated with a scenario of common descent than areal diffusion; see, e.g., Kortlandt (2010) on grammatical evidence and Kassian, Zhivlov & Starostin (2015) on evidence from the basic lexicon. Indo-European/Uralic lexical isoglosses often tend to be semantically conservative (cf. PIE **wed-r-*: PU **wete* ‘water’, PIE **wedh-*: PU **wetä* ‘to lead’, PIE **klew-*: PU **kule* ‘to hear’, etc.), and this could theoretically help recover at least a few agricultural matches, had they been present in Proto-Indo-Uralic. However, as we have already stated above, no agricultural lexicon is reconstructible for Proto-Uralic, and what is even worse, no agricultural terms reliably reconstructed for Proto-Indo-European, such as **yew-* ‘cultural crop’ or **arə-* (**h₂erh₃-* in the laryngealistic notation) ‘to plough’, seem to find any solid etymological cognates in Proto-Uralic, which inevitably leads to the suspicion that many, if not most, of these terms may have developed their agricultural meanings through internal Indo-European semantic shifts or introduced into Proto-Indo-European from an outside source, and that Proto-Uralic really did not lose the original agricultural lexicon, but never had it in the first place.

With Altaic, the situation is even more complicated. Compared to the vast amount of literature on the reconstruction of Proto-Altaic phonology, morphology, and basic lexicon, as well as the comparable amount of critical debate on Altaic, pro-Altaicist works attempting to apply the *Wörter und Sachen* method to Altaic are few and far between. The single largest corpus of Altaic etymologies up to date (Starostin, Dybo & Mudrak 2003) contains hundreds of comparanda that certainly belong to the cultural lexicon, but it is often these particular comparanda that draw particularly strong critical fire due either to sheer implausibility, e.g. an attempt to reconstruct the word ‘bridle’ for Proto-Altaic, or semantic permissiveness; see the anti-Altaicist review in Vovin (2005) and the pro-Altaicist analysis of the evidence in Robbeets (2005) for various specific criticisms.

A useful, though not easily accessible, overview of the various spheres of Proto-Altaic cultural lexicon is provided in Dybo (2000), where the author summarizes the results as indicative of a culture well familiar with hunting and pastoralism, but, at best, a rudimentary understanding of agriculture, with but one verb that could be interpreted as having something to do with tilling the soil and two names for possible tools, as well as possible equivalents for ‘millet’ and ‘barley’. The verb in question, reconstructed as **t^hiora*, is based on Proto-Turkic **TArī-* ‘to cultivate’, Proto-Mongolic **tarija-n* ‘crops’, and Proto-Japanese **tā* ‘cultivated field’, but the comparison with Japanese is phonetically problematic, and without it, the etymology is reduced to a Turkic-Mongolic match that not only could be interpreted

in terms of contacts, but even if it were genetic, would only be reconstructible on the much more shallow Turko-Mongolic level rather than Altaic in general. Likewise, the word **arp^ha* is reconstructed based on Proto-Turkic **arpa* ‘barley’, Proto-Mongolic **arbaj* ‘barley’, and Proto-Japanese **ápá* ‘millet’ (Manchu *arfa* ‘barley’, also adduced in the etymology, is almost unquestionably a Mongolism), but the Japanese word is semantically different and the connection is dubious, once again reducing the etymology to an areal Turko-Mongolic isogloss that may have spread as a result of diffusion, especially if its roots, as is sometimes suggested, really lie in an older Proto-IE stem (Robbeets 2017: 28).

On the whole, the vast majority of agricultural isoglosses between the various branches of Altaic seem to extend to no more than two branches, at least as far as phonetically and semantically acceptable parts of the etymologies are concerned, usually those in geographical proximity to each other. Despite this, Robbeets (2017) still makes an interesting attempt to associate Proto-Altaic with the millet-cultivating cultures of Manchuria and Inner Mongolia since the 7th millennium BC, offering a special etymology in support of this hypothesis that unites Proto-Tungusic **pise* ‘seed, millet’ with Proto-Korean **pisi* ‘seed’ ~ **pihi* ‘barnyard millet’ and Proto-Japanese **piy-* ‘barnyard millet’ (the Tungusic-Korean match goes all the way back to G. J. Ramstedt’s old works, but the Japanese connection is new). Unfortunately, no parallels are attested in either Turkic or Mongolic, once again making this an areal isogloss that could have very easily become diffused across the Manchurian/Korean region.

Summing up, we should probably acknowledge that evidence for associating either the origins of agriculture in the Near East or its initial spread across Eurasia with languages of the Nostratic lineage, regardless of whether the Nostratic hypothesis is understood in genetic or in areal terms, is quite flimsy at best. Interfamily connections in the sphere of agricultural lexicon here tend to be sporadic, dubious, and usually geographically contiguous, as in the case of various branches of Altaic.

B. Sino-Caucasian

“Sino-Caucasian” is the common designation of a proposal by Sergei Starostin (1984), who, building upon the earlier comparisons of several researchers (e.g. Bouda 1936, 1957), claimed to have had established regular phonetic correspondences and discovered a core of several hundred common etymologies for three far-flung linguistic families of the Old World: North Caucasian (a somewhat problematic taxon by itself, since many Caucasologists remain skeptical about a genetic connection between its two primary constituents, Northeast and Northwest Caucasian), Yeniseian, whose only remaining modern descendant is Ket, and

Sino-Tibetan. Since then, following up on various ideas dating back to the early 20th century, the hypothesis has been expanded by also including such Eurasian isolates as Basque and Burushaski, as well as the Na-Dene languages of North America, which is why the macrofamily is often referred to as “Dene-Caucasian”, particularly in Western sources; see Bengtson 2008a for a general overview of the expanded hypothesis.

Of the six potential branches of Sino-Caucasian, two are traced back to non-agricultural lineages: Yeniseian and Na-Dene; not coincidentally, the alternate “micro”-hypothesis of a binary connection specifically between these two families, recently put forward by Edward Vajda (2011), avoids any references to agriculture. This is not too surprising, considering the proposed original homelands for both families (Central Siberia, far more suitable for a hunting-gathering than for an agricultural lifestyle, in the case of Yeniseian; Beringian region for Na-Dene), and, as in the case of Uralic, should not be taken as proof that Proto-Sino-Caucasian had no knowledge of agriculture. On the other hand, Proto-Sino-Tibetan is commonly agreed to have had at least a few terms for culture crops (Sagart 2003), and Proto-North Caucasian is reconstructed by Sergei Starostin and Sergei Nikolayev as reflecting a very elaborately developed agricultural and pastoral lifestyle, with numerous terms for cereals and agricultural processes (Starostin & Nikolayev 1994; Starostin 2004).

Regarding Sino-Caucasian itself, the etymological corpus assembled for the macrofamily by Starostin (2005) does include a certain number of terms that could be interpreted in agricultural terms, e.g. Proto-NC **Hra:jcu:* ‘wooden plough, mattock’: Proto-ST **ru:jH* ‘part of a plough’: Burushaski **hars* ‘plough’: Proto-NC **s:wizer* ‘a k. of cereal (barley, millet)’: Proto-ST **si* ‘fruit, grain, seed’; amusingly, even Proto-NC **regwe* ‘yoke’: Old Chinese **rək* id. (!). However, the absolute majority of all such matches are questionable in terms of semantics, phonetics, and interfamily distribution, and do not seem too impressive even to general supporters of the hypothesis, myself included. For Sino-Tibetan, in particular, it seems more prudent to look for potential parallels in the sphere of agricultural lexicon in other language families of the Southeast Asian region, such as Austronesian; cf. in this respect the research conducted by L. Sagart (2003, 2005), who leans towards a genetic connection between Sino-Tibetan and Austronesian, although the best etymological parallels in his works can easily have an areal interpretation as well.

However, leaving out the potential “Eastern” branches of Sino-Caucasian for the moment, it is quite instructive to pay closer attention to the Caucasian part of the equation. Proto-North Caucasian is reconstructed by S. Starostin as a language abundant in agricultural terminology (Starostin 2004), including terms for activities (‘to plough’, ‘to reap’, ‘to thresh’), tools (‘ploughshare’, ‘sickle’), and various cereals (‘barley’, ‘wheat’, ‘millet’). While it is true that proper criteria for distinguishing

between etymological cognates and borrowings in North Caucasian have not been established to general satisfaction, the sheer number of agricultural terms that are well distributed across various branches and generally obey the system of regular correspondences set up for the family is far more impressive than the situation in Proto-Altaiic or even in Proto-Indo-European. Particularly important is the fact that the majority of these terms are *always* featured in daughter languages with specifically agricultural semantics – cf., for instance, Proto-NC **V:rLV* ‘to thresh’ → Proto-Nakh **ʔa:rL-*, Proto-Andi **=il-*, Proto-Tsez **=oL-* ‘to thresh’, Proto-Lezgian **jiʔ:* ‘threshing (noun)’, without any attested reflexes meaning ‘to hit’, ‘to pound’, etc., which could hint at a secondary origin for the agricultural meanings.

It should be pointed out that the seeming discrepancy between the current geographical distribution of most of the North Caucasian-speaking populations and the glottochronological dating of Proto-NC (at least 6000 years BP, if not more, according to S. Starostin’s calculations; at the same time, well-developed agriculture in the North Caucasian region seems to appear no earlier than 5000 years BP) can be resolved within a scenario that postulates the original North Caucasian homeland not in the Caucasian mountains themselves, but further south, perhaps closer to the Zagros range and, consequently, also closer to the original agriculture spread zone.

In any case, despite all the remaining problems, a deep-running association between North Caucasian (or at least Northeast Caucasian, i. e. Nakh-Daghestanian) and an agricultural as well as pastoralist lifestyle is practically undeniable at this point. Whether these ties go even deeper, all the way to the highly hypothetical Sino-Caucasian level, is unclear and dubious. However, at least one particular link deserves very careful consideration – namely, the connection between North Caucasian and Basque, going all the way back to the old “Ibero-Caucasian” hypothesis and recently refined by John Bengtson within his “Euskaro-Caucasian” theory (Bengtson is a supporter of the Dene-Sino-Caucasian connection, but his primary research for the past several decades has focused on the binary link between North Caucasian and Basque).

As in the case of Altaic, it should be remarked that certain cultural isoglosses between North Caucasian and Basque may deserve attention even regardless of whether the overall evidence for a genetic relationship between these two taxa is found convincing or not. The really important thing is that the agricultural isoglosses should not be between modern Basque and any of the modern North Caucasian languages, but between “Proto-Basque” (the oldest reconstructible form of Basque, furthermore, restricted to roots that cannot be shown to have been borrowed from a Latin/Romance source) and Proto-North Caucasian (i. e. roots attested in at least several different branches of the family).

The following etymologies, assembled by Bengtson (partially published in Bengtson 2008b and also available online as part of the general Sino-Caucasian

etymological database at the “Tower of Babel” website; all North Caucasian correlates are taken from Starostin & Nikolayev 1994), seem to be especially promising:

- a. Basque *gari* (*gal-* in compounds; reconstructed as **gali* in Trask 2008:200) ‘wheat’; cf. Proto-NC **G:o:lʔe* → Proto-Lezghian **q:ol*, Proto-Andian **qʔ:iru* ‘wheat’ – this is the only direct isogloss between two intermediate protolanguages meaning specifically ‘wheat’ and a strong candidate for the main designation of ‘wheat’ in Proto-North Caucasian;
- b. Basque *larrain* ‘threshing floor’; cf. Proto-NC **=V:rLV* ‘to thresh’ (see above). For Basque, one has to assume plausible metathesis, possibly due to the simplification of a complex cluster, and suffixal derivation of the noun from an original verb;
- c. Basque *e(i)ho* ‘to grind’; cf. Proto-NC **Hemχ:wV* ‘to grind’ → Proto-Nakh **ʔah-*, Proto-Avar-Andi **ʔiχʷVn-*, Proto-Lezghian **reχ:lʷa* ‘mill’ (from **r=Hemχ:wV* with a nominal prefix), etc. It must be noted that the reconstruction **Hemχ:wV* is based on certain oblique considerations (none of the daughter branches actually preserve the alleged labial nasal), and that a simpler **Hex:wV* is also possible, which would agree with the Basque form even better.

These comparanda, though limited in number (hardly surprising, considering the time and space that lies in between), are semantically exact, phonetically compatible, and present no problems in terms of distribution and topology, being reconstructible in this meaning at least for Proto-Nakh-Daghestanian, and most likely belonging to the inherited lexical stratum in Basque. Of course, there are also multiple additional parallels from the same semantic field in Bengtson’s works that are phonetically, semantically, or distributionally weaker, but still plausible to a degree – altogether, in my opinion, the evidence is sufficient to be taken very seriously, if not necessarily as conclusive proof for a common Euskaro-Caucasian agricultural basis.

If this hypothesis checks out, implications for our understanding of European prehistory would be huge, almost on the point of sensational. It is curious that a very rough lexicostatistical assessment of the hypothetical common ancestor of Basque and North Caucasian, recently conducted by the author of the present paper based on a comparison of the most stable half of the Swadesh 100-item wordlist, yielded a glottochronological dating of about 9000 years BP – more or less the same time as the appearance of the first signs of agriculture in the Balkans. Instead of trying, like Renfrew, to associate this spread with early waves of Indo-European migrations (a point of view hotly contested by the majority of historical linguists), it would then seem far more logical to ascribe it to a “Euskaro-Caucasian” wave, whose cultural lineages may have flourished all across Central and Western Europe before the subsequent waves of Indo-European migrants several millennia later, to

be eventually replaced everywhere except for a small “refugium” in the Pyrenees. At the very least, this seems like a reasonable scenario well worth exploring, and one that should also stimulate research on possible “Caucasian” origins of some of the enigmatic substrate lexicon, found in large numbers in various branches of Indo-European languages across Europe.

C. Afroasiatic

As of now, the Afroasiatic hypothesis, formulated already in the 19th century and significantly refined and elaborated over the course of the last hundred years, remains the only macrofamily-level hypothesis on languages of Eurasia and, in this case, partially also Africa that is viewed as generally accepted by the comparative-historical linguistic community, largely due to certain impressive grammatical isoglosses between its most distant members (see brief, but comprehensive overview in Hayward 2000), although certain problems with defining the borders of Afroasiatic remain pertinent, e.g., the status of the Omotic languages in Ethiopia (see Theil 2006, where the same type of critique is applied to the Afroasiatic status of Omotic as one usually encounters in works critical of Altaic, Nostratic, or Sino-Caucasian).

Common agricultural lexicon is perfectly well reconstructible for Proto-Semitic (Agmon 2010), as well as for multiple small Afroasiatic taxa in Africa; however, due to a certain amount of negligence as to the lexical reconstruction of individual daughter branches of Afroasiatic outside of the context of Proto-Afroasiatic, it is not easy to delineate and present, beyond reasonable doubt, a solid agricultural vocabulary for such deep levels as Proto-Chadic or Proto-Cushitic. (The most recent etymological dictionary of Proto-Chadic, Stolbova 2016, only presents scarce and circumstantial lexical evidence that could be unambiguously evaluated as “agricultural”; as for Cushitic, it seems to be such a chronologically deep linguistic entity that no etymological corpus for Proto-Cushitic has ever been produced independently of a corpus for the entire Afroasiatic macrofamily.)

Nevertheless, a serious attempt has been made by Alexander Militarev (2002) to demonstrate, based on a series of phonetically rigorous and semantically plausible comparanda, that an extensive agricultural as well as pastoral vocabulary is reconstructible for the deepest level of the Afroasiatic macrofamily, including reflexes in such remote branches as Cushitic and Omotic. Correlating the evidence with his own glottochronological dating of Proto-Afroasiatic to the early Neolithic epoch (Militarev 2000), Militarev proposes equating it with the language of the Natufian culture, i.e. the originators of agriculture in the Near East.

This hypothesis surmises a somewhat complex migration scenario. Although the only Asian branch of Afroasiatic is Semitic, there is no evidence of a binary split between Semitic and “Hamitic” languages – lexicostatistical and etymological data rather suggest a binary split between Cushitic (and maybe also Omotic), on one hand, and Semitic-Berber-Chadic-Egyptian, on the other, which would imply that, if Militarev’s hypothesis is to be accepted, we have to assume either that the ancestors of Proto-Semitic originally migrated into Africa from the Levant, separated in Africa from their closest relatives and then went back to the Near East – or that, as the Afroasiatic unity gradually split into several different lineages, there were several waves of migration to Africa: first the Cushites, then the Chadic, Berber, and Egyptian-speaking groups.

Naturally, this rather convoluted scenario runs against the intuitive opinions of many scholars, including Christopher Ehret, whose conception of Afroasiatic includes an African, rather than Near Eastern, homeland, and subsequent migration of the Semites to Asia, with agricultural practices and terminology developing independently across various lineages already after the split and much later than the 10th–8th millennia BC; according to Ehret, “the proto-Afroasiatic vocabulary included ... no words at all implying the herding of animals or the cultivating of crops” (Ehret 2000: 290–291). It should be kept in mind that Ehret’s statement is based on his own reconstruction of Proto-Afroasiatic (Ehret 1995), which has been frequently criticized on methodological grounds, especially for semantic overpermissiveness and lack of detailed attention to intermediate levels of reconstruction; however, these criticisms would rather be valid in the event of alleged positive rather than negative evidence for Afroasiatic agriculture, and, of course, the argument from topology, where the majority of scholars today come up with very similar tree diagrams for Afroasiatic (see the comparison between the lexicostatistical models of A. Militarev and G. Starostin in Blažek 2012), would seem to agree very strongly with Ehret’s etymological observations.

The real solution, however, lies not in deciding which of the proposed scenarios is more economic, but in whether the evidence presented by Militarev may really be interpreted, unambiguously and reliably, as beyond-reasonable-doubt proof of agricultural knowledge in Proto-Afroasiatic society. In order to do so, we should be sure that the evidence at least satisfies the same criteria that are found satisfactory for North Caucasian: namely, there should be at least several terms that would be formally reconstructible to the top levels of several primary branches of Afroasiatic (e.g. Proto-Cushitic vs. Proto-Semitic) with specifically agricultural semantics either directly matching across lineages or at least relatable to each other through trivial, typologically common semantic shifts (e.g. ‘to plough’ – ‘/a/ plough’, ‘wheat’ – ‘grain /of wheat/’, etc.). Considering the sheer number of Afroasiatic languages and the massive size of the dictionaries for some of them, most notably

Semitic, finding accidental look-alikes with vaguely resembling agricultural semantics is not in itself a difficult task; far more difficult is to demonstrate that these look-alikes are really traceable back in time across lineages and agree with each other in a concise, agreeable scenario of semantic shifts, replacements by other terms in some daughter branches and preservation in others.

Even a cursory analysis of the 32 terms in Militarev 2002, singled out as the best evidence for Proto-Afroasiatic farming practices, shows that, while they certainly do not *disprove* Proto-Afroasiatic agriculture (many of the matches look phonetically and semantically plausible in theory), they cannot be accepted as uncontroversial proof of its existence, either. Most worrisome is the near-complete lack of terms for cultural plants with unambiguously determined semantics. The only such item is **ĉarVy-* ‘barley’, and even that is reconstructed based on Proto-Semitic **šarVr-* ‘barley; grass, straw’, with Ethiosemitic and Modern South Arabic reflexes usually restricted to ‘grass, straw’; some scattered Chadic parallels meaning either ‘yam’ or ‘okra’; and an alleged Proto-Cushitic **ʔaĉar*, reflected as Beja *ešerri* ‘maize’ and as *ašaru-* ‘barley’ only in one East Cushitic language (Kambaata). The direct semantic isogloss is therefore confined to a few Semitic languages and one Cushitic language, making the semantic reconstruction ‘barley’ highly dubious. The degree of semantic lenience and topological scattering in other terms is even higher: for instance, **bar-* ‘a cereal’ has the meanings ‘wheat’, ‘maize’, ‘threshed grain’, ‘sorgho’, ‘yam’, ‘millet’, ‘ground nut’, ‘oats’, ‘stalk’, ‘straw’ in daughter languages – clearly, a huge amount of low-level semantic reconstruction is necessary here to ascertain which of these reflexes deserve to be grouped together and which ones are the result of accidental phonetic resemblance.

Likewise, names for alleged agricultural practices also leave a lot to be desired: for instance, PAA **ʔry/w* is reconstructed with the mixed semantics of ‘to gather, reap, cultivate’, and indeed, in many of the compared languages the meaning is simply ‘to gather’, but since the typological shift ‘to reap (cultured cereals, etc.)’ → ‘to gather (any plants, including wild)’ is virtually unknown, if all the compared etyma are indeed genetically related, the original meaning of the term surely must have been simply ‘to gather’, and would rather point to individual shifts to the meaning ‘to reap’ in those branches whose speakers had gradually shifted to agricultural practices. (For the sake of accuracy, the meaning ‘to gather’ is attested in languages whose speakers were or still are farmers, so there can be no explanation through “regression” to a foraging lifestyle in this case). Akkadian *šakāku* ‘to harrow’ (no reliable parallels in other Semitic languages) is compared with Chadic **swk* ~ **skw* that means ‘to sow’ in most of the languages where it is attested – again, not an impossible connection, but the two processes are really quite distinct, and whether they can go back to a common semantic denominator is unclear: ‘harrowing’ is usually connected with ‘hitting’ or ‘piercing’, whereas ‘sowing’ is closely associated

with ‘throwing’. These examples could be easily multiplied, but they are really quite typical of the overall material.

On the whole, I have to state that there is not a single common Afroasiatic root that would be phonetically, semantically, and distributionally comparable with the several examples of common “Euskaro-Caucasian” agricultural lexicon listed above. Particularly telling is the lack of convincing parallels between Cushitic, Omotic, and “Narrow Afroasiatic”, comprising all the other branches; at the same time, a few cultural isoglosses between Semitic and Chadic (more rarely, Berber and Egyptian) inspire more confidence about agriculture on the “narrow” level, already after the separation of Cushitic (such as Semitic **tizin-* ‘fig-tree’, Berber **tiHVyn-* ‘fig tree’, Chadic **tizun-/ *tizan-* ‘mahogany, fig tree’, etc.), but even on that level it is hard to find one etymon that would unquestionably qualify as the optimal candidate for a specific agricultural meaning on the level of at least two proto-branches.

An additional problem is separating borrowings from inherited lexicon: in many cases, particularly where areas of intense linguistic contact are concerned, such as Cushitic – Ethiosemitic/Arabic, Cushitic – Omotic, Arabic and the other Semitic languages, etc., the author diligently marks potential situations of borrowing, but when in doubt, always seems to favor the genetic solution. For instance, he doubts that a certain batch of Cushitic and Omotic forms, reduced to the prototype of **ǰVrɪ/y/w-* ‘seed; to sow’ could be borrowed from Ethiosemitic reflexes of Proto-Semitic **zrɪ* (Militarev 2002: 148), since the forms are fairly widely distributed across several branches of Cushitic. He does not, however, take into consideration that the only branch of Cushitic where these forms are not attested is Southern Cushitic – precisely the one branch that lacks any contact with speakers of Ethiosemitic. Furthermore, it does not strike him as suspicious that there are almost no potential cognates for this root in Berber, Chadic, and Egyptian – languages that, according to consensus topology, are much closer to Semitic than to Cushitic and Omotic, and should be expected to preserve more reliable traces of such an important root than Pero (a single West Chadic language in Nigeria) *ǰúra* ‘groundnuts’ (!) and late (!) Egyptian *zɪ* ‘a kind of field’ (!). In my opinion, the facts inescapably point here towards an internal Semitic origin for this root, later borrowed into some non-Semitic languages of Ethiopia.

As an interesting curio, it could be instructive to mention a recent study (Agmon & Bloch 2013) that used statistical methods to ascertain that various terms reflecting hunting and foraging activities in Semitic tend to be shorter, i.e. are more frequently represented by archaic biconsonantal roots than agricultural terms, which, conversely, tend to be almost always represented by longer, triconsonantal roots. If this study checks out through detailed etymological research, this could be a serious argument in favor of a relatively late origin of agricultural terminology for ancestors of Proto-Semitic. For now, we simply have to accept the

fact that a lot of research on various subgroups of Afroasiatic is still necessary in order to properly resolve the issue – and that, for the moment, strong evidence for agriculture in Proto-Afroasiatic is non-existent.

Conclusion

As skeptical as one can be about the idea of reconstructing agricultural lexicon on chronological levels of such depth where reconstruction of *any* lexicon is usually considered problematic, it can hardly be denied that different linguistic lineages, submitted to etymological scrutiny, do not always yield the same result. In some cases, such as Nostratic, reported results amount to zero; in other cases, such as Altaic or Afroasiatic, potential comparanda have been found, but their interpretation remains ambiguous, with genetic, areal, and chance similarities hard to distinguish from each other; finally, in one case, namely, “Euskaro-Caucasian”, results of comparison are clearly more impressive than in any other case, and seem to be relevant even if one wishes to interpret them through a contact scenario, because in this case, one would still have to assume that the original speakers of Basque once dwelled in close proximity to speakers of North Caucasian languages.

In any case, explorations like these are probably as close as we can ever get in the quest to align archaeological evidence with linguistic and cultural lineages. The following tasks, in particular, can be seen as challenging, but potentially quite rewarding:

1. further research on agricultural terminology in Proto-North Caucasian, aiming at a more detailed inventory of inherited agricultural lexicon in modern languages (based on a wealth of new sources published in recent decades, already after the appearance of Starostin and Nikolayev’s dictionary), a refining of the currently available list of reconstructions, and direct confirmation, by means of internal analysis, that the Proto-North Caucasian agricultural terminology is truly archaic, rather than generated during some late stage of North Caucasian through semantic shifts;
2. further research on the connections between North Caucasian and Basque that could also, perhaps, be strengthened by an analysis of the substrate lexicon in various Indo-European languages of Europe – it would indeed be a surprise if none of that lexicon retained any terms for various crops that should have been widespread on the continent prior to the Indo-European arrival;
3. transformation of research on potential Afroasiatic agricultural terminology into a step-by-step process, where reliably reconstructible semantic fields in such families as Semitic, Berber, and Chadic (maybe even three different

subgroups of Chadic, considering how vast the family is) could be first compared directly to each other on binary levels: for instance, if the closest relative of Semitic is Berber, then, logically, the amount of common reliable Semito-Berber agricultural terms should be *higher*, not *lower*, than the corresponding amount for Afroasiatic in general. As of now, this does not seem to be the case, but this might also be due to the simple fact that nobody really tried applying such a binary approach in the first place;

4. analysis of attested agricultural lexicons in such families as Indo-European, Kartvelian, Dravidian, etc. (also not forgetting about such ancient Eurasian isolates as Sumerian or Elamite) on the subject of areal diffusion. For Indo-European, in particular, proposals have been made about the borrowing of at least a part of its agricultural lexicon from Semitic (Gamkrelidze & Ivanov 1995: 768–773) or from North Caucasian (Starostin 1988), which would agree with the general idea of speakers of these stocks triggering the spread of agriculture in the Near East.

Needless to say, none of these tasks can really be performed outside of the general historical-linguistic context – all of them are dependent on such basic things as systems of regular phonetic correspondences between compared languages; phylogenetic/topological analysis of the internal structure of the compared families; and degrees of typological plausibility of various semantic shifts in the cultural lexicon, to make it understandable, for instance, which types of cultural crops can easily “morph” into each other over time and which ones have no basis for being compared whatsoever. However, none of these problems seem to be theoretically insurmountable, as long as there exists a concise framework within which they may all be resolved. I can only hope that this brief overview may serve as a small contribution towards establishing such a framework for future research.

Abbreviations

NC	North Caucasian
PAA	Proto-Afroasiatic (Afrasian)
PIE	Proto-Indo-European
PU	Proto-Uralic
ST	Sino-Tibetan

References

- Agmon, Noam. 2010. Materials and language: Pre-Semitic root structure change concomitant with transition to agriculture. *Brill's Annual of Afroasiatic Languages and Linguistics* 2: 23–79. doi:10.1163/187666310X12688137960669
- Agmon, Noam & Bloch, Yigal. 2013. Statistics of language morphology change: From biconsonantal hunters to triconsonantal farmers. *PLOS One* 8 (12). doi:10.1371/journal.pone.0083780.
- Anthony, David W. 2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *Journal of Language Relationship* 9: 1–21.
- Bengtson, John D. 2008a. Materials for a comparative grammar of the Dene-Caucasian [Sino-Caucasian] languages. In *Aspects of Comparative Linguistics* 3, George Starostin (ed.), 45–118. Moscow: RSUH Publishers.
- Bengtson, John D. 2008b. A genetic comparison of Basque and Caucasian vocabulary. In *Linguistic Fossils: Studies in Historical Linguistics and Paleolinguistics*, John D. Bengtson (ed), 45–54. Calgary: Theophania Publishing.
- Blažek, Václav. 2012. Afroasiatic migrations: Linguistic evidence. Ms, paper presented at the conference Rethinking Africa's Transcontinental Continuities in Pre- and Protohistory, African Studies Centre, Leiden.
- Bomhard, Allan. 2014. *A Comprehensive Introduction to Nostratic Comparative Linguistics*. Charleston SC.
- Bouckaert, Remco *et al.* 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097): 957–960. doi:10.1126/science.1219669
- Bouda, Karl. 1936. Jenisseisch-tibetische Wortgleichungen. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 90(1): 149–159.
- Bouda, Karl. 1957. Die Sprache der Jenissejer Genealogische und morphologische Untersuchungen. *Anthropos* 52(1–2): 65–134.
- Dolgopolsky, Aharon. 1998. *The Nostratic Macrofamily and Linguistic Palaeontology*. Cambridge: The McDonald Institute for Archaeological Research.
- Dolgopolsky, Aharon. 2008. *Nostratic Dictionary*. Oxford: McDonald Institute for Archaeological Research.
- Dybo, Anna. 2000. Mir praaltajcev: Wörter und Sachen (The world of Proto-Altai speakers: Wörter und Sachen). In *Problemy izuchenija dal'nego rodstva jazykov na rubezhe tretjego tys'acheletija (Problems of the Study of Long Distance Relationships between Languages on the Threshold of the Third Millennium)*, George Starostin & Sergei Starostin (eds), 38–48. Moscow: RSUH Publishers.
- Ehret, Christopher. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian). Vowels, Tone, Consonants, and Vocabulary*. Berkeley CA: University of California Press.
- Ehret, Christopher. 2000. Language and history. In *African Languages: An Introduction*, Bernd Heine & Derek Nurse (eds), 272–297. Cambridge: CUP.
- Fähnrich, Heinz. 2007. *Kartwelisches Etymologisches Wörterbuch*. Leiden: Brill. doi:10.1163/ej.97890004161092.1-874
- Gamkrelidze, Thomas V. & Ivanov, Vjačeslav V. 1995. *Indo-European and the Indo-Europeans. A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture, Part I: The Text*. Berlin: Mouton de Gruyter.
- Gell-Mann, Murray, Peiros, Ilia & Starostin, George. 2009. Distant language relationship: The current perspective. *Journal of Language Relationship* 1: 13–30.

- Gray, Russell D. & Atkinson, Quentin D. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965): 435–439.
doi:10.1038/nature02029
- Hajdu, Peter. 1975. *Finnno-Ugrian Languages and Peoples*. London: André Deutsch.
- Harris, David R. 1989. An evolutionary continuum of people-plant interaction. In *Foraging and Farming: The Evolution of Plant Exploitation*, David R. Harris & Gordon C. Hillman (eds), 11–26. London: Unwin Hyman.
- Hayward, Richard J. 2000. Afroasiatic. In *African Languages: an Introduction*, Bernd Heine & Derek Nurse (eds), 74–98. Cambridge: CUP.
- Illich-Svitych, V. M. 1971–1984. *Opyt sravnenija nostraticheskix jazykov. Sravnitel'nyj slovar' (An Attempt at a Comparative Dictionary of the Nostratic Languages)*. Moscow: Nauka.
- Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? In *The Quasiquicentennial of the Finnno-Ugrian Society* [Mémoires de la Société Finnno-Ougrienne 258], Jussi Ylikoski (ed.), 57–78. Helsinki: Suomalais-Ugrilainen Seura.
- Kassian, Alexei, Zhivlov, Mikhail & Starostin, George. 2015. Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3–4): 301–347.
- Kortlandt, Frederik. 2010. *Studies in Germanic, Indo-European and Indo-Uralic*. Amsterdam: Rodopi.
- Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages*. Cambridge: CUP.
doi:10.1017/CBO9780511486876
- Kroonen, Gus. 2012. Non-Indo-European root nouns in Germanic: Evidence in support of the Agricultural Substrate Hypothesis. In *A Linguistic Map of Prehistoric Northern Europe* [Mémoires de la Société Finnno-Ougrienne 266], Riho Grünthal & Petri Kallio (eds), 239–260. Helsinki: Suomalais-Ugrilainen Seura.
- Larson, Greger *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* 111(17): 6139–6146.
doi:10.1073/pnas.1323964111
- Militarev, Alexander. 2000. Towards the chronology of Afrasian (Afroasiatic) and its daughter families. In *Time Depth in Historical Linguistics*, Vol. 1: Colin Renfrew, April McMahon & Larry Trask (eds), 267–307. Cambridge: McDonald Institute for Archaeological Research.
- Militarev, Alexander. 2002. The prehistory of a dispersal: The Proto-Afrasian (Afroasiatic) farming lexicon. In *Examining the Farming/Language Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 135–150. Cambridge: McDonald Institute for Archaeological Research.
- Pereltsvaig, Asya & Lewis, Martin. 2015. *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: CUP. doi:10.1017/CBO9781107294332
- Renfrew, Colin. 1987. *Archaeology and Language: the Puzzle of Indo-European Origins*. Cambridge: CUP.
- Robbeets, Martine Irma. 2005. *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Harrassowitz.
- Robbeets, Martine Irma. 2017. Proto-Transeurasian: Where and when? *Man in India* 97(1): 19–46.
- Sagart, Laurent. 2003. The vocabulary of cereal cultivation and the phylogeny of East Asian languages. *Bulletin of the Indo-Pacific Prehistory Association* 23(1): 127–136.
- Sagart, Laurent. 2005. Sino-Tibetan-Austronesian: An updated and improved argument. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 161–176. London: Routledge Curzon.
doi:10.4324/9780203343685_chapter_9

- Starostin, George. 2014. Macro-comparative linguistics in the 21st century: State of the art and perspectives. *Journal of Language Relationship* 11: 1–12.
- Starostin, Sergei. 1984. Gipoteza o geneticheskikh sv'az'akh sino-tibetskikh jazykov s jenisejskimi i severnokavkazskimi jazykami (A hypothesis about the genetic connections between Sino-Tibetan, Yeniseian, and North Caucasian languages). In *Lingvisticheskaja rekonstrukcija i drevnejshaja istorija Vostoka (Linguistic Reconstruction and the Prehistory of the East)*, Igor Vardul (ed.), 19–38. Moscow: Institute of Oriental Studies.
- Starostin, Sergei. 1988. Indoevropejsko-severnokavkazskije izoglossy (Indo-European-North Caucasian isoglosses). In *Drevnij Vostok: etnoku'lturnyje sv'azi (The Ancient Orient: Ethnocultural Ties)*, Grigory Bongard-Levin & Vladislav Ardzinba (eds), 112–163. Moscow: Nauka.
- Starostin, Sergei. 2004. The cultural vocabulary in the common North Caucasian lexical stock. In *Russian Oriental Studies*, Vitaly Naumkin (ed.), 211–245. Leiden: Brill.
- Starostin, Sergei. 2005. Sino-Caucasian [comparative phonology, glossary]. Ms, <<http://starling.rinet.ru/Texts/scc.pdf>>
- Starostin, Sergei, Dybo, Anna & Mudrak, Oleg. 2003. *Etymological Dictionary of the Altaic Languages*. Leiden: Brill.
- Starostin, Sergei & Nikolayev, Sergei. 1994. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk Publishers.
- Stolbova, Olga. 2016. *Etimologicheskij slovar' chadskih jazykov (Chadic Etymological Dictionary)*. Moscow: Institute of Oriental Studies.
- Theil, Rolf. 2006. *Is Omotic Afroasiatic? A Critical Discussion*. Ms, paper presented at the David Dwyer Retirement Symposium, Michigan State University, October.
- Trask, Larry R. 2008. *Etymological Dictionary of Basque*, Max W. Wheeler (ed.). Ms.
- Vajda, Edward. 2011. A Siberian link with Na-Dene languages. In *The Dene-Yeniseian Connection*, Jim Kari & Ben Potter (eds), 33–99. Fairbanks AK: ANLC.
- Vovin, Alexander. 2005. The end of the Altaic controversy. *Central Asiatic Journal* 49(1): 71–132.

Were the first Bantu speakers south of the rainforest farmers?

A first assessment of the linguistic evidence

Koen Bostoen and Joseph Koni Muluwa

UGent Centre for Bantu Studies (BantUGent)

Popular belief has it that the Bantu Expansion was a farming/language dispersal. However, there is neither conclusive archaeological nor linguistic evidence to substantiate this hypothesis, especially not for the initial spread in West-Central Africa. In this chapter we consider lexical reconstructions for both domesticated and wild plants in Proto-West-Coastal Bantu associated with the first Bantu speech communities south of the rainforest about 2500 years ago. The possibility to reconstruct terms for five different crops, i.e. pearl millet (*Pennisetum glaucum*), okra (*Hibiscus/Abelmoschus esculentus*), cowpea (*Vigna unguiculata*), Bambara groundnut (*Vigna subterranea*) and plantain (*Musa spp.*), indicates that by that time Bantu speakers did know how to cultivate plants. At the same time, they still strongly depended on the plant resources that could be collected in their natural environment, as is evidenced by a preliminary assessment of reconstructible names for wild plants. Agriculture in Central Africa was indeed “a slow revolution”, as the late Jan Vansina once proposed, and certainly not the principal motor behind the early Bantu Expansion.

Keywords: Bantu Expansion, West-Coastal Bantu, agriculture, foraging, hunter-gatherers, lexical reconstruction, plant names

1. Introduction

The Bantu Expansion is no doubt the most important linguistic, cultural and demographic process in Late Holocene Africa. It has sparked intense debate across disciplines and far beyond Africanist circles. Several generations of linguists, archaeologists, anthropologists, geneticists and many more have debated on how the Bantu language family, which is not older than 5000 years, could spread over such disproportionately large parts of Central, Eastern and Southern Africa; see

Figure 1. As it often happens with hotly debated issues, certain widely held beliefs threaten to become “factoids”, because they are no longer critically questioned and start to lead a life of their own that bears little relation to any factual reality. One of the commonest conjectures about the Bantu Expansion certainly is that it would have been a farming/language dispersal with agriculture as the principal motor behind large-scale language spread (e.g. Bellwood & Renfrew 2002; Diamond & Bellwood 2003; Phillipson 2003). Both phenomena are so strongly tied up in the minds of certain scholars that they simply consider Bantu language phylogenies (e.g. Holden 2002) or archaeology-based phylogeographies (e.g. Russell et al. 2014) as mirroring the spread of farming without even discussing the slightest evidence for food production. The equation between Bantu and agriculture is also taken for granted by most geneticists who consistently adopt a dichotomy between “Bantu (speaking) farmers” and autochthonous foragers, i.e. the “Pygmies” in Central Africa and “(Khoi)San” in Southern Africa (e.g. Destro-Biso et al. 2004; Quintana-Murci et al. 2008; de Filippo et al. 2010; Barbieri et al. 2014; Patin et al. 2014). However, as we have extensively argued elsewhere (Kahlheber et al. 2009; Neumann et al. 2012a; Bostoen et al. 2013a; Bostoen 2014; Bostoen et al. 2015), both direct archaeological evidence and indirect linguistic evidence concur to question the plausibility of agriculture as the main driving force behind the Bantu Expansion, especially as far as its initial phases are concerned.

On the other hand, it is increasingly recognized that the Bantu Expansion was facilitated and even accelerated through climate-induced openings of the Central African rainforest block (Brncic et al. 2009; Ngomanda et al. 2009; Maley et al. 2012; Neumann et al. 2012b; Hubau et al. 2015), rather than that migrating Bantu speech communities themselves would have caused deforestation (Bayon et al. 2012). Schwartz (1992) was the first to link the dispersal of Bantu languages with climate change around 3000 BP. We have deepened and revised this hypothesis through an extensive review of evidence from biogeography, palynology, geology, historical linguistics, and archaeology that led to a new interdisciplinary reconstruction of the palaeoclimatic context in which the early Bantu Expansion took place (Bostoen et al. 2015). Palaeoenvironmental data indicate that a climate crisis affected the equatorial rainforest during the Holocene, first its periphery around 4000 BP and later its core around 2500 BP. Both phases had an impact on the Bantu Expansion, but in different ways. The climate-induced extension of savannas at the periphery of the rainforest, for instance in the Sanaga-Mbam confluence area in central Cameroon, around 4000–3500 BP probably facilitated the settlement of early Bantu-speech communities in the region of Yaoundé in present-day Cameroon and later along the coast of Equatorial Guinea and Gabon and inland along the Ogooué River, but did not lead to a large-scale geographic expansion of Bantu-speaking settlements in Central Africa. It was only when the core of the

Central African rainforest was affected around 2500 BP that such a rapid eastward and southward expansion occurred. The rapidness of this initial migration through the forest is also indicated by genetic data suggesting that most admixture between various groups of hunter-gatherers and neighboring communities took place within the past 1000 years (Patin et al. 2014). Contacts seem to have intensified only once Bantu speech communities were firmly settled in the rainforest. Metallurgy and domesticated plants from the savannah, such as pearl millet, also spread through Central Africa around 2500 BP (Kahlheber et al. 2009; Clist 2012; Neumann et al. 2012a; Kahlheber et al. 2014) to become part of the cultural package which Bantu speakers took further East and South.

Using a dated phylogeny of more than 400 Bantu languages calibrated through archaeological dates and combined with contemporary geographical information and appropriate statistical modelling, Grollemund et al. (2015) try to demonstrate that early Bantu-speaking populations did indeed not expand from their ancestral homeland in a “random walk” but, rather, that they followed emerging savannah corridors, with rainforest habitats repeatedly imposing temporal barriers to movement. The Sangha River Interval, in particular, may have been a crucial passageway for the start of the gradual colonization of the Inner Congo Basin by Bantu speakers as well as for their initial north-south migration across the Equator (Bostoen et al. 2015; Grollemund et al. 2015). It is precisely that last movement which would have led to the introduction of the Bantu language ancestral to the present-day “West-Western” clade (Grollemund et al. 2015), aka “West-Coastal” (Vansina 1995), into the area North of the Malebo Pool on the Congo River. The homeland of this major Bantu clade, on which the current chapter focuses, has been tentatively situated between the Bateke Plateau, a huge highland straddling three countries (Gabon and both Congo), and the Bandundu region (Democratic Republic of the Congo), i.e. around 3°S and between about 14°E and 17°E; see Figure 1. These ancestral “West-Western” or “West-Coastal” Bantu speakers were the first Bantu speakers south of the forest.

In this chapter, we review subsistence-related vocabulary that can be reconstructed in Proto-West-Coastal Bantu in order to get a better understanding of the subsistence economy of the first Bantu speakers south of the rainforest and to make a first assessment of whether they had become farmers by that time. We will exclusively focus here on plant vocabulary by relying mainly on the comparative word lists that were included in the PhD dissertation of the second author (Koni Muluwa 2010). The fieldwork data from the Nsong, Ngong, Mpiin, Mbuun and Hungan languages, all spoken in the Kwilu Province (Democratic Republic of the Congo), were subsequently published in Koni Muluwa (2014). More comparative cultural vocabulary from languages spoken in that area was included in Koni Muluwa and Bostoen (2015).

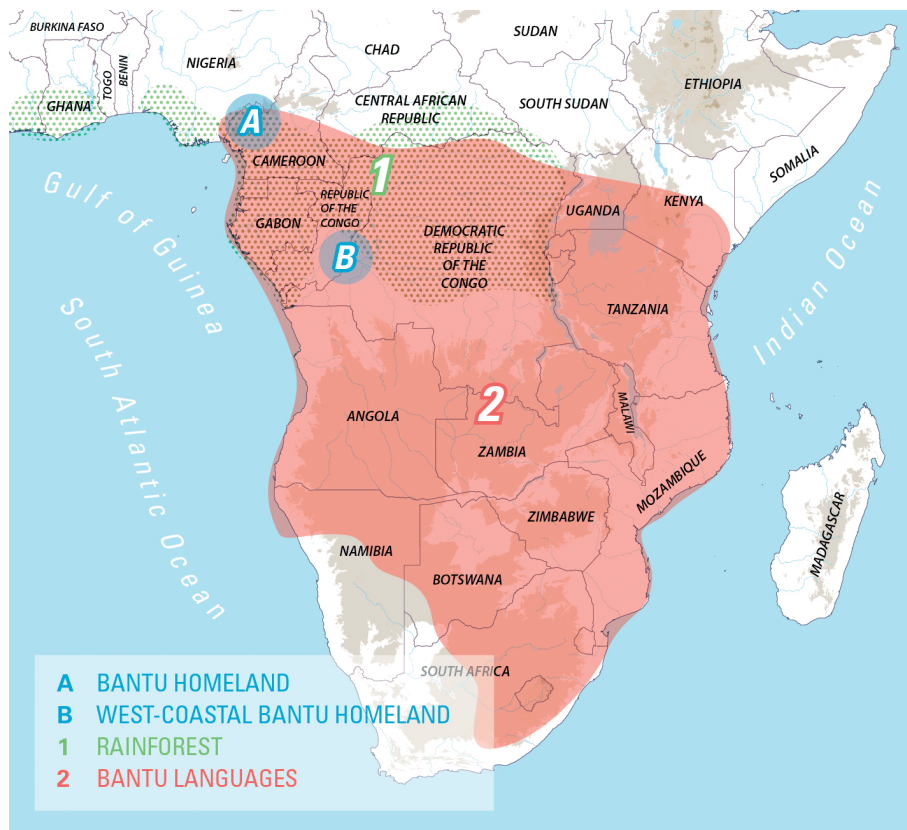


Figure 1. Approximate distribution of the Bantu languages and location of the Bantu and West-Coastal Bantu homelands

Of the five Kwilu Bantu languages mentioned above, Hungan is the only one to belong to the so-called “Kikongo Language Cluster” (KLC), which is the main sub-branch of West-Coastal Bantu in terms of the number of languages and their distribution. The Kikongo Language Cluster spread from the inland homeland south of the rainforest towards the Atlantic Coast and covers today major parts of southern Gabon, the southern Republic of the Congo, the southwestern Democratic Republic of the Congo and northern Angola including Cabinda. Within the Kikongo Language Cluster, Hungan belongs to the “Kikongoid” sub-clade, the first to split off from the common core (de Schryver et al. 2015). Nsong, Ngong, Mpiin and Mbuun, from their side, are part of the Yanzi group, a second sub-branch of West-Coastal Bantu, which springs from an ancestor language that moved east of the Congo River somewhere in between the Kwango and Kwilu Rivers in the Bandundu region of the Democratic Republic of the Congo. The third sub-branch

of West-Coastal Bantu consists of the Nzebi-Mbete-Teke languages, which are still spoken today in the vicinity of the Bateke plateau, close to the West-Coastal Bantu homeland. Plant vocabulary attested in each of these three sub-branches will be considered here for reconstruction in Proto-West-Coastal Bantu.

In Section 2, we review the evidence available for the assumption that the Bantu Expansion would have been a language/farming dispersal. In Section 3, we assess the crop plant vocabulary that can be reconstructed in Proto-West-Coastal Bantu. In Section 4 we consider Proto-West-Coastal Bantu wild plant vocabulary. Conclusions are presented in Section 5.

2. Reviewing the evidence for the Bantu Expansion as a language/farming dispersal

Direct archaeological evidence for food production and domestication in Central Africa is still very scarce, substantially younger than the assumed start of the Bantu Expansion, i.e. some 4000 to 5000 years ago (Vansina 1995; Blench 2006; Bostoen 2007), and discovered far from the Bantu homeland, which is situated in the Nigerian-Cameroonian borderland (Greenberg 1972); see Figure 1. Domesticated pearl millet (*Pennisetum glaucum*) was found in three sites from southern Cameroon, all dated between 2350–2200 BP, and in one site in the Democratic Republic of the Congo on the Lulonga River dated around 2200 BP (Eggert et al. 2006; Kahlheber et al. 2009; Kahlheber et al. 2014). In another South-Cameroonian site, remains of the pulse species Bambara groundnut (*Vigna subterranea*) dated around 1750 BP were found (Eggert et al. 2006). Both crop species originate from more northerly savannah regions and are adapted to drier environmental conditions. They do not belong to the crop inventory of current-day Central African rainforest agriculture which is mainly based on *Musa* species (plantain) and several tuber plants like cassava (*Manihot esculenta*), taro (*Colocasia esculenta*), tannia (*Xanthosoma sagittifolium*, *Xanthosoma poeppigii*), sweet potato (*Ipomoea batatas*) and yams (*Dioscorea spp.*) as the principal providers of carbohydrate, whereas the cultivation of maize (*Zea mays*) and Asian rice (*Oryza sativa*) is only occasional. Only certain yams are indigenous to Africa, but the role of these tubers in past subsistence economies is difficult to assess archaeologically, since yam starch does not leave easily detectable traces in Africa (Neumann 2005:262). The only early evidence available for forest crops are banana phytoliths from Cameroon dated between 2750 and 2350 BP (Mbida Mindzié et al. 2000) and from Uganda dated to the 6th millennium BP (Lejju et al. 2005; Lejju et al. 2006). Such early dates for a domesticated plant of Southeast Asian origin has caused a great deal of controversy (Vansina 2003; Mbida Mindzié et al. 2005; Neumann & Hildebrand 2009). They call

for corroborating evidence from other Central African sites, which has not been found so far, among other things because fieldwork in Central Africa specifically targeting archaeobotanical remains is recent and not yet standard. The role of animal domestication in early Bantu-speaking societies is also difficult to assess due to the poor preservation of bones, particularly in open-air sites. The little evidence available suggests the presence of small livestock in Central Africa by the mid-third millennium BP, but at the same time the minor importance of domesticated animals in the earliest phases of the Bantu Expansion (Van Neer 2000). As things stand today, the Late Holocene archaeology of Central Africa provides no convincing evidence for farming as the principal driving force behind the Bantu Expansion.

Calling the earliest Bantu speakers “farmers” is also unjustifiable from a linguistic viewpoint. The only crops for which vocabulary can be reconstructed in Proto-Bantu are yams and possibly two *Vigna* species, i.e. the cowpea (*Vigna unguiculata*) and the Bambara groundnut (*Vigna subterranea*) (Philippon & Bahuchet 1994–1995; Bostoen 2014). The high number of lexical reconstructions for yams suggests that different *Dioscorea* species were indeed on the menu (Maniacky 2005). They were no doubt the main starch ingredient with which early Bantu speakers prepared their staple porridge as a mash (Ricquier & Bostoen 2011). Moreover, all Proto-Bantu yam terms were inherited from an older language stage, strongly suggesting that yams were already part of the diet before the ancestors of Bantu speakers reached the Bantu homeland in the Nigerian-Cameroonian borderland (Maniacky 2005; Blench 2006). However, since the wild ancestors of domesticated African yams also occur in the rainforest, these lexical reconstructions cannot be taken as evidence for plant cultivation, even not indirectly. The reconstruction of Proto-Bantu *Vigna* vocabulary could be in line with the archaeological evidence discussed above except for the chronology since the first and only archaeobotanical attestation of the Bambara groundnut (*Vigna subterranea*) is less than 2000 years old. An in-depth study is needed to corroborate whether the words reconstructed for these pulse species really referred from the very start to these domesticates exogenous to Central Africa and can indeed be seen as indirect evidence for food production. It should be excluded that they did not originally designate local wild plants and only became vernacular *Vigna* names through semantic shift as commonly happened for crops imported in Africa (Pasch 1979). Vocabulary for pearl millet and bananas cannot be regularly reconstructed to Proto-Bantu, but only appears in more recent ancestral language stages which suggests that Bantu speakers only integrated them in their culinary traditions in the course of their expansion (Bostoen 2006–2007; Blench 2009). However, they did already exploit fruit-bearing trees before leaving their homeland, and quite extensively to judge from the number of reconstructions, which would even increase provided that more dedicated historical linguistic research was done. Proto-Bantu vocabulary includes names

for several wild species, which have been widely protected and cultivated in equatorial Central Africa, but have never become domesticates, such as the oil palm (*Elaeis guineensis*), the bush-candle (*Canarium schweinfurthii*), the African plum (*Dacryodes edulis*), and the umbrella tree (*Musanga cecropioides*) (Bostoen et al. 2013a; Bostoen 2014).

The early economic importance of the oil palm and the bush-candle is well attested in the archaeological record of Western and Central Africa, where the remains of both oleaginous plants have often been found from ca. 5000 BP onwards in association with other indicators of plant food-processing, such as pounding/grinding equipment, polished stone tools and pottery (de Maret 1994–1995; D’Andrea et al. 2006). Other nuts have been found in archaeological deposits around 2000 BP in Cameroon, Equatorial Guinea and Gabon, like *Antrocaryon micraster*, *Chytranthus macrobotrys*, *Coula edulis* (African walnut), *Panda oleosa* (Clist 2005). Recently, scholars working in the Democratic Republic of the Congo succeeded for the first time to recover *Musanga cecropioides* diaspores (Kahlheber et al. 2014) and charred wood remains (Hubau et al. 2014) from archaeological deposits.

In sum, both the archaeological and linguistic evidence currently available urge us to seriously question the widely held belief that the Bantu Expansion is a textbook case of a farming/language dispersal. Both bodies of evidence rather suggest that the earliest Bantu speakers chiefly relied on non-domesticated foods and had a lifestyle that was situated towards the foragers’ side of the “middle ground”, i.e. “the large transitional zone in the continuum between hunter-gatherers on the one hand and agriculturalists largely depending on domesticated crops on the other (...)” (Neumann 2005: 249).

3. Crop vocabulary in Proto-West-Coastal Bantu

We tentatively propose five crop names for reconstruction in Proto-West-Coastal Bantu:

- *-cángú ‘pearl millet’ (*Pennisetum glaucum*)
- *-kòndò ‘plantain’ (*Musa spp.*)
- *-gómbo ‘okra’ (*Hibiscus/Abelmoschus esculentus*)
- *-kóndè ‘cowpea’ (*Vigna unguiculata*)
- *-jùgú ‘Bambara groundnut’ (*Vigna subterranea*)

As we have extensively demonstrated elsewhere (Bostoen 2006–2007; Kahlheber et al. 2009), the noun stem *-cángú can be reconstructed to Proto-Bantu, where it referred to grains of some kind, though not specifically to pearl millet (*Pennisetum glaucum*). It only became associated with this particular domesticated of West

African origin after Bantu speakers had started to emigrate southwards from their homeland and their ancestral language had started to diverge into distinct sub-branches. Regular reflexes designating this cereal are attested in present-day Bantu languages belonging to the South-Western Bantu, Central-Western Bantu and West-Coastal Bantu branches, which all split off after the Bantu languages started their rapid dispersal through the rainforest. The late semantic shift or narrowing towards ‘pearl millet’ is well in line with the currently available archaeobotanical evidence indicating that this cereal only appeared in Central Africa after 2500 BP once the core of the rainforest underwent a climate-induced crisis associated with a more accentuated seasonality, which is needed for the cultivation of pearl millet. With current-day reflexes in all three West-Coastal sub-branches, **-cángó* ‘pearl millet’ can safely be reconstructed into Proto-West-Coastal Bantu. Today, however, reflexes of **-cángó* more commonly refer to maize (*Zea mays*), which West-Coastal Bantu speech communities acquired as part of the Columbian exchange and whose cultivation is nowadays more widespread than that of pearl millet. The lexical reconstructions **-kú* ‘millet; eleusine’ and **-pòndó* ‘millet’, proposed by Bastin et al. (2002), reflect other innovations in the cereal cultivating traditions of West-Coastal Bantu speakers. The two terms seem innovations that are posterior to Proto-West-Coastal Bantu, but more dedicated study is needed to establish both the time depth of their introduction and the specific cereal species to which they initially referred.

Apart from recent archaeobotanical finds of pearl millet (Eggert et al. 2006; Kahlheber et al. 2014), other evidence for early plant cultivation in western Central Africa comes from the identification of banana phytoliths by Mbida Mindzié et al. (2000). According to Blench (2009:363), plantains arrived in West Africa earlier than 3000 BP along with taro and water yam and the cultivation of these crops made possible the effective exploitation of the dense equatorial rainforest. He identifies one widespread term for plantain, which also occurs across the zone where the greatest degree of somatic variation is found, i.e. the northeastern Democratic Republic of the Congo (DRC). However, the “most prominent reconstructible” form **-ko[n]do* which he proposes is not a true reconstruction. It rather reflects the phonological irregularity, which this term manifests across languages, suggesting that its initial diffusion was contact-induced and not the consequence of language spread and divergence. This is well in line with the conclusion of Philippon and Bahuchet (1994–1995) that reconstructing a regularly inherited term for plantain or banana to Proto-Bantu is not possible. On the other hand, West-Coastal Bantu languages do share a cognate term that seems to be regularly inherited from their most recent common ancestor and corresponds to the reconstruction **-kòndò* ‘banana: Musaceae’ (Bastin et al. 2002). It is widely attested in languages of all three West-Coastal Bantu sub-branches and respects regular sound correspondences between them, as some examples in Table 1 illustrate. The final nasal-consonant

cluster reduction and the apocope of the final syllable observed in the West-Coastal Bantu languages not belonging to the Kikongo Language Cluster is a sound shift regularly shared amongst them (Daeleman 1977; Hombert 1986; Koni Muluwa 2010). A systematic comparison of all available attestations is needed to establish a firm Proto-West-Coastal Bantu reconstruction, but the lexical evidence in Table 1 suggests that by the time the first Bantu speakers reached south of the rainforest, bananas of some kind had become regular part of their diet. Along with their languages, West-Coastal Bantu speakers further spread them towards the Atlantic Coast in the west and the Bandundu in the east.

Table 1. Reflexes of *-kòndò ‘banana’ in present-day West-Coastal Bantu languages

Sub-branch	Language	Country	Term	Source
KLC	Ntandu	DRC	<i>dinkóndo</i>	(Daeleman & Pauwels 1983:203)
	Suku	DRC	<i>dinkondu</i>	(Bunkheti 1997: 114)
Yanzi	Nsong	DRC	<i>ékɔ́:n</i>	(Koni Muluwa 2014: 70)
	Nzadi	DRC	<i>ikwɔ</i>	(Crane et al. 2011:283)
Nzebi-Mbete-Teke	Nzebi	Gabon	<i>ləkɔ</i>	(Blanchon & de Nadaillac 1987:65)
	Teke	Gabon	<i>kó</i>	(Fontaney 1984:57)

Blench (2006: 121) rightfully observes that no Proto-Bantu reconstructions are available for ancient African domesticates, such as okra (*Hibiscus/Abelmoschus esculentus*), roselle (*Hibiscus sabdariffa*) and amaranth (*Amaranthus sp.*). Such does not seem to be entirely the case for Proto-West-Coastal Bantu.

As for okra, a crop whose center of domestication is still uncertain but definitely outside the Bantu area (Hamon & Charrier 1997: 322–323), a cognate term reconstructible as *-gómbo is widespread in two sub-branches of West-Coastal Bantu, i.e. the Kikongo Language Cluster and the Yanzi subgroup (see Table 2). The tone pattern of the reflex in Ntandu, whose correspondences with tone in Bantu lexical reconstructions are best known (Daeleman 1983), does not allow to discriminate between *HH and *HL. For the time being, no reflex could be identified in the Nzebi-Mbete-Teke sub-branch. Boma, for instance, has *lɔnalɔ́:n* (Koni Muluwa and Bostoen 2015: 102), which seems to have several cognates among languages of the Yanzi sub-group, e.g. Nzadi *dɔ́yɔ́n* (Crane et al. 2011: 292). See also Koni Muluwa and Bostoen (2015: 102) for Yans, Mpur, Lwel and Ngwi. Outside West-Coastal Bantu, it occurs in Lingala, for instance: *dɔngɔ́dɔngɔ́* (van Everbroecke 1985). The *-gómbo term for ‘okra’ is also attested outside West-Coastal Bantu, i.e. mainly in South-Western Bantu languages, e.g. Kimbundu *kingombo* (Gossweiler 1953: 39) and Lucazi *cingombo* (Storrs 1995). This makes it a likely candidate for reconstruction in Proto-West-Coastal Bantu. It is also this specific term which made it to the other side of the Atlantic as part of the Columbian exchange. In Creole culinary

culture, *gumbo* has become a signature dish consisting of stew made of okra and bits of meat and poultry or shellfish, served as a soup or with rice (McCann 2009: 171). It is important to stress that the **gómbo* reconstruction for okra has nothing to do with the *tukuru* form, which Blench (1994–1995) proposes as going back as far as Proto-Benue-Congo, the proto-language ancestral to Proto-Bantu itself. This great time depth is likely to be exaggerated and in need of serious reconsideration.

Table 2. Reflexes of **-gómbo* ‘okra’ in present-day West-Coastal Bantu languages

Sub-branch	Language	Country	Term	Source
KLC	Ntandu	DRC	<i>góombo</i>	(Daeleman & Pauwels 1983: 196)
	Samba	DRC	<i>kingómbu</i>	(Koni Muluwa & Bostoen 2015: 102)
Yanzi	Ngong	DRC	<i>kéngómb</i>	(Koni Muluwa 2014: 37)
	Mbuun	DRC	<i>íngomb</i>	(Koni Muluwa 2014: 37)

As for roselle (*Hibiscus sabdariffa*) and amaranth (*Amaranthus* sp.), West-Coastal Bantu languages do share some terms that seem to have a certain time-depth, but for the time being none of them is eligible for a solid reconstruction in Proto-West-Coastal Bantu.

The term referring to amaranth which several languages spoken in the Kwilu Province (Democratic Republic of the Congo) share is reminiscent of the regional reconstruction **dèngàdèngà* proposed by Bastin et al. (2002) on the basis of data from eastern Bantu languages: Mpiin *mulín*, Nsong *ɔlɛŋ*, Mbuun *ɔlɛŋ*, Ngong *mólé*, Hungan *muléŋ* (Koni Muluwa 2010: 479; 2014: 39). Similar words occur in South-Western Bantu languages, such as Cokwe and Kanyok, i.e. respectively *mulenje* (Gossweiler 1953: 392) and *múlé:ŋ* (Kabinda 1988). However, for now, no other attestations were found elsewhere in West-Coastal Bantu, among other things because the vocabulary concerned is not well documented. It is therefore hard to say whether the amaranth terms attested in Mpiin, Nsong, Mbuun, Ngong and Hungan are retentions from Proto-West-Coastal Bantu or rather the outcome of contact with South-Western Bantu languages spoken in the neighborhood. More dedicated data collection and language comparison is needed here.

With regard to roselle, the Yanzi languages from the Kikwit area also share a term that seems to be attested outside West-Coastal Bantu but nowhere else inside. Nsong and Ngong have *bɔkwɛs*, Mpiin *bukwés* and Mbuun *ɔkwɛs* (Koni Muluwa 2010: 487; 2014: 62). Possible cognates are attested in the South-Western Bantu languages Kimbundu and Cokwe, i.e. respectively *use* and *kise* (Gossweiler 1953: 156). However, West-Coastal Bantu languages of the Kikongo Language Cluster and the Nzebi-Mbete-Teke subgroup have a *-kulu* stem for this plant, e.g. Bembe *kinkulú* (Kouarata 2016: 81), Punu *ábukúlu* (Blanchon 1991: 57), Vili

búk^húlú (Ndinga-Koumba-Binza 2000), Latege *lánkùlú* (Linton 2016: 21), Iyaa *ikùlú* (Mouandza 1991: 103). More research is needed to establish whether this stem is eligible for reconstruction in Proto-West-Coastal Bantu or whether it is a later innovation excluding the Yanzi languages from the Bandundu.

Finally, it is worth noting that the reflexes of the lexical reconstructions proposed for the pulses *Vigna unguiculata* (cowpea) and *Vigna subterranea* (Bambara groundnut), i.e. **-kòndè* and **-jògú* respectively (Philippson & Bahuchet 1994–1995), occur only marginally in West-Coastal Bantu. Koni Muluwa (2010: 313; 2014: 85) reports *ékú:nd* in Nsong, where it designates both *Vigna unguiculata* and *Phaseolus vulgaris* or common bean, the latter being imported through the Columbian exchange. Several other languages of the Yanzi subgroup designate the common bean with a cognate form (Koni Muluwa & Bostoen 2015: 106). However, in the Yanzi subgroup and the Kikongo Language Cluster, cognate forms of Ntandu *nkása* (Daeleman & Pauwels 1983: 212) are prevalent for both *Vigna unguiculata* and *Phaseolus vulgaris* (Koni Muluwa & Bostoen 2015: 106). This *-kasa* stem appears to be an innovation posterior to Proto-West-Coastal Bantu, along with *-deeso*, which also refers to *Phaseolus vulgaris* and is especially pervasive within the Kikongo Language Cluster, but equally occurs elsewhere inside and outside West-Coastal Bantu (Koni Muluwa & Bostoen 2015: 106; Ricquier 2016: 118). A similar widespread innovation, i.e. *-guba*, exists for both *Vigna subterranea* (Bambara groundnut) and *Arachis hypogaea* (peanut), the latter also being an import of American origin. It is particularly prevalent within the Kikongo Language Cluster (Ricquier 2016: 138), while the more archaic stem **-jògú* has been maintained in the other West-Coastal Bantu branches (Koni Muluwa & Bostoen 2015: 55). Although it mainly refers to the peanut in present-day languages, it is also still associated in some of them with the Bambara groundnut, which is nowadays less commonly cultivated. The Ngong people from the Kwilu area, for example, call it *lodzú la ngó*, because they consider it to be their signature crop (Koni Muluwa 2014: 86). That is why we would tentatively propose – in anticipation of more in-depth analysis – **-jògú* as a Proto-West-Coastal Bantu reconstruction for *Vigna subterranea* along with **-kòndè* for *Vigna unguiculata*.

In sum, the comparative lexical data considered above allow for the tentative reconstruction of Proto-West-Coastal Bantu terms for at least five crops, i.e. pearl millet (*Pennisetum glaucum*), okra (*Hibiscus/Abelmoschus esculentus*), cowpea (*Vigna unguiculata*), Bambara groundnut (*Vigna subterranea*) and plantain (*Musa spp.*). All are crops whose center of domestication is situated beyond the Bantu distribution area. In other words, if the first Bantu speakers south of the rainforest had vocabulary for these crops, they probably knew how to cultivate plants in their West-Coastal Bantu homeland. In this regard, the lexical evidence available for the reliance on domesticated crops is definitely more conclusive at the stage

of Proto-West-Coastal Bantu than at the earlier stage of Proto-Bantu, even if the number of such crops in their diet was still fairly limited. Moreover, given that most of these crop names do not seem to be West-Coastal Bantu innovations, but terms also attested in other major Bantu branches, especially in South-Western and Central-Western Bantu, it is quite likely that ancestral West-Coastal Bantu speakers had integrated the cultivation of these crops in their subsistence strategies before they arrived in their homeland south of the rainforest.

4. Wild plant vocabulary in Proto-West-Coastal Bantu

The possibility to reconstruct at least five crop names in Proto-West-Coastal Bantu is an important progress with regard to Proto-Bantu. However, this number is still fairly low, especially if compared with the number of wild plant names reconstructible in Proto-West-Coastal Bantu. On the basis of our preliminary comparative research, we could propose not less than 42 tentative Proto-West-Coastal Bantu reconstructions referring to different kinds of wild trees, shrubs and other plants occurring in different types of habitats. This number does not include those for (wild) yams and for wild trees, such as oil palm (*Elaeis guineensis*), bush-candle (*Canarium schweinfurthii*), African plum (*Dacryodes edulis*), umbrella tree (*Musanga cecropioides*) and cola nut tree (*Cola sp.*), which were reconstructed earlier on for Proto-Bantu (Maniacky 2005; Bostoen et al. 2013a; Bostoen 2014) and several of which were retained in Proto-West-Coastal Bantu. It would go beyond the scope and the page constraints of this chapter to present all 42 new lexical reconstructions. We refrain ourselves to some case studies which are illustrative of the natural environment in which Proto-West-Coastal Bantu speakers lived, of the different purposes for which they relied on wild plants and of the different ancestral stages in which these plant names were acquired.

Firstly, a series of Proto-West-Coastal Bantu plant names are actually retentions from Proto-Bantu. It concerns series of cognate terms that are attested in those Bantu branches which split off first, such as Mbam-Bubi and/or North-Western Bantu (Grollemund et al. 2015), as well as in several later major branches, such as Central-Western Bantu, West-Coastal Bantu, South-Western Bantu and/or East Bantu. Some of these lexical reconstructions already figure in Bastin et al. (2002), but were not yet solidly reconstructed into Proto-Bantu; others were never proposed before. One of the latter kind is a term referring to the kapok tree or *Ceiba pentandra* (Malvaceae). This tree is, just like *Elaeis guineensis*, *Canarium schweinfurthii* and *Musanga cecropioides* (Bostoen et al. 2013a), a pioneer species that naturally colonizes clearings in the tropical forest zone. In Central African societies, this tree traditionally is multifunctional: the wood is used for carvings, coffins and

dugout canoes, the fibres for bedding and life preservers, the oil in the seeds for soap, the bark as a purgative and to cause vomiting in the event of poisoning and the leaves for different kinds of medical treatment, such as for haemorrhoids, asthenia, heartburn, etc. (Raponda-Walker & Sillans 1995; Latham 2004; Koni Muluwa 2014). As shown in Table 3, cognate forms for this tree occur in languages belonging to North-Western, Central-Western, South-Western and West-Coastal Bantu. We tentatively propose the reconstruction **-kùmà* for this comparative series. The reconstructed LL tone pattern is based on the tones of the Mongo reflex, which should be morphologically analysed as *b(o)-uma*. It is well known that Proto-Bantu **k* has become \emptyset in Mongo and the language directly reflects Proto-Bantu tones (Hulstaert 1941; de Rop 1953, 1958). Being represented in all three West-Coastal Bantu sub-branches, this term can also be reconstructed to their most recent common ancestor as a retention from Proto-Bantu.

Table 3. Reflexes of **-kùmà* ‘kapok tree’ in Bantu languages belonging to distinct major branches

Branch	Language	Country	Term	Source
NW	Mpiemo	Cameroun	<i>dumɔ</i>	(Thornell 2004: 66)
	Tsogo	Gabon	<i>ogumà</i>	(Raponda-Walker & Sillans 1995: 106)
CW	Mongo	DRC	<i>buma</i>	(Hulstaert 1957: 455)
	Turumbu	DRC	<i>lihuma</i>	(SPIAF 1988: 8)
SW	Kimbundu	Angola	<i>mufuma</i>	(Gossweiler 1953: 154)
	Cokwe	Angola	<i>kafuma-fuma</i>	(Gossweiler 1953: 154)
WCB	Mbede	Gabon	<i>okuma</i>	(Raponda-Walker & Sillans 1995: 106)
	Nsong	DRC	<i>ópfum</i>	(Koni Muluwa 2014: 47)
	Hungan	DRC	<i>múpfum</i>	(Koni Muluwa 2014: 47)

Secondly, a series of Proto-West-Coastal Bantu plant names seem to be retentions from an ancestral stage posterior to Proto-Bantu. They are attested in several Bantu branches other than West-Coastal Bantu, but are not sufficiently widespread to be reconstructed into Proto-Bantu, especially because they are absent from the branches that split off first, i.e. Mbam-Bubi and North-Western Bantu. Several names of useful plants are shared between East-Bantu and all western Bantu branches except Mbam-Bubi and North-Western Bantu. This is in line with the claim that East-Bantu is a late offshoot that emerged from western Bantu (Grollemund et al. 2015). Two reconstructions already proposed by Bastin et al. (2002) on the basis of reflexes from these four branches fit into this category, i.e. **-dódò* ‘*Annona senegalensis*’ and **-pòmí* ‘*Erythrophleum suaveolens*’.

The first one, also known as “African custard-apple” is a common savannah species whose fruits are edible. The young leaves and roots are used to treat, among

other things, constipation, gastritis, diabetes, painful joints, anaemia and epilepsy, and the gum is applied to cuts and wounds to seal them. It also hosts edible caterpillars (Latham 2004; Koni Muluwa 2014). Table 4 presents reflexes of **-dódò* in a series of Bantu languages belonging to different major branches. It should be noted that it does not always refer to *Annona senegalensis* itself in present-day West-Coastal and other Bantu languages, but sometimes to closely related species, such as the *Annona stenophylla* and the *Annona arenaria*. As a consequence, it is safer to associate the value ‘*Annona sp.*’ with the Proto-West-Coastal Bantu reconstruction **-dódò*. Moreover, in several northern languages of the Kikongo Language Cluster, the term was also adopted to designate the papaya, a fruit of American origin, at the time of its introduction as part of the Columbian exchange (Ricquier 2016: 130).

Table 4. Reflexes of **-dódò* ‘*Annona sp.*’ in Bantu languages belonging to distinct major branches

Branch	Language	Country	Term	Source
E	Shona	Zimbabwe	<i>muroro</i>	(Hannan 1974: 936)
	Fwe	Zambia	<i>muroro</i>	(Bingham 2005)
CW	Tetela	DRC	<i>ɔlɔ́</i>	(Hagendorens 1975: 328)
SW	Kimbundu	Angola	<i>dilolo</i>	(Gossweiler 1953: 137)
	Cokwe	Angola	<i>mulolo</i>	(Gossweiler 1953: 137)
	Kwamashi	Zambia	<i>diróró</i>	(Bostoën fieldwork 2007)
WCB	Ntandu	DRC	<i>kilólo; ñlóló</i>	(Daeleman & Pauwels 1983: 168)
	Mpiin	DRC	<i>múlbl</i>	(Koni Muluwa 2014: 40)

The second one (*Erythrophleum guineense*) is also known as the “ordeal tree”, because it produces a poison that is used for ordeals throughout Central Africa. This is a widespread and ancient ritual tradition among western Bantu speech communities (Vansina 1990: 300; MacGaffey 1991: 9). As Vansina (1990: 300) notes, apart from **-pòmí*, of which he observed reflexes in West-Coastal, Central-Western, South-Western and Eastern Bantu languages, a second term tentatively reconstructed as **-kaca* is widespread among western Bantu languages, especially in West-Coastal and Central-Western Bantu languages. Table 5 presents reflexes of both roots in West-Coastal Bantu languages. The **-pòmí* stem seems to prevail in the Yanzi subgroup, while the **-kaca* stem is predominant in the two other West-Coastal Bantu sub-branches. Relying on their attestations outside West-Coastal Bantu, both stems appear to be reconstructible to the most recent common ancestor of the Kikongo Language Cluster, Nzebi-Mbete-Teke and Yanzi subgroups. Remarkably, several other species, such as *Elaeis guineensis*, *Canarium schweinfurthii* and *Musanga cecropioides*, similarly have two widespread stems with a partially complementary distribution within western Bantu (Bostoën et al. 2013a). In certain present-day

languages, such as Ntandu and Yombe in Table 5 below, the term actually refers to the closely related species *Erythrophleum suaveolens*, which is used for the same purposes. Hence, in this case, rather than being true synonyms, the two terms possibly used to be near-synonyms, which subsequently started to designate the same species.

Table 5. Reflexes of *-pòmí/*-kaca ‘*Erythrophleum guineense/suaveolens*’ in West-Coastal Bantu

Sub-branch	Language	Country	Term	Source
Yanzi	Nsong	DRC	<i>épwim</i>	(Koni Muluwa 2014: 58)
	Mpiin	DRC	<i>kípwm</i>	(Koni Muluwa 2014: 58)
	Yans	DRC	<i>nkay; ipæm</i>	(Koni Muluwa & Bostoen 2015: 145)
KLC	Ntandu	DRC	<i>nkása</i>	(Daeleman & Pauwels 1983: 176)
	Yombe	DRC	<i>nkaása</i>	(De Grauwe 2009: 83)
Nzebi-Mbete-Teke	Duma	Gabon	<i>mukasa</i>	(Raponda-Walker & Sillans 1995: 227)
	Nzebi	Gabon	<i>mukasa</i>	(Raponda-Walker & Sillans 1995: 227)
	Ndumu	Gabon	<i>okasa</i>	(Raponda-Walker & Sillans 1995: 227)

Finally, a certain number of names for useful wild plants seem to be Proto-West-Coastal Bantu innovations in the sense that they occur in West-Coastal Bantu sub-branches, but are not attested outside West-Coastal Bantu. One such case is the common name for the oil bean tree (*Pentaclethra macrophylla*), which is a fast-growing tree to 25 m high that is multifunctional among West-Coastal Bantu speech communities. The timber is used for construction works, for the fabrication of utensils, such as mortars, and for the production of charcoal. The seed pods can be used for fuel and also yield lye used for soap. The leaves host edible caterpillars and are used to produce a decoction for treating diarrhea or headache, while the bark serves in infertility treatments (Raponda-Walker & Sillans 1995; Latham 2004; Koni Muluwa 2014). As shown in Table 6, a cognate term for this tree is recurrent in West-Coastal Bantu. We tentatively propose the reconstruction *-pánji for this comparative series. The reflexes from the Kikongo Language Cluster clearly indicate an initial consonant *p (cf. Bostoen et al. 2013b: 64), which was retained as such in the Ntandu term *mpáansa*, which actually refers to the seed pods and not to the tree itself which is called *ngáansi*. The difference in stem-initial consonant can be accounted for by the fact that the name for the pods takes a noun prefix of classes 9/10 (singular/plural), which is a non-syllabic nasal having a conservative effect on the following consonant, while the tree name takes a noun prefix of classes 3/4 (singular/plural), which is a syllabic nasal not having this conservative effect, because it originally had a vowel following the nasal, i.e. *-mv (3), *-mi (4) (cf. Bostoen & de Schryver 2015). As for the alternation in final vowel observed between the two

Ntandu terms, this seems to be a variation that is recurrent across West-Coastal Bantu. However, the umlaut of the initial vowel observed in several languages of the Yanzi and Nzebi-Mbete-Teke subgroups calls for the reconstruction of an initial low vowel **a* and a final front vowel **ɪ*, as this is a regular sound shift among these languages (Bostoen & Koni Muluwa 2014). The final vowel *a* is mainly observed within the Kikongo Language Cluster and is probably a later innovation. As for the tones, the Ntandu reflexes manifest the same tone pattern as the **-gómbo* reflex. As it is impossible to discriminate between **HL* and **HH*, we only reconstruct a high tone for the first syllable for the time being.

Table 6. Reflexes of **-pánji* ‘*Pentaclethra macrophylla*’ in West-Coastal Bantu

Sub-branch	Language	Country	Term	Source
Yanzi	Nsong	DRC	<i>mówendz</i>	(Koni Muluwa 2014: 73)
	Mpiin	DRC	<i>múwendz</i>	(Koni Muluwa 2014: 73)
	Ngong	DRC	<i>mówándz</i>	(Koni Muluwa 2014: 73)
KLC	Hungan	DRC	<i>múwándz</i>	(Koni Muluwa 2014: 73)
	Yombe	DRC	<i>mváanza</i>	(De Grauwe 2009: 75)
	Ntandu	DRC	<i>ngáansi</i> ; <i>mpáansa</i>	(Daeleman & Pauwels 1983: 201)
	Laadi	Congo	<i>kihanzi</i>	(Adjanohoun 1998)
Nzebi-Mbete-Teke	Punu	Gabon	<i>muvandji</i>	(Raponda-Walker & Sillans 1995: 244)
	Duma	Gabon	<i>mupandji</i>	(Raponda-Walker & Sillans 1995: 244)
	Nzebi	Gabon	<i>muwendji</i>	(Raponda-Walker & Sillans 1995: 244)
	Laali	Congo	<i>muwai</i>	(Adjanohoun 1998)

5. Conclusions

The comparative lexical data considered in this article suggest that the first Bantu speakers who emerged south of the rainforest about 2500 years ago knew how to cultivate plants. The circumstantial evidence supporting this conclusion is the reconstruction of names for five distinct crops into Proto-West-Coastal Bantu, i.e. pearl millet (*Pennisetum glaucum*), okra (*Hibiscus/Abelmoschus esculentus*), cowpea (*Vigna unguiculata*), Bambara groundnut (*Vigna subterranea*) and plantain (*Musa spp.*). Since none of these crops were domesticated in Bantu-speaking Central Africa, the possibility to reconstruct names for them in an ancestral Bantu language is a strong indication of the fact that by that time Bantu speakers not only consumed crops, but also cultivated plants. This conclusion founded on lexical data is in line with the appearance of pearl millet and plantain in the archaeological record of Central Africa around the same period, i.e. 3000 to 2500 years ago. The

presence of domesticated plants in the archaeological record is conclusive evidence for cultivation. While cultivation refers to “any human activity that increases the yield of harvested or exploited plants” and “can be practiced with wild or domesticated plants”, domestication is a process which “only occurs under cultivation” and leads to “genetic, morphological and physiological changes of plants” (Neumann 2005: 250). Such conclusive evidence – both direct archaeological and indirect linguistic – is missing for the era corresponding to the assumed start of the Bantu Expansion, i.e. around 5000 years ago. The fact that the crop names reconstructible to Proto-West-Coastal Bantu do not date back to Proto-Bantu but are still shared with certain other Bantu branches fits in rather well with the hypothesis that Bantu speech communities acquired them in the course of their rapid migration through the Central rainforest block, which was facilitated thanks to the climate-induced opening of the forest around 2500 years ago. This climate change also induced the increased seasonality as well as savannah environment that was needed for the cultivation of crops such as pearl millet.

The considerable lapse of time between the beginning of the Bantu Expansion and the first conclusive evidence for plant cultivation and domestication, i.e. at least two millennia, suggests that the emergence of agriculture in Central Africa was indeed “a slow revolution” (Vansina 1994–1995). Its contribution to the subsistence of early Bantu speech communities grew only very steadily. Farming can therefore not have been the principal driving force behind the initial phases of the Bantu Expansion. Before Bantu speakers started to cultivate domesticated crops, as they certainly did as soon as they arrived south of the rainforest, they no doubt protected and increased the yield of wild plants available in their natural habitat, such as yams and several tree species for which vocabulary can be reconstructed in Proto-Bantu. The recurrent finds of oil palm (*Elaeis guineensis*) and bush-candle (*Canarium schweinfurthii*) remains in archaeological sites associated with early Bantu-speaking village communities may indeed point towards early arboriculture, even if it is hard to tell from the archaeobotanical record whether people just harvested from wild stands or already managed their forests, as present-day rainforest dwellers commonly do (Kahlheber et al. 2009: 261).

Moreover, the possibility to reconstruct crop names in Proto-West-Coastal Bantu, along with the archaeobotanical evidence for some of these crops from roughly the same period, should not be taken yet as evidence for agricultural intensification and surplus creation, often seen as pathways to societal complexity (McIntosh 1999: 4). As Neumann (2005: 250) puts it, “a single grain of domesticated sorghum does not justify calling the corresponding human population ‘farmers’”. Such is true for a single grain of pearl millet and for a single banana phytolith and even more for the reconstruction of some crop names. Although the first Bantu speakers south of the rainforest knew how to cultivate certain crops, they still

exploited intensively the different ecosystems to which they had access as part of their subsistence economy and their wider culture. Even if they had slightly moved towards the agriculturalists' side of the large continuum between hunter-gatherers and farmers in comparison with their ancestors, they still largely depended on the plant resources that they could collect in their natural environment, as is evidenced by a preliminary assessment of wild plant names that can be reconstructed to Proto-West-Coastal Bantu. While the reconstructible crop vocabulary is fairly limited, inherited names for wild plants shared between West-Coastal Bantu languages are numerous and would still increase if better ethnobotanical data were available. Wild or semi-domesticated plants were not only used for nutritional purposes, but also had various material-cultural, medicinal and ritual applications, many of which have persisted until today.

The fact that the vocabulary for different crops, such as cowpea, Bambara groundnut, okra, amaranth and roselle, still underwent considerable innovation in distinct branches of West-Coastal Bantu suggests that plant cultivation systems were still subject to important changes after West-Coastal Bantu speech communities had left their ancestral homeland south of the forest. Farming only became a more predominant subsistence strategy once they had started to migrate towards the Atlantic coast in the West and the Bandundu region in the East and it was definitely further boosted at the time of the first contacts with Europeans, i.e. from the late 15th century onwards. Many present-day crops, such as maize, cassava, sweet potato, peanut, common bean, etc., were introduced in Central Africa as part of the Columbian exchange and were often designated by inherited Bantu names which underwent semantic shift, e.g. 'pearl millet' > 'maize'; 'yam' > 'sweet potato'; 'Bambara groundnut' > 'peanut'; 'cowpea' > 'common bean', etc.

As Katharina Neumann has recently put it in a comment on Bostoen et al. (2015: 374), "basic questions on diet and subsistence of the 'Bantu' immigrants are still completely open". In order to answer these basic questions not only more dedicated historical linguistic research, but also – and first and foremost – more dedicated archaeobotanical research in Central Africa is needed, for instance to establish whether the plants for which we could reconstruct vocabulary in Proto-West-Coastal Bantu can also be retrieved in the archaeological record. It is only through such a joint cross-disciplinary approach that we will succeed in transforming our understanding of how the "middle ground" looked like in early Bantu speech communities and how it evolved through time. While archaeologists will focus on the means of subsistence that have left retrievable remains in Central African soils, historical linguists will additionally – but not exclusively – reconstruct the vocabulary for those plants (and animals) that are now archaeologically invisible. Such thinking across the disciplines will prove indispensable in order to conceive language dispersals beyond farming.

Acknowledgements

We wish to thank Sebastian Dom and Heidi Goes for their reading of an earlier version of this chapter. The usual disclaimers apply.

Abbreviations

CW	Central-Western Bantu
E	Eastern Bantu
KLC	Kikongo language cluster
NW	North-Western Bantu
SW	South-Western Bantu
WCB	West-Coastal Bantu

References

- Adjanohoun, E. J. 1998. *Médecine traditionnelle et pharmacopée. Contribution aux études ethnobotaniques et floristiques en R.P. du Congo*. Brazzaville: ACCT.
- Barbieri, Chiara *et al.* 2014. Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in southern Africa. *PLoS ONE* 9: e99117. doi:10.1371/journal.pone.0099117
- Bastin, Yvonne, Coupez, André, Mumba, Evariste & Schadeberg, Thilo C. 2002. *Bantu Lexical Reconstructions* 3. Tervuren: Royal Museum for Central Africa. <linguistics.africamuseum.be/BLR3.html> (30 March 2015).
- Bayon, Germain *et al.* 2012. Intensifying weathering and land use in Iron Age Central Africa. *Science* 335: 1219–1222. doi:10.1126/science.1215400
- Bellwood, Peter & Renfrew, Colin. 2002. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- Bingham, M. G. 2005. Western Province Species List. Ms, Lusaka.
- Blanchon, Jean Alain. 1991. Le pounou (B 43), le mpongwe (B 11a) et l'hypothèse fortis/lenis. *Pholia* 6: 49–83.
- Blanchon, Jean Alain & de Nadaillac, Luc. 1987. Malcolm Guthrie et la tonalite des nominaux nzèbi. *Pholia* 2: 47–73.
- Blench, Roger. 1994–1995. Evidence for the inception of agriculture in the Nigeria-Cameroon borderland. *Azania* 29–30: 83–102.
- Blench, Roger. 2006. *Archaeology, Language and the African Past*. Lanham MD: Altamira Press.
- Blench, Roger. 2009. Bananas and plantains in Africa: Re-interpreting the linguistic evidence. *Ethnobotany Research & Applications* 7: 363–380. doi:10.17348/era.7.0.363-380
- Bostoën, Koen. 2006–2007. Pearl millet in early Bantu speech communities in Central Africa: A reconsideration of the lexical evidence. *Afrika und Übersee* 89: 183–213 (published in 2010).
- Bostoën, Koen. 2007. Pots, words and the Bantu problem: On lexical reconstruction and early African history. *Journal of African History* 48: 173–199. doi:10.1017/S002185370700254X

- Bostoen, Koen. 2014. Wild trees in the subsistence economy of early Bantu speech communities: A historical-linguistic approach. In *Archaeology of African Plant Use*, Chris J. Stevens, Sam Nixon, Mary Anne Murray & Dorian Q. Fuller (eds), 129–140. Walnut Creek CA: Left Coast Press.
- Bostoen, Koen, Bernard Clist, Charles Doumenge, Rebecca Grollemund, Jean-Marie Hombert, Joseph Koni Muluwa & Jean Maley. 2015. Middle to Late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of West Central-Africa. *Current Anthropology* 56: 354–384. doi:10.1086/681436
- Bostoen, Koen & de Schryver, Gilles-Maurice. 2015. Linguistic innovation, political centralization and economic integration in the Kongo kingdom: Reconstructing the spread of prefix reduction. *Diachronica* 32: 139–185. doi:10.1075/dia.32.2.01bos
- Bostoen, Koen, Grollemund, Rebecca & Koni Muluwa, Joseph. 2013a. Climate-induced vegetation dynamics and the Bantu expansion: Evidence from Bantu names for pioneer trees (*Elaeis guineensis*, *Canarium schweinfurthii* and *Musanga cecropioides*). *Comptes Rendus Geoscience* 345: 336–349. doi:10.1016/j.crte.2013.03.005
- Bostoen, Koen & Koni Muluwa, Joseph. 2014. Umlaut in the Bantu B80 languages of the Kwilu (DRC). *Transactions of the Philological Society* 112: 209–230. doi:10.1111/1467-968X.12033
- Bostoen, Koen, Ndonga Tshiyayi, Odjas & de Schryver, Gilles-Maurice. 2013b. On the origin of the royal Kongo title ngangula. *Africana Linguistica* 19: 53–83.
- Brcic, Terry M., Willis, Kathy J., Harris, David J., Telfer, Matt W. & Bailey, Richard M. 2009. Fire and climate change impacts on lowland forest composition in northern Congo during the last 2580 years from palaeoecological analyses of a seasonally flooded swamp. *Holocene* 19: 79–89. doi:10.1177/0959683608098954
- Bunkheti, Kifindi. 1997. *Recherches en grammaire du suku: langue bantu (H.32) de la vallée du Kwango (Angola-Kongo Zaïre)*. Paris: Paris III, Sorbonne Nouvelle.
- Clist, Bernard. 2005. *Des premiers villages aux premiers européens autour de l'estuaire du Gabon: quatre millénaires d'interactions entre l'homme et son milieu*. Bruxelles: Université libre de Bruxelles, thèse de doctorat.
- Clist, Bernard. 2012. Vers une réduction des préjugés et la fonte des antagonismes: Un bilan de l'expansion de la métallurgie du fer en Afrique sud-saharienne. *Journal of African Archaeology* 10: 71–84. doi:10.3213/2191-5784-10205
- Crane, Thera M., Hyman, Larry M. & Tukumu, Simon Nsielanga. 2011. *A Grammar of Nzadi [B.865]: A Bantu Language of the Democratic Republic of Congo with an Appendix of Proto-Bantu-Nzadi Sound Correspondences by Clara Cohen*. Berkeley CA: University of California Publications.
- D'Andrea, A. Catherine, Logan, Amanda L. & Watson, Derek J. 2006. Oil palm and prehistoric subsistence in tropical Africa. *Journal of African Archaeology* 4: 195–222. doi:10.3213/1612-1651-10072
- Daeleman, Jan. 1977. A comparison of some zone B languages in Bantu. *Africana Linguistica* 7: 93–144. doi:10.3406/aflin.1977.899
- Daeleman, Jan. 1983. Les réflexes du proto-bantu en ntandu (dialecte koongo). *Mélanges de culture et de linguistique africaines publiés à la mémoire de Leo Stappers*, Clémentine Faïk-Nzuzi Madiya & Erika Sulzmann (eds), 331–397. Berlin: Dietrich Reimer.
- Daeleman, Jan & Pauwels, Luc. 1983. Notes d'ethnobotanique ntandu (Kongo). Principales plantes de la région de Kisaantu: noms ntandu et noms scientifiques. *Africana Linguistica* 9: 149–256. doi:10.3406/aflin.1983.920

- de Filippo, Cesare, Heyn, Patricia, Barham, Larry, Stoneking, Mark & Pakendorf, Brigitte. 2010. Genetic perspectives on forager-farmer interaction in the Luangwa Valley of Zambia. *American Journal of Physical Anthropology* 141: 382–394.
- De Grauwe, Jan. 2009. *Lexique yoómbé-français, avec index français-yoómbé (bantu H16c)* [Tervuren Series for African Language Documentation and Description]. Tervuren: Royal Museum for Central Africa.
- de Maret, Pierre. 1994–1995. Pits, pots and the far-west streams. *Azania* 29–30: 318–323.
- de Rop, Albert Jozef. 1953. Het toonsysteem van het Lomongo en het Oerbantu. *Kongo-Overzee* 19: 413–419.
- de Rop, Albert Jozef. 1958. *Éléments de phonétique historique de Lomongo*. Léopoldville: Université Lovanium.
- de Schryver, Gilles-Maurice, Grollemund, Rebecca, Branford, Simon & Bostoen, Koen. 2015. Introducing a state-of-the-art phylogenetic classification of the Kikongo language cluster. *Africana Linguistica* 21: 87–162.
- Destro-Biso, Giovanni *et al.* 2004. The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *The American Naturalist* 163: 212–226. doi:10.1086/381405
- Diamond, Jared & Bellwood, Peter. 2003. Farmers and their languages: The first expansions. *Science* 300: 597–603. doi:10.1126/science.1078208
- Eggert, Manfred K. H. *et al.* 2006. Pits, graves and grains: Archaeological and archaeobotanical research in southern Cameroon. *Journal of African Archaeology* 4: 273–298. doi:10.3213/1612-1651-10076
- Fontaney, V. Louise. 1984. Notes towards a description of Teke (Gabon). *Pholia* 1: 47–70.
- Gossweiler, John. 1953. *Nomes indígenas de plantas de Angola*. Luanda: Imprensa Nacional.
- Greenberg, Joseph Harold. 1972. Linguistic evidence regarding Bantu origins. *Journal of African Languages and Linguistics* 13: 189–216.
- Grollemund, Rebecca *et al.* 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the United States of America* 112: 13296–13301. doi:10.1073/pnas.1503793112
- Hagendorens, Joseph. 1975. *Dictionnaire otetela-français* [Série III v. 5]. Bandundu: Ceeba.
- Hamon, Serge & Charrier, André. 1997. Les gombos. *L'amélioration des plantes tropicales*, André Charrier, Michel Jacquot, Serge Hamon & Dominique Nicolas (eds), 313–333. Montpellier: CIRAD, ORSTOM.
- Hannan, Michael. 1974. *Standard Shona Dictionary*. Salisbury: Mardon printers Ltd for Rhodesia Literature Bureau.
- Holden, Claire Janaki. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society of London Series B-Biological Sciences* 269: 793–799. doi:10.1098/rspb.2002.1955
- Hombert, Jean-Marie. 1986. The development of nasalized vowels in the Teke language group. *The Phonological Representation of Suprasegmentals*, Koen Bogers, Harry van der Hulst & Marten Mous (eds), 359–379. Dordrecht: Foris. doi:10.1515/9783110866292-019
- Hubau, Wannes *et al.* 2014. Archaeological charcoals as archives for firewood preferences and vegetation composition during the late Holocene in the southern Mayumbe, Democratic Republic of the Congo (DRC). *Vegetation History and Archaeobotany* 23: 591–606.
- Hubau, Wannes *et al.* 2015. Charcoal-inferred Holocene fire and vegetation history linked to drought periods in the Democratic Republic of Congo. *Global Change Biology* 21: 2296–2308. doi:10.1111/gcb.12844

- Hulstaert, Gustaaf. 1941. Tonetiek van Lomongo en Tshiluba. *Aequatoria* 4: 56–58.
- Hulstaert, Gustaaf. 1957. *Dictionnaire lomongo-français* [*Annales du Musée royal du Congo belge, Série in-8p0s, Sciences de l'homme, Linguistique* 16]. Tervuren: Musée royal du Congo belge.
- Kabinda, Kongolo. 1988. *La dendronymie du kanyok. Essai d'analyse morpho-sémantique et socio-culturelle*. Lubumbashi: Université de Lubumbashi, Mémoire de licence.
- Kahlheber, Stefanie, Bostoën, Koen & Neumann, Katharina. 2009. Early plant cultivation in the Central African rain forest: First millennium BC pearl millet from South Cameroon. *Journal of African Archaeology* 7: 253–272. doi:10.3213/1612-1651-10142
- Kahlheber, Stefanie, Eggert, Manfred K. H., Seidensticker, Dirk & Wotzka, Hans-Peter. 2014. Pearl millet and other plant remains from the Early Iron Age site of Boso-Njafo (Inner Congo Basin, Democratic Republic of the Congo). *African Archaeological Review* 31: 479–512. doi:10.1007/s10437-014-9168-1
- Koni Muluwa, Joseph. 2010. Plantes, animaux et champignons en langues bantu. Etude comparée de phytonymes, zoonymes et myconymes en nsong, ngong, mpiin, mbuun et hungan (Bandundu, RD Congo). Thèse de doctorat, Université libre de Bruxelles.
- Koni Muluwa, Joseph. 2014. *Noms et usages de plantes, animaux et champignons chez les mbuun, mpiin, ngong, nsong et hungan en RD Congo* [Tervuren Series for African Language Documentation and Description]. Tervuren: Musée royal de l'Afrique centrale.
- Koni Muluwa, Joseph & Bostoën, Koen. 2015. *Lexique comparé des langues bantu du Kwilu (République démocratique du Congo)*. Cologne: Rüdiger Köppe.
- Kouarata, Guy Noël. 2016. *Dictionnaire beembe-français*. Saint-Denis: Edilivre.
- Latham, Paul. 2004. *Useful Plants of Bas-Congo Province, Democratic Republic of the Congo*. London: United Kingdom Department for International Development.
- Lejju, Julius B., Robertshaw, Peter & Taylor, David. 2006. Africa's earliest bananas? *Journal of Archaeological Science* 33: 102–113. doi:10.1016/j.jas.2005.06.015
- Lejju, Julius B., Taylor, David & Robertshaw, Peter. 2005. Late-Holocene environmental variability at Munsu archaeological site, Uganda: a multicore, multiproxy approach. *The Holocene* 15: 1044–1061.
- Linton, Pauline. 2016. *Le système verbal de la langue latege* [Deuxième Révisio]. Léconi: Global Outreach Mission.
- MacGaffey, Wyatt. 1991. *Art and Healing of the Bakongo, Commented by Themselves: Minkisi from the Laman Collection* [Monograph series no 16]. Stockholm: Folkens museum-Etnografiska.
- Maley, Jean, Giresse, Pierre, Doumenge, Charles & Charly, Favier. 2012. Comment on intensifying weathering and land use in Iron Age Central Africa. *Science* 337: 1040. doi:10.1126/science.1221820
- Maniacky, Jacky. 2005. Quelques thèmes pour igname en bantu. *Studies in African Comparative Linguistics, with Special Focus on Bantu and Mande*, Koen Bostoën & Jacky Maniacky (eds), 165–187. Tervuren: Royal Museum for Central Africa.
- Mbida Mindzié *et al.* 2005. The initial history of bananas in Africa. A reply to Jan Vansina. *Azania* 40: 128–135. doi:10.1080/00672700509480419
- Mbida Mindzié, Christophe, Neer, Wim Van, Doutrelepon, Hugues & Vrydaghs, Luc. 2000. Evidence for banana cultivation and animal husbandry during the first millennium BC in the forest of southern Cameroon. *Journal of Archaeological Science* 27: 151–162. doi:10.1006/jasc.1999.0447
- McCann, James. 2009. *Stirring the Pot: A History of African Cuisine*. Athens OH: Ohio University Press.

- McIntosh, Susan Keech. 1999. Pathways to complexity: An African perspective. *Beyond Chiefdoms: Pathways to Complexity in Africa*, Susan Keech McIntosh (ed.), 1–30. Cambridge: CUP. doi:10.1017/CBO9780511558238.002
- Mouandza, Jean-Daniel. 1991. *Esquisse phonologique du iyaa (Parler bantu du Congo)*. Brazzaville: Université Marien Ngouabi, mémoire de DES.
- Ndinga-Koumba-Binza, Hugues Steve. 2000. *Phonologie du civilisé de mayumba: Langue bantu du Gabon (H12a)*. Libreville: Université Omar Bongo, mémoire de maîtrise.
- Neumann, Katharina. 2005. The romance of farming: Plant cultivation and domestication in Africa. *African Archaeology: A Critical Introduction*, Ann Brower Stahl (ed.), 249–275. Malden, MA: Blackwell Pub.
- Neumann, Katharina *et al.* 2012a. First farmers in the Central African rainforest: A view from southern Cameroon. *Quaternary International* 249: 53–62. doi:10.1016/j.quaint.2011.03.024
- Neumann, Katharina *et al.* 2012b. Comment on intensifying weathering and land use in Iron Age Central Africa. *Science* 337: 1040. doi:10.1126/science.1221747
- Neumann, Katharina & Hildebrand, Elisabeth. 2009. Early bananas in Africa: The state of the art. *Ethnobotany Research & Applications* 7: 353–362. doi:10.17348/era.7.0.353-362
- Ngomanda, Alfred, Neumann, Katharina, Schweizer, Astrid & Maley, Jean. 2009. Seasonality change and the third millennium BP rainforest crisis in southern Cameroon (Central Africa). *Quaternary Research* 71: 307–318. doi:10.1016/j.yqres.2008.12.002
- Pasch, Helma. 1979. *Linguistische Aspekte der Verbreitung lateinamerikanischer Nutzpflanzen in Afrika*. PhD dissertation, University of Cologne.
- Patin, Etienne *et al.* 2014. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nature Communications* 5: 3163. doi:10.1038/ncomms4163
- Philippson, Gérard & Bahuchet, Serge. 1994–1995. Cultivated crops and Bantu migrations in central and eastern Africa: A linguistic approach. *Azania* 29–30: 103–120.
- Phillipson, David W. 2003. Language and farming dispersals in sub-Saharan Africa, with particular reference to the Bantu-speaking peoples. *Examining the Farming/Language Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 177–187. Cambridge: McDonald Institute for Archaeological Research.
- Quintana-Murci, L. *et al.* 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences USA* 105: 1596–1601. doi:10.1073/pnas.0711467105
- Raponda-Walker, André & Sillans, Roger. 1995. *Les Plantes utiles du Gabon* [Encyclopédie Biologique 56.]. Libreville: Fondation Raponda-Walker Edition, Sépia & Centre Culturel Saint-Exupéry.
- Ricquier, Birgit. 2016. A foodie's guide to Kongo language history: Early events, North versus South, and the innovative West. *Africana Linguistica* 22: 107–146.
- Ricquier, Birgit & Bostoen, Koen. 2011. Stirring up the porridge: How early Bantu speakers prepared their cereals. *Windows on the African Past: Current Approaches to African Archaeobotany*, Ahmed G. Fahmy, Stefanie Kahlheber & Catherine D'Andrea (eds), 209–224. Frankfurt: Africa Magna Verlag.
- Russell, Thembi, Silva, Fabio & Steele, James. 2014. Modelling the spread of farming in the Bantu-speaking regions of Africa: An archaeology-based phylogeography. *PLoS ONE* 9: e87854. doi:10.1371/journal.pone.0087854

- Schwartz, Dominique. 1992. Assèchement climatique vers 3000 B.P. et expansion Bantu en Afrique centrale atlantique: Quelques réflexions. *Bulletin de la Société Géologique de France* 163: 353–361.
- SPIAF. 1988. *Liste des essences forestières du Zaïre*. Kinshasa: Département des Affaires Foncières, Environnement et Conservation de la Nature.
- Storrs, Adrian E. G. 1995. *Know Your Trees. Some of the Common Trees Found in Zambia*. Nairobi: Regional Soil Conservation Unit.
- Thornell, Christina. 2004. Wild plant names in the Mpiemo language. *Africa & Asia: Göteborg Working Papers on Asian and African Languages and Literatures* 4: 57–89.
- van Everbroecke, René V. 1985. *Maloba ma lokóta lingála / Dictionnaire lingála: lingála-français, français-lingála*. Limete-Kinshasa: Editions l'Épiphanie.
- Van Neer, Wim. 2000. Domestic animals from archaeological sites in West and Central Africa. *The Origins and Development in African Livestock: Archaeology, Genetics, Linguistics, and Ethnography*, Roger Blench & Kevin C. MacDonald (eds), 163–190. London: University College of London Press.
- Vansina, Jan. 1990. *Paths in the Rainforest: Toward a History of Political Tradition in Equatorial Africa*. Madison: University of Wisconsin Press.
- Vansina, Jan. 1994–1995. A slow revolution: Farming in subequatorial Africa. *Azania* 29–30: 15–26.
- Vansina, Jan. 1995. New linguistic evidence and the Bantu expansion. *Journal of African History* 36: 173–195. doi:10.1017/S0021853700034101
- Vansina, Jan. 2003. Bananas in Cameroon c. 500 BCE? Not proven. *Azania* 38: 174–176. doi:10.1080/00672700309480367

Expanding the methodology of lexical examination in the investigation of the intersection of early agriculture and language dispersal

Brian D. Joseph

The Ohio State University

Analysis of agricultural vocabulary remains one of the most compelling methodologies bearing on Renfrew's Farming/Language Dispersal Hypothesis, by which the reconstructed lexicon for a proto-language of a well-dispersed language family is predicted to contain several agricultural items. Mostly, though, this methodology has involved noting the presence or absence of particular lexical items for a given proto-language and drawing inferences from that, or working out root derivations and drawing appropriate inferences. I propose here two new types of lexically based argument, by way of expanding the methodology of lexical examination and analysis, looking first at derivational processes involved in the creation of relevant words and the meaning that such processes add to the derivative, and then at religious rituals and mythology to examine the embedding of agricultural vocabulary into the religious practices and mythological tales associated with early Indo-European culture. Ultimately, then, I argue that it is not enough to just look at the meanings of particular words and to try to develop a sense of what they originally meant, nor is it enough to determine the source of the words (derivation, etymology). Rather, one also has to look at how the words were used, what is reconstructible about the use and form of the word, and what the cultural context was for the words. Only then can insights derived from lexical examination be used in developing a sense of prehistory.

Keywords: Farming/Language Dispersal Hypothesis, lexical analysis, derivation, etymology, Indo-European, religious ritual, mythology

1. Introduction

Analysis of agricultural vocabulary remains one of the most compelling methodologies bearing on Renfrew's Farming/Language Dispersal Hypothesis, by which the reconstructed lexicon for a proto-language of a well-dispersed language family is predicted to contain several agricultural items. However, for the most part, this methodology has involved three different types of analysis. In one type, which can be called the "Proto-language Lexeme" approach, the presence or absence of particular lexical items for a given proto-language is noted, and appropriate inferences are drawn from that; in a second type of analysis, which can be called the "Root Etymology" approach, if root derivations for agriculturally relevant words can be worked out, then one can get a glimpse into the cultural mindset, so to speak, underlying the formation of a given item, as well as into the technology involved in such a derivation, thus undertaking a kind of "Wörter und Sachen" analysis; finally, in a third type, which can be called the "Loanword" approach, if borrowings can be detected that bear on agriculture, then one presumably has direct evidence for a particular kind of diffusion of agricultural knowledge.¹

These varied lexical methodologies are useful and have led to interesting insights over the years, but I suggest here that there are yet more ways to use lexical evidence. In particular, I propose two further types of lexically based argumentation, by way of expanding the methodological range of lexical examination and analysis that pertains to farming vocabulary and the inferences that may be derived from it.

First, though, it is useful to exemplify these types of analysis and offer a critique of them, so that the novel suggestions have a standard against which they can be compared.

2. Lexical analysis exemplified, and critiqued

In this section, I use material from the Indo-European family first to illustrate the various types of lexical analysis and then to provide the basis for a critical appraisal of the forms in question and of their value for deductions about agriculture among the Proto-Indo-Europeans, the speakers of the reconstructed Proto-Indo-European

1. Both the "Root Etymology" and the "Loanword" approaches could be considered subtypes of a general approach seeking the ultimate source of particular reconstructed proto-language lexemes.

language. In doing so, I give an assessment as well of the methodology involved, in a sense, then, first offering reconstruction and then offering deconstruction.²

2.1 The Proto-language Lexeme approach

For the first type of analysis, we can consider the following. There is an eminently reconstructible word for Proto-Indo-European for a farming tool, namely the plow, that has the form $*H_2erH_3-tro-m$ (with neuter nominative/accusative singular $*-m$), created from the root $*H_2erH_3-$ with the instrument-noun suffix $*-tro-$. This reconstruction is indicated by the cognate set of Greek $\acute{\alpha}\rho\omicron\tau\rho\omicron\nu$ (*arotron*), Old Irish *arathar*, Armenian *arawr* (< $*ar\bar{a}-tro-$), and Latin *arātrum*; relevant here too are forms with well-instantiated variants of the $*-tro-$ suffix, namely Lithuanian *árklas*, with $-kl-$ from $*-tl-$, and the Slavic forms with the $*-d^hlo-$ variant found regularly in Slavic, such as Serbian *rālo* and Czech *rādlo*, from Proto-Slavic $*ordlo$ (from $*H_2erH_3-d^hlo-$). The root might well mean ‘to plow’, so that the derived word would be ‘the instrument through which plowing takes place’, but given that the root is the basis for the Hittite word for ‘rake’ (discussed below, in § 3), the original meaning may have been ‘to break ground’ (as Tischler 1983: 122 suggests).

Moreover, with the same instrumental suffix, one finds evidence for another agricultural tool, specifically one that is grain-related, in various cognate words for ‘sieve’, an implement used in harvesting grains: Old Irish *criathar* ‘sieve’ from (full-grade) $*krei-trom$, where the root is $*krei-$ ‘select’, and Old English *hridder* (with a secondary variant *hriddel*) from a zero-grade ($*kri-tro-$); relevant here too is Latin *cribrum* ‘sieve’, from the same root but with a variant form of the $*tro-$ suffix,

2. I work with a somewhat traditional but, I believe, widely accepted phonological system for Proto-Indo-European; see Fortson (2010: 53–74), for this view and an explication of the motivation for it. The symbol <’> indicates a stop at the palatal point of articulation, so that <ǵ> is a voiced unaspirated palatal stop. <H> stands for a laryngeal consonant, one of three such sounds reconstructed for Proto-Indo-European, the phonetics of which are somewhat unclear (but are certainly not “laryngeal” consonants phonetically); I use <H₁> for the laryngeal that has no vowel-coloring effect on an adjacent $*e$, <H₂> for the laryngeal that colors an adjacent $*e$ to [a], <H₃> for the laryngeal that colors an adjacent $*e$ to [o], and <H> for a laryngeal whose vowel-coloring properties are indeterminate. All other phonetic symbols have their usual interpretation. I use the terms “full-grade” and “zero-grade” to refer to different ablaut grades of Proto-Indo-European roots and suffixes, the former referring to root forms that have the vowel $*e$ and the latter referring to root forms lacking the full-grade vowel $*e$. I give Greek forms in Greek letters with a transliteration following in parentheses.

specifically from **krei-d^hrom*. In each case, the meaning of the derivative would be ‘the instrument through which a certain kind of selection takes place’.³

The reconstructibility of a word for ‘plow’ can be taken as *prima facie* evidence supporting the hypothesis that the Proto-Indo-European community had a knowledge of cultivation and agriculture; moreover, a reconstructible word for ‘sieve’ would focus attention on grain-related farming.⁴ Indeed, a number of words for grains can be reconstructed for Proto-Indo-European; Kölligan 2017 gives the following summary:⁵

The PIE people were agricultors as can be seen in inherited terms for ‘grain’ such as **ǵr̥h₂-no-* ..., orig. ‘ground’, a verbal adjective built to the root **ǵerh₂-* ‘grind’ (that might be identical with **ǵerh₂-* ‘make/get old, wear down’ ...), **ieuo-* ‘corn, barley, spelt’ ..., **puHro-* ‘wheat’ ... (perhaps from **peuH-* ‘purify’, Skt. *punāti, pāvate*, i.e. that which is purified on the threshing floor), and **d^hoh₁neh₂-* ‘corn, seed’ ... (perhaps from **d^heh₁-* ‘put’ [sc. into the ground]). Also attested, though with more limited distribution, are **urug^hio-* ‘rye’ ... and **b^har-es-* ‘barley’ ...

Still, there are potential issues that prevent one from wholeheartedly endorsing these results. Most significantly, the **-tro-* suffix (with variants, as in Slavic) is well represented across the various branches of Indo-European and can be considered to be somewhat productive (Meillet 1964: 273). As such, it could be used to form an instrument noun at any time and could therefore presumably have been created in individual branches. Moreover, if the original meaning of **H₂erH₃-* were ‘to break ground’, then ‘plow’ could be a specialization of a noun meaning ‘the instrument through which breaking of ground takes place’. This raises the possibility that even though it is attested in several distinct points within Indo-European, both east (Slavic, Armenian) and west (Latin, Irish), the ‘plow’ meaning for this word could represent the result of independent semantic shifting within each point.⁶ Such considerations would mean that, strictly speaking, **H₂erH₃-tro-m* need not have been a part of Proto-Indo-European proper. Similarly, since the words for ‘sieve’ occur

3. The apparently metaphorical use of *sift* or *winnow* in English today, as in *to sift through / winnow the application files for the best candidate*, attests to the closeness of selection in general and selecting the most suitable grains via physical sifting.

4. See below, however, for a reconsideration of the basic root for the ‘plow’ word and its derivation, and also some discussion of ‘sieve’ in Hittite and elsewhere.

5. See also Mallory & Adams (1997: 51–2 (s.v. BARLEY), 236–7 (s.v. GRAIN), 409 (s.v. OATS), 491–2 (s.v. RYE)).

6. Moreover, if Armenian and Greek are developments from a deeper “Helleno-Armenian” dialect within Indo-European and Italic and Celtic share a deeper “Italo-Celtic” connection, the number of distinct points is reduced.

in the areally close Italic, Celtic (or Italo-Celtic, see footnote 6), and Germanic, one might suppose that they belong to a western Indo-European grouping, not necessarily representing a common innovation so much as possibly showing diffusion from one branch to another; in that case, it too would not necessarily be reconstructible for Proto-Indo-European itself. The same can be said for some of the grain-words that Kölligan reconstructs, especially those with a “more limited distribution”, such as ‘rye’ and ‘barley’ (though see § 3 for more on ‘barley’).

It must be admitted, though, that given their respective distributions, ‘plow’ would seem to have a better chance of being of Proto-Indo-European age than ‘sieve’. And one can easily suppose that the meaning of the root $*H_2erH_3-$ was specialized to ‘to plow’ in Proto-Indo-European times. Since it is hard to imagine that there was a verb meaning ‘to plow’ without the primary instrument for effecting the action of that verb, $*H_2erH_3-tro-m$ as a Proto-Indo-European word for ‘plow’ becomes a more compelling reconstruction. Nonetheless, the more general methodological caveat here is that positing specific words as members of a proto-language lexicon is fraught with difficulty, so that drawing inferences about cultural or technological history from the presence or absence of particular lexical items is equally fraught.

2.2 The root etymology approach

As for the second type of analysis, Kölligan’s assessment contains some speculation about the roots involved in nouns for grains. It should be noted, though, that if the grain-words represent derivatives of roots that have nothing to do with agriculture, e.g. ‘wear down’, ‘purify’, ‘put’, it could be that the specialization of their meanings to grain-related senses was a later phenomenon that occurred either post-Proto-Indo-European after the dispersal of the individual branches, or at a late stage within the proto-language.

A somewhat more complicated case that presents a wide range of caveats even in the face of a seemingly strong representation across the family and a clear root derivation is that of the word for ‘field’. A careful consideration of the issues it raises is important, however, for the methodological lessons to be learned from it.

Based on the equation offered by words for ‘(arable) field’ in various languages, specifically Latin *ager*, Greek ἀγρός (*agrós*), Sanskrit *ajra-*, and Gothic *akrs*, a reconstructible word for Proto-Indo-European, $*H_2eǵ-ro-$, with the meaning ‘(arable) field’, appears to be well called for. At this point this exercise appears to be like the proto-language lexeme approach discussed in § 2.1, with the reasoning being that if the Proto-Indo-European speech community had a word with such semantics, then arability of a field must have been a relevant notion for the Proto-Indo-Europeans,

and consequently the tools for making fields arable would also have been available to them.

However, one can go further, as this word appears clearly to be derived within Proto-Indo-European from a root $*H_2eǵ-$ – a derivation evident in each language too, cf. Latin *agō*, Greek ἄγω (*agō*), Sanskrit *ajā(mi)*, Old Norse *aka* – a root that means ‘to drive, to lead’ in the individual languages. Assuming – as one would in the “root etymology approach” – that this meaning is valid for Proto-Indo-European would entail that the derivative probably originally meant ‘driving-place, i.e. place where animals are driven’,⁷ as in plowing; this derivation would thus suggest established agricultural practices for the proto-language whereby this noun could be associated with this meaning of the verbal root.⁸

What makes it complicated is that all aspects of the derivation raise concerns; it is thus a particularly important lexical item to consider from a methodological standpoint. For instance, in Vedic Sanskrit, the earliest Sanskrit available,⁹ *ajra-* means ‘plain’ or ‘grassy field’, as contrasted with mountains (cf. Masica 1979 on this, drawing on Brandenstein). That detail could indicate that the meaning ‘arable field’ represents a later semantic shift, and therefore it is not to be reconstructed for Proto-Indo-European, despite the match across the languages. Indeed, traces of that presumed original meaning are found in derivatives in other languages, especially Greek ἄγριος (*ágrios*), ‘wild’ (i.e., “of the field”), which is matched exactly in form, and closely in meaning, by Vedic Sanskrit *ajriya-* ‘being in or connected with a field or plain’ (Monier-Williams 1899: s.v. *ajrya-*).

However, if $*H_2eǵ-ro-$ is a derivative from $*H_2eǵ-$ ‘to drive’, as one looking for evidence of agriculture in Proto-Indo-European society might posit, it is fair to ask how $*H_2eǵ-ro-$ could have at first had the meaning ‘grassy field, plain’. A semantic shift from something like “driving place” to “grassy field” does not seem particularly reasonable or well motivated.

A possible solution here might be to consider both meanings to be reconstructible for Proto-Indo-European, but at different chronological layers of Proto-Indo-European. This is especially feasible if we assume that what we call “Proto-Indo-European” actually represents a speech community that existed over

7. As Pokorny (1959: 6) puts it, “Ort, wo das Vieh hinausgetrieben wird”.

8. In a sense, the discussion concerning the derivation of the noun ‘plow’ from the verbal root ‘to plow’ in § 2.1 overlaps with this “root etymology approach”, except that with ‘plow’, the semantics of the verbal root made for a more obvious connection to the noun than with ‘field’ and ‘to drive’.

9. Vedic Sanskrit refers to the Sanskrit as found in the hymns of the Rigveda and related materials. The Rigveda is conventionally dated to about 1200 BC, though parts are clearly much older, showing cognate phraseology – not just words but full phrases but even thematic parallels – in other ancient Indo-European material, such as Homeric Greek or Hittite rituals.

a long time span and thus that semantic shift could have taken place in the course of what we still label as Proto-Indo-European. This is a distinct possibility, to be sure, but is essentially untestable. Moreover, the original impetus for the semantic derivation and the putative connection with the root $*H_2eǵ-$ in the meaning ‘to drive’ and subsequent semantic shifts would remain to be explained. Typological lexical semantics, the exploration of what sorts of semantic shifts are attested and are plausible and thus waiting to be invoked as parallels to a putative shift in reconstructed items or in derivatives, can be of assistance here, though nothing relevant immediately suggests itself here.

While such issues may suggest that the agricultural meaning is original after all, it could also mean that the derivation from $*H_2eǵ-$ ‘to drive’ needs to be reconsidered. And, indeed, from a formal standpoint, quite apart from the semantics, the derivation of $*H_2eǵ-ro-$ from $*H_2eǵ-$ ‘to drive’ is somewhat problematic. In particular, the suffix $-ro-$ usually created adjectives, not nouns, and usually had zero-grade of the root it attached to (Meillet 1964:267), as shown by such forms as Avestan *tiy-ra-* ‘sharp’ (root $*(s)teig-$ ‘to stick, to be sharp’), and Vedic Sanskrit *ug-rá-* ‘powerful, fierce’ (root $*H_2eug-$ ‘to increase’), *ṛj-rá-* ‘shining’, among others, this last with an exact cognate in Greek ἀργός (*argós*) ‘bright’ (from a presumed $*\text{ἀργρός}$ ($*argrós$)). While it is hard to see what other derivation for $*H_2eǵro-$ might be possible,¹⁰ the fact of a problematic derivation coupled with the semantic issues must give one pause in drawing too solid an inference about Proto-Indo-European agriculture from $*H_2eǵro-$, and thus more generally, from placing too much store in deriving cultural information from root etymologies. As seen in § 2.1, a shaky linguistic foundation for a cultural inference means that the inference itself is diminished in value.

2.3 The loanword approach

The loanword approach seeks to identify borrowings in the proto-language that allow for inferences about, in this case, agriculture and related matters. As such, it has a more direct cultural basis, as the borrowing of lexical items implies contact between speakers of different languages, and thus of different social groups.

By way of illustrating this approach, one can cite the word for ‘a kind of harmful insect’, reconstructed for Proto-Indo-European as $*math-$ by Pokorny (1959:700) on

10. Romain Garnier (p.c., September 2016), noting the unusual *e*-grade, speculated that perhaps one should reckon with a different root and a different segmentation altogether for $*H_2eǵro-$. For instance, if $*H_2eǵro-$ were segmented $*H_2e-ǵr-o-$, one might suppose it is composed of a preverb $*H_2e$ and a root $*ǵer-$; however, no known Proto-Indo-European preverb has that shape and no known Proto-Indo-European root has a reasonable semantic fit here.

the basis of the apparent cognate forms Armenian *mat'il* 'louse' and Gothic *maþa* 'moth', with a host of inner-Germanic cognates, including Old Icelandic *maþkr*, Old Swedish *matk*, and English *moth*. This reconstructed form word is phonologically unusual for Proto-Indo-European in two respects: the occurrence of a voiceless aspirate,¹¹ and the occurrence of **a*, a vowel which is rejected altogether for Proto-Indo-European by some Indo-Europeanists (see, e.g., Beekes 1995: 138–9) or recognized as occurring mainly in words that are marked in some way, e.g., as described by Meillet (1964: 99), "mots de caractère populaire, technique ou affectif". Beekes (ibid.) suggests also that words with **a* might be very old loans, a reasonable view inasmuch as phonological oddities are often associated with loan words. Thus, this word may well have been a borrowing into Proto-Indo-European; in this regard, Finnish *matikka* 'little worm' is relevant, as it is an apparent loanword from Swedish (as suggested by Pokorny) and thus shows that this is the sort of word that can be borrowed. Moreover, and more to the point for the discussion here, while there are many types of moths and harmful insects, particularly common among moths are those that attack grains, such as the Indian mealmoth (*plodia interpunctella*) and the Angoumois grain moth.¹² The argument here is that from some external source, Proto-Indo-European itself (as opposed to Armenian and Germanic independently) acquired this loanword designating an agricultural pest, which would suggest that Proto-Indo-European society had the sort of agriculture that would attract such pests. While it is of course a bit of speculation that the relevant pests were grain moths, associating this loanword with agriculture would provide a motivation for the borrowing, which otherwise would just be a random event.

A more specifically grain-related Proto-Indo-European lexeme that has been considered to be a borrowing is **b^har(e)s-* 'barley'. As Mallory and Adams (1997: 51) put it, "This word is found in the west and center of the IE world and is often taken to be a borrowing". They go on to mention Proto-Semitic **burr-/barr-* 'grain, threshed grain' as a possible source, though they note (ibid.) that "the distribution of cognates within Indo-European does not support direct connections with the Near East". As an alternative, they state that it could be a substratum word of "central or

11. The prevailing view about the Proto-Indo-European phonological system is that it did not have phonemic voiceless aspirated stops (see Fortson 2010: 56), though there are a few cognate sets that are suggestive of the need for reconstructing such sounds. See Joseph (1985) for some relevant discussion as well, especially pertaining to this word for 'moth'.

12. As a rule, I consider it bad form to cite Wikipedia as a source for anything linguistic, but I am out of my element when it comes to the entomological (as opposed to the etymological) side of moths, and have found the material and the links provided by relevant Wikipedia pages to be very helpful, e.g. <<https://entomology.ca.uky.edu/ef156>> and <https://en.wikipedia.org/wiki/Indian_mealmoth>.

western Europe”, but if so, they suggest, “it is a very old borrowing, taken across at a time when the various Indo-European dialects were not very much differentiated” (ibid.). If a borrowing, and if the source can be identified, then inferences can be drawn about agriculture and early Indo-European societies, but there is not necessarily great clarity here as to which of these hypotheses is correct.

Thus, there are several assumptions needed to make such borrowing-based inferences work, especially involving the identification of the ultimate sources of the loanwords and their original meaning. Such assumptions, if too many, might well prove ultimately to undermine the value of looking to loanwords for inferences about cultural diffusion. Thus loanword analysis, like the other types of lexical analysis surveyed in § 2.1 and § 2.2, is only as strong as the linguistic foundation it is built on.

2.4 Assessment

The upshot of this survey of various kinds of lexical analysis is that as potentially useful as these typical types of analysis are, other methods are needed to supplement them. While some such “other methods” might be envisioned that are of a nonlinguistic nature, lexical analysis offers yet other dimensions that can be exploited that are linguistic in nature. In the sections that follow, I present, discuss, and highlight further types of lexically based methods of analysis that illustrate other means of developing insights into a proto-language from an examination of proto-language lexical material.

3. Derivation

As the example involving ‘field’ (Greek ἀγρός, etc.) showed, examining the internal source of a word can potentially offer some insight into the reconstructed proto-lexicon, even if that particular example had some issues. Still, we can draw a distinction between determining the etymology of a word – identifying the root that underlies it – and studying the details of its derivation. That is, understanding a given item’s word-formation details, that is, looking at the precise derivational processes involved, can be helpful in developing a picture of the proto-language lexical stock. For instance, Latin *rāstrum* ‘drag-hoe’ derives from the verb *rādō* ‘scrape’ with the aforementioned **-tro-* suffix, but there is reason to believe that that suffix was “moribund in Latin” (Weiss 2009: 283), suggesting that this noun is an old form whose derivation can be projected back into Proto-Indo-European, or at least pre-Latin, despite its relative isolation within Indo-European.

Similarly, the noun **yugom* ‘yoke’, derived from **yeug-* ‘to yoke, to join’, is used in reference to yoking oxen to a plow, with widespread cognates across the family, including Sanskrit *yugam*, Hittite *iukan*, Greek ζυγόν (*zugón*), Lat. *iugum*, English *yoke*, Old Welsh *iou*. As a derivative, it would appear that **yugom* must be very old, as derivationally, it involves the formation of a thematic noun from a verbal root (**yeug-*, cf. Skt *yuj-*, Greek ζεύγ- (*zeúg-*), Latin *iung-*) by internal derivation – with zero-grade ablaut – and with no specialized suffix beyond the thematic vowel **-o-*.¹³ Indeed, Mallory and Adams (1997:655) include this noun as among the reconstructible Proto-Indo-European agricultural terminology, as does Kölligan 2017. However, even if to be posited as part of the Proto-Indo-European lexicon on the basis of its derivational pattern, the original sense could have been for yoking a team to a chariot, as suggested by Vedic Sanskrit terminology, and not for yoking a team to a plow.

Nonetheless, the methodological step of looking to the details of derivation and the processes involved – more a matter of Proto-Indo-European word-formation per se than just (root) etymology – shows promise as a type of lexical analysis, if the right words and the right manner of derivation are summoned forth. I offer here such a case in point, involving a Proto-Indo-European derivational process, namely reduplication, due to its possible involvement in terms for various items of agricultural relevance.

Drawing here on Joseph (1992), I suggest that reduplication as a morphological process employed in word-formation in Proto-Indo-European lies at the intersection of various Indo-European words for grain and for instruments, especially agricultural instruments. Such a nexus allows for the hypothesis that reduplication was specialized for use in Proto-Indo-European with agricultural terminology.

The relevant evidence comes out of a consideration of Hittite *memal* ‘grits, meal’ and Armenian *mamul* meaning ‘press, vice’. Both forms are built on the root **melH₂-* for ‘to mill, to grind’ (Rix & Kümmel 2001: s.v., ‘zerreiben, mahlen’), seen in Hittite *malli*, Latin *molō*, Old Irish *melid*, inter alia. Both show reduplication in their derivation, but they have different functions, different kinds of meaning related to milling. In particular, *memal* is a result noun, in particular referring to grain – grits or meal taken as the results of milling – whereas *mamul* is an instrument noun, a related kind of machine or tool.

Reduplication occurs across the Indo-European family in grain-related words and in Hittite and maybe elsewhere on several grain/agriculture-related instrument nouns. Regarding the former, grain-related words, there are the following to take note of:

13. The thematic vowel itself could well have had a semantic function in derivation but it more usually serves just a classificatory function as an indicator of a particular pattern of inflection.

1. Greek *παιπάλη* (*paipálē*) ‘finest meal’, with variant *πασπάλη* (*paspálē*),¹⁴ all of which are related within Greek to (derived from) *πάλλω* (*pállō*) / *παιπάλλω* (*paipállō*) ‘to quiver, to shake’,¹⁵ from the Proto-Indo-European root **pel(H)-* ‘to pour, to flow, to fill’.
2. Latin *furfur* ‘bran’, from a Proto-Indo-European root **gher-* ‘rub’, seen in Lithuanian *gūrti* ‘crumble’, and in the initial cluster of English *grind*.
3. Sanskrit *kiknasa-* ‘particles of ground corn’, most likely from a Proto-Indo-European root **knes-* ‘scratch’, an enlargement of **ken-*, as found in Greek *κνέωρος* (*knéōros*) ‘spurge-flax’; a possibly relevant form is *cikkasa-* ‘barley meal’, which appears to show reduplication, though its base root is uncertain.

Regarding the latter, instrument words, the following can be cited:

1. Armenian *mamul* ‘press, vice’, related within Armenian to the verbs *malem* ‘to smash, to crumble, to chop’ and *mlmlem* ‘to rub’, and the noun *mul-* ‘mill’, and outside Armenian to Old High German *muljan* ‘to smash, to crumble’, and Greek *μύλη* (*múlē*) ‘mill’, all from a Proto-Indo-European root **melH₂-* ‘to grind, to mill’ (and see above regarding *memal*).
2. Hittite ^{GIS}*šešarul*¹⁶ ‘sieve’ (with a related verb *šešarie-* ‘to sift’) < PIE **srew-* ‘to flow’ (an enlarged form of **ser-* ‘to flow’), with a ‘sieve’ representing the instrument through which a certain type of flowing, e.g. of grain, is accomplished.
3. Hittite ^{GIS}*ḫah(ha)r(a)-* ‘rake’ (with derived denominal verb *ḫahḫarie-* ‘to rake’) < **H₂erH₃-* ‘to plow, to break ground’ (so Tischler 1983: 122).

It may also be the case that the celebrated Proto-Indo-European word for ‘wheel’, **k^we-k^wl-o-*, belongs here too. Its reconstruction is guaranteed by the equation of Sanskrit *cakra-*, Greek *κύκλος* (*kúklos*), and Old English *hweo(wo)l*, and it derives from the root **k^wel-* ‘turn’. This noun can embody an instrument function, with a wheel being something by which turning is accomplished, perhaps originally ‘the turner’, as far as its meaning is concerned. Its Proto-Indo-European age is suggested also by the fact that it has an apparently archaic structure, with reduplication and

14. Greek also shows a synonymous nonreduplicated form *πάλη* (*pálē*).

15. This verb admittedly shows reduplication, but the reduplication here presumably reflects another cross-linguistically common function for this process, namely intermittent action.

16. The superscript element “^{GIS}” in this word and the next, here and throughout, indicates a Sumerian cuneiform logographic symbol (meaning ‘wood’, literally) that is used as a determiner of a class of noun, in this case instruments made with wood; the noun itself here is written out syllabically in Hittite (e.g., as *šešarul*). Such “Sumerograms” are frequent with certain words and are typically cited, as here, as part of the Hittite representation of the word even though they have no phonological relevance.

zero-grade, traits that individually but especially together are somewhat uncommon among Indo-European nominal forms. As an instrument, the wheel was surely materially involved in agriculture, as it provided the possibility of carts and wagons to haul the results of harvesting grain and other crops, as well as manure to be used as fertilizer.¹⁷

All that is seen here for the semantics and function of these reduplicated terms across Indo-European is consistent with cross-linguistic uses of reduplication, going with nouns for items taken in collectivity in many little bits and pieces, like grains, and for repeated actions (cf. Moravcsik 1978), so that the possibility of independent use of reduplication in each linguistic tradition cannot be dismissed. However, it can be speculated that reduplication is perhaps especially well suited as a derivational process with agricultural terms, since the actions involved in agriculture, including tilling, plowing, and sifting, require repeated actions in ways that the tasks involved in, say, animal husbandry, do not, and the results of agriculture, especially involving grains, lead to collections of multiple small items. If this is so, then we can say that even though **k^we-k^wl-o-* is not found in Anatolian ('wheel' is *hurki-*), the Proto-Indo-European agricultural instrument derivational process is present nonetheless via ^{GIS}*šešarul* 'sieve' and ^{GIS}*hah(ha)r(a)-* 'rake'.

It must be admitted, of course, that reduplication as a process has other functions in Proto-Indo-European, most notably the grammatical functions of being one of the distinctive marks of the perfect tense, as seen (with the reduplicative syllables in bold), e.g., in the equation of Greek *λέ-λοιπ-ε* (*lé-loip-e*) 's/he has left', Sanskrit *ri-rec-a*, from **le-loik^w-e*, and of being a key element in some present tense formations, as seen (ditto), e.g., in the equation of Greek *δι-δω-σι* (*dī-dō-si*) 's/he gives', Sanskrit *da-dā-ti*.¹⁸ And, it figures in the more lexical derivation of intensive stems, to judge from Sanskrit forms such as *jān-ghan-ti* 'he strikes repeatedly' (root *han-* from **g^when-* 'to strike') and parallel Greek forms like *παμ-φαίν-ει* (*pam-phain-ei*) 'it shines forth' (root *φαν-* (*phan-*) built on **bheH₂-* 'to shine'). Moreover, it is true as well that reduplication does not occur in all agricultural terms; indeed, some of the reconstructible words for grains and tools already discussed, e.g. **ieuo-* 'corn, barley, spelt' or **H₂erH₃tro-* 'plow' show no reduplication. Nonetheless, the clustering of reduplication in various terms for grain and instrumentation for grain and agricultural across the family is striking, and would go unnoticed without the

17. The wheel can of course be used in grinding grain but unfortunately there is no archaeological evidence suggesting that the Indo-Europeans used wheels in that way; that particular use seems to have been an invention in Hellenistic Greek times.

18. The difference in the reduplicative vowel – *a* in Sanskrit versus *i* in Greek – while a real issue to be tackled in reconstructing the details of present-stem reduplication, is irrelevant for the equatability of the stem-formation type.

impetus provided by lexical analysis of derivational patterns and their possible relation to a specific semantic class of words. The argument, then, from this observation, for agriculture in Proto-Indo-European would be that the specialization of a derivational process for use with agriculturally related terminology would only be possible in a society in which there was agriculture; that is, one needs to have the technology first within a society for there to be a derivational process specialized for vocabulary associated with that technology.

4. The lexicon of ritual

A further type of lexical analysis looks at the use of particular agricultural words in context. In particular, the language of Proto-Indo-European religious rituals and mythology gives evidence in them, as argued by Watkins (1978) in his discussion of “famous grains” of Proto-Indo-European, of the embedding of agricultural vocabulary into the religious practices and mythological tales associated with early Indo-European culture. This usage can be taken to demonstrate how ingrained (so to speak!) the practice of agriculture must have been for the Indo-Europeans if it is able to penetrate into their holiest and most sacred practices.

In particular, Watkins draws attention to a number of ways in which grains figure in references to rituals and myths associated with rituals in early Indo-European texts, especially Homeric Greek epic, the sacred Sanskrit hymns of the Rigveda (RV) and Atharvaveda (AV), and passages in the ancient Iranian language Avestan. I give here a sampling of the remarkable collection of relevant material that Watkins assembles in support of his hypothesis of the prominence of grains in Proto-Indo-European religious culture.

For instance, Watkins (1978: 10–14) notes what he refers to as “the solemn utterance ἄλφι καὶ ὕδωρ [(*álphi kai húdōr*)] ‘barley and water’ of the goddess of grain herself, in the Homeric Hymn to Demeter 208”.¹⁹ And, in the Atharvaveda, hymn 6.14, “*yáva* ‘barley’ is the addressee of a hymn” and is referred to as *devam* ‘divine’.²⁰ Watkins observes, concerning that hymn, that “agricultural *carmina* such as AV 6.14 are deeply rooted in the Indo-European tradition”. He further states that the combination of *yava-*, and its Avestan cognate counterpart *yauuō*, with the verb *karš-* ‘plow’, Avestan *karš-*, is a Common Indo-Iranian verb phrase, and its occurrence in “an important passage in the *Vidēvdāt* ... shows the religiosity

19. Greek ἄλφι (*álphi*) is cognate with Albanian *elb* ‘barley’ and some modern Iranian forms, e.g. Pashto *ōrbaše* (PL.) ‘barley’ (Mallory & Adams 1997: 51).

20. Sanskrit *yava-* is cognate with Greek ζειαί (*zeiaí*), Hittite *euwan*, Tocharian B *yap*.

of the cognate *yaauō* in Iranian”.²¹ He goes on to develop a line of argumentation showing that “not only has barley a genealogy, but also a mythology”. Among the myths associated with barley is that in RV 1.117.21 in which *yava-* is said to be spread by “the two *Aśvins* ploughing and sowing *with a wolf*” (so also RV 8.22.6); other animals are mentioned in connection with sowing barley in other passages: the bull (*vṛṣan-*), in RV 1.176.2, and cattle (*gav-*), in RV 1.23.15. Watkins elaborates on the role of grain, saying that “in the Indo-Iranian world barley has its place not only as a foodstuff, and not only in the cosmology and mythology, but also in cultic practice”. One finds the ritualistic mixing of barley with milk, a product of related agricultural practice, in both Indic and Iranian sources, and “roasted barleycorns ... are eaten by Indra as a *garnish* to the soma drink itself” as a part of the soma ritual. Importantly, Watkins finds parallel practices and phraseology to the grain-related aspects of the soma ritual in Homeric passages, e.g. in book 10 of the *Odyssey* (especially lines 233–236 and 316), where there is “the description of Circe’s magic potion that turns men into swine”. Finally, there is parallel in the mixing of barley and water (the ἄλφι καὶ ὑδωρ (*álphi kai húdōr*) cited above) in the Homeric Hymn to Demeter, about which Watkins opines: “There can be no doubt that we have an extremely archaic piece of traditional lore, both linguistically and thematically”. In summation, taking in parallels not discussed here, Watkins (1978: 17) offers the following particularly compelling statement:

My conclusion is dictated by the basic tenets of the comparative method: the soma ritual of Vedic and Indo-Iranian, by men for men, but symbolically by women; the ritual act of communion of the Eleusinian mysteries, by women for women; and a warrior ritual in archaic Greece, by women for men; all of these must go back to a single common Indo-European liturgical cultic practice. The number and the precision of the agreements between Indo-Iranian and Greek, and their articulation as a structure, a *total social fact*, are too striking for a fortuitous resemblance to be plausible.

The fact that grains and other agriculturally related entities are embedded in these cultic practices and religious rituals raises the question, hinted at in the beginning of this section, of how this mytho-religio-linguistic embedding could have occurred, that is, how such items – the lexemes and the real-world entities that they represent – could have become such an important part of this cultural context. The answer seems clear: it could only have happened if grains were a part of Proto-Indo-European culture already in the reconstructed proto-language, the language ancestral to Anatolian as well as Greek, and Italic, that is, “classical” Proto-Indo-European

21. The *Vidēvdāt* is a subpart of the Avestan corpus that deals with ways of counteracting evil demons.

(or Proto-Indo-Hittite, so to speak), and if they were a key part of life for the Indo-European speech community. The specificity of the parallels that Watkins notes, both as to practice and as to diction, is what – according to the dictates of the comparative method – allows one to locate grains chronologically in Proto-Indo-European society; it is difficult to suppose that such precise correspondences could have arisen independently in each branch. The examination of the context in which the relevant lexemes occur, then, in reconstructible Proto-Indo-European text and practice thus becomes a tool for learning about prehistoric agriculture as far as the Indo-Europeans were concerned.

5. Conclusion

Ultimately, then, my claim is that it is not enough to just look at the semantics of particular words and to try to develop a sense of what they originally meant (this type of grain or that, this type of fruit or that, etc.), nor is it enough to determine the source of the words (derivation, etymology, including borrowing). Rather, one also has to look at how the words were used, what details are reconstructible about the words, including the derivational processes involved in their formation, and the use of the words, including the cultural context in which they occur. If we are armed with such a fuller perspective, then the insights we derive from lexical examination that are used in developing a sense of prehistory can take on a greater degree of credibility.

It is important to realize that the extensions of previous lexically based methodologies advocated here may not be applicable in all cases or in all language families. With Indo-European, we are blessed with an abundance of ancient testimony to work with, and thus we can milk that material for all it is worth, so to speak. However, since part of the argumentation here comes from mythological and ritualistic uses of particular language, even cultures without a deep written history could have a deep oral tradition to draw on.²² The dimensions to lexical analysis discussed here, therefore, represent ways of getting more out of this material than a focus simply on vocabulary inspection or root derivations or etymology would allow.

22. In this regard, it is instructive to remember that although there is now a written tradition for the transmission of the Vedic hymns, for millennia they were – and still are, even now with written forms to work with – passed down orally.

Abbreviations

AV	Atharvaveda
(P)IE	(Proto)-Indo-European
RV	Rigveda

References

- Beekes, Robert S. P. 1995. *Comparative Indo-European Linguistics. An Introduction*. Amsterdam: John Benjamins. doi:10.1075/z.72
- Fortson, IV, Benjamin, W. 2010. *Indo-European Language and Culture. An Introduction*, 2nd edn. Oxford: Wiley-Blackwell.
- Joseph, Brian D. 1985. Proto-Indo-European consonantism: Methodological and further typological concerns. In *Papers from the 6th International Conference on Historical Linguistics* [Current Issues in Linguistic Theory 34], Jacek Fisiak (ed.), 313–321. Amsterdam: John Benjamins.
- Joseph, Brian D. 1992. On some Classical Armenian reduplicative nouns: *Mamul*, *mamur*, and *mamuR*. In *Proceedings of the 4rd International Conference on Armenian Linguistics*, John Greppin (ed.), 101–114. Delmar NY: Caravan Books.
- Kölligan, Daniel. 2017. The lexicon of Proto-Indo-European. In *Handbook of Comparative and Historical Indo-European Linguistics*, Jared Klein, Brian Joseph & Matthias Fritz (eds). Berlin: De Gruyter Mouton.
- Mallory, James P. & Adams, Douglas Q. 1997. *Encyclopedia of Indo-European Culture*. London: Fitzroy Dearborn.
- Masica, Colin P. 1979. Aryan and non-Aryan elements in North Indian agriculture. In *Aryan and Non-Aryan in India*, Madhav M. Deshpande & Peter Edwin Hook (eds), 69–70. Ann Arbor MI: Center for South and Southeast Asian Studies.
- Meillet, Antoine. 1964[1922]. *Introduction à l'étude des Langues Indo-Européennes*. University AL: University of Alabama Press.
- Monier-Williams, Sir Monier. 1899. *A Sanskrit-English Dictionary: Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Oxford: OUP.
- Moravcsik, Edith A. 1978. Reduplicative constructions. In *Universals of Human Language*, Vol. 3, Joseph H. Greenberg (ed.), 297–334. Stanford CA: Stanford University Press.
- Pokorny, Julius. 1959. *Indogermanisches Etymologisches Wörterbuch*. Bern: Francke.
- Rix, Helmut & Kümmel, Martin. 2001[1998]. *Lexikon der Indogermanischen Verben*, 2: Erweiterte und Verbesserte Auflage. Wiesbaden: Reichert.
- Tischler, Johann. 1983. *Hethitisches etymologisches Glossar, Teil I: (a-k)*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Watkins, Calvert W. 1978. Let us now praise famous grains. *Proceedings of the American Philosophical Society* 122 (1): 9–17.
- Weiss, Michael. 2009. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor MI: Beech Stave Press.

Agricultural terms in Indo-Iranian

Martin Joachim Kümmel

Friedrich-Schiller-Universität Jena

The article investigates the agricultural lexicon of Indo-Iranian, especially its earlier records, and what it may tell us about the spread of farming. After some general remarks on “Neolithic” vocabulary, a short overview of the animal husbandry terminology shows that this field of vocabulary was evidently well-established in Proto-Indo-Iranian, with many cognate terms. Words for cattle, horses, sheep and goats are well developed and mostly inherited, while evidence for pigs is more limited, and the words for donkey and camel look like common loans. A more extensive discussion of plant terminology reveals that while some generic terms for grain are inherited, more specific words for different kinds of cereals show few inherited terms and/or irregular variation, and the same is even clearer for pulses and some other vegetables. The terminology for agricultural terminology is largely different from that of most European branches of Indo-European. The conclusion is that the cultural background behind these linguistic data points to spreading of a mainly pastoralist culture in the case of Indo-Iranian.

Keywords: Indo-Iranian, husbandry terms, plant cultivation, agricultural technology, pastoralist

1. Introduction

Indo-Iranian (II) is the major Southeastern branch of Indo-European. In antiquity, their territory covered much of the Western steppe, Western Central Asia, most of the North of the Indian subcontinent and most of the lands to the West of it, until Eastern Anatolia. Indo-Iranian consists of two main subbranches, Indo-Aryan and Iranian. Some intermediate modern languages in Nuristan in present-day Afghanistan appear to have separated from one of the main subbranches very early, so that they practically constitute a third subbranch, Nuristanic. The language family is attested since around 1400 BCE, when some words of Indo-Iranian origin appeared in sources of the Near Eastern Hurrian state of Mittani, in a language

very close to Old Indo-Aryan as attested in India by the Vedic texts from ca. 2200–1500 BCE (orally transmitted until much later), composed in a language known as Sanskrit which has remained a classical literary language until today (here I will use *Vedic* for the earlier stage and use *Sanskrit* only for post-Vedic times). The first Old Iranian texts are roughly as old: the corpus of Avestan is contemporary to Vedic, and the Old Persian inscriptions date from 525–300 BCE. These early texts of both branches show languages that are grammatically extremely similar. From them and later data, a not too far common protolanguage can be reconstructed, Proto-Indo-Iranian (PII), probably spoken around 2500–2000 BCE somewhere in Western Central Asia. After the oldest period, a variety of Middle Iranian¹ and Middle Indo-Aryan² languages are attested, but many modern languages have no attested direct ancestors. It is most often assumed that Indo-Iranian had its origins in the western or central Eurasian steppe and then spread east and south (cf. Kuz'mina 2007; Parpola 2012), building on a primarily pastoralist economy. However, if an Anatolian homeland of Proto-Indo-European is assumed, Indo-Iranian may also have had its origin south of the Caucasus and then spread east together with agriculture before it spread northward into the steppe zone. Both scenarios should be distinguishable in the lexicon of Proto-Indo-Iranian and early Indo-Iranian. In the first case, the pastoralist terminology would be expected to be more stable and easier to reconstruct, i.e. mainly terms for domestic animals and their products. In the second case, an ancient and rather stable terminology of plant cultivation should be easier to reconstruct, i.e. terms for the most important crops and the relevant technology (esp. ploughing).

The present article investigates the agricultural lexicon of Indo-Iranian, especially its earlier records, and what it may tell us about the spread of farming. The primary data are taken from the usual dictionaries, most notably Mayrhofer (1992–2001); Bailey (1979); Bartholomae (1904); Morgenstierne (1929, 1938, 1974, 2003); Abaev (1979); Rastorgueva & Edelman (2000–2007); Edelman (2011). Data from later languages are normally only adduced if there is no attestation in Old Iranian or Old Indo-Aryan, or if they provide additional information.

1. “Western” Middle Iranian languages are represented by Middle Persian, Parthian, and a particular early stage attested by loanwords in Armenian; the “Eastern” group comprises Alanic, Xwarezmian, Sogdian, Bactrian, and Saka (Khotanese and Tumshuqese).

2. The most important representative of the oldest stage of Middle Indo-Aryan is Pāli, the language of the Buddhist Theravada canon; the second stage is represented by the so-called Prakṛts, the best attested being Ardhamagadhi, used for the large corpus of Jaina texts.

2. “Neolithic” vocabulary

2.1 General remarks

The lexicon of Indo-Iranian, like that of Indo-European in general, presupposes a Neolithic stage of cultural development. The terminology for animal husbandry and pastoralism is well developed and easily reconstructible. Terminology for different aspects of plant cultivation is also present, including terms for grain, pulses, and technology such as ploughing – however, it is more difficult to reconstruct, as we shall see. While it is disputed if wheeled vehicles can be assumed for the Proto-Indo-European level, there is no questions that they were known already in Proto-Indo-Iranian, including the chariot (PII **rátha-*).

In the following, a short overview about animal husbandry is presented first, before a more detailed treatment of plants and plant cultivation is given.

2.2 Terms for domestic animals

As already mentioned, this semantic field is well attested and contains many assured Proto-Indo-Iranian terms with rather fine-grained distinctions. This is valid for cattle, horses and sheep, while terms for goats are already a little more varied, and the words for the “southern” animals, donkey and camel, seem to be loanwords. For ‘pig’, the evidence points to a rather marginal role.

(1) ‘Cattle’ *Bos (primigenius) taurus*

Generic (epicene) PII **gʷ- f./m.* ‘cow, bull, ox’: Av. *gāuu-* = Ved. *gāv-*; < PIE **gʷ- ów-*

Young: PII **watsá-* m. ‘calf’: Clr. **wasa-* = Ved. *vatsá-*; < PIE **wets-ó-* ‘belonging to the (current) year’ (cf. Vine 2009: 213–8)

Female: PII **(H)ajʰi-* f. ‘cow’: Av. *azī-* = Ved. *ahí-*;

PII **dʰa(H)inú-* f. ‘milch cow’: Av. *daēnu-* = Ved. *dhenú-*

Young: PII **wacá-* f. ‘young cow’: Ved. *vaśá-*; < PIE?, cf. Lat. *uacca*;

?PII **gr̥stí-* > Ved. *gr̥stí-* f. ‘young cow’³

3. This word has no cognates outside Indo-Aryan but looks like an ancient formation.

Male: PII *wṛśān- m. ‘male, bull, stallion’: Clr. *wəršan- = Ved. vṛśan-; < PIE *wérson-;

PII *rśān- m. ‘male, bull, stallion’: Av. aršan-, cf. Ved. ṛṣa-bhá-; < PIE *h₁rśén-

Young: PII *hukśān-⁴ m. ‘young bull’: Av. uxšān- = Ved. ukśán-; < PIE *h₁uksén-

(2) ‘Horse’ *Equus ferus caballus*

Generic and male: PII *ácwa- m. ‘horse’ > Av. aspa- = Ved. áśva-; < PIE *h₁ékw(o)-

Young: unclear; Persian *kurna- ≠ Ved. kiśorá- ‘foal’

Female: PII *ácwā- f. ‘mare’ > Av. aspā- = Ved. áśvā-; < (P)IE *h₁ékwah₂-

Male: PII *ácwa-; PII *wṛśān-, PII *rśān- (as for ‘bull’, see above)

In addition, there are other, more poetic terms like ‘runner’.

(3) ‘Sheep’ *Ovis (aries) aries*

Generic and female: PII *háwi- f. ‘sheep, ewe’ > Ved. ávi-; < PIE *h₂ówi-/h₂awi-

Young: PII *wár(h)an- m. ‘lamb’ > Clr. *waran- = Ved. úran-; < PIE *wṛh₁én-

Female: PII *mayšī- f. ‘ewe’ > Av. maēšī- = Ved. mešī-, derived from *mayśá- m.

Male: PII *mayśá- m. ‘ram’ > Av. maēša- = Ved. meśá-; < PIE *mojsó-;

PII *wṛšni- m. ‘ram’ > Av. varšni- = Ved. vṛṣni-, derived from *wṛśān-

(4) ‘Goat’ *Capra aegagrus hircus*

Generic and male: PII *hájá- m. ‘goat’ > Av. aza-⁵ = Ved. ajá-; < (P)IE *h₂aǵó-;

PII *b^húja- m. ‘goat’ > Av. būza-; < (P)IE *b^huǵo-

Generic and female (P)II *hájá- f. ‘she-goat’ > Ved. ajá-;

(P)II *b^hújá- f. ‘she-goat’ > Clr. *buzā-; both derived from masc.

PII *scagā- f. ‘(she-)goat’ > Clr. *sagā- > Oss. D. sæγæ (cf. Ved. chagalá- ‘kid’, chāgā- ‘she-goat’)⁶

Young: PII *skáni- m.? > Av. scaini- ‘kid’; < PIE *(s)ken(H)-

?Male (P)II *pazdá- > Ved. bastá- m. ‘he-goat’

4. Since there is some evidence that *h₂ was partly preserved as *h in Proto-Indo-Iranian and even in Proto-Iranian (see Kümmel 2016: 81–83), I reconstruct PII *h- in words with PIE *h₂-, although there might be no Inner-Indo-Iranian evidence for *h- (e.g., no continuants with Persian h-/x-).

5. Common Iranian aza- and/or *azā- was also borrowed into Tocharian, cf. Toch. A ās ‘(she-)goat’, derivative Toch. B ašīye ‘goat’s’ = A āšī* (Pinault 1997: 200–204; Adams 1999: 32; Carling 72–73).

6. Cf. Lubotsky (2001: 309, 312); also LW in Uralic, cf. Rédei (1986: 59).

(5) ‘Pig’ *Sus scrofa*

Generic and female: PII **súH-* F. ‘pig, sow’ > Av. *huš*; < PIE **súH-*
 derivative PII **suHka-* m. > Khot. *hva-*, MP. *hwg*; cf. Ved. *sūka-rá-* m.
 ‘(wild) pig, boar’

Male and wild: PII **warā́j⁶á-* m. ‘boar’ (LW)⁷ > Av. *varāza-* = Ved.
varāhá-; also loanwords in Uralic⁸

Young: PII **parća-* m. ‘pig(let)’ > Av. **par²sa-*, Khot. *pāsa-*; also loanwords in
 Uralic;⁹ < PIE **pórko-*

Pigs were apparently not important in early Indo-Iranian culture, and the words may mostly refer to the wild boar.

(6) ‘Donkey’ *Equus africanus asinus*

Generic and male PII **khará-* m. ‘donkey’ > Av. *xara-* = Ved. (AV+) *khara-*
 (probably a loanword, cf. Lubotsky 2001: 311)

Ved. *gardabhá-* m.; *rāsabha-* m. ‘donkey’

Female derived: Av. *xarā-* f.; Ved. (AV+) *gardabhī-* f.; Skt. also *khari-*;
rāsabhī-

→ ‘Mule’: derivatives of ‘horse’ or ‘donkey’:

PII (?) **aćwa-tara-* m. ‘mule’ > Clr. **aswatara-* > Bactr. *aspodaro* = Ved.
 (AV+) *aśvatará-*

Plr. **khara-tara-* m. ‘mule’ > Clr. **xaratara-* > Khot. *khaḍara-*

Female derived: Ved. (AV+) *aśvatarī-*

(7) ‘Camel’ *Camelus bactrianus/dromedarius*

Generic and male PII **húštra-* m. ‘camel’¹⁰ > Av. *uštra-* = Ved. *úṣṭra-*

Female derived: Av. *uštrā-* F. ≈ (Late) Ved. *uṣṭrī-* f.

There is also a rather well-developed terminology for animal products, but this cannot be treated here (for ‘milk’ cf. the contribution by Garnier & Sagart in this volume).

7. Cf. Lubotsky (2001: 309, 312).

8. Cf. Rédei (1986: 54; 1988: 720); Zhivlov (2014: 139).

9. Cf. Bailey (1979: 235); Katz (2003: 205–206). However, Hyllested (2017: 192–193) has argued that these words may ultimately represent loans from Turkic (into different Indo-European and Uralic languages), so that the apparently perfect equation between Avestan, Khotanese and the European languages would be a mirage.

10. This word is probably a loanword from a Central Asian contact language; the reconstruction with **h-* is motivated by the Iranian personal name **jarat-huštra-* > **jaraθuštra-* > Avestan *Zaraθuštra-* etc. (cf. Lubotsky 2001: 313; Kümmel 2016: 82).

2.3 Agricultural plants

2.3.1 Cereals

A recent comprehensive treatment for all of Indo-European is Witczak (2003), although it is overconfident in reconstructing inherited words; an English summary and valuable comments are found in the review by Blažek (2005).

1. General terms

The more general terms appear to be inherited Indo-European words. The most central term appears to be PII **yáwa-* ‘grain, cereals, barley’ > Av. *yauua-* = Ved. *yáva-* < PIE **jéwo-*, cf. Hitt. *ewa-*, Lith. *jawai* etc.,¹¹ also borrowed into Uralic (**jewá* > Finn. *jyvä* etc.);¹² also in Clr. **yawa-arta(ka)-* ‘grain-flour’ = ‘grain, cereals’ > Parthian *ywrđw*, Bactrian *iaoapδaoi* etc. (cf. also Blažek 2017: 54–5). A more specialized word is PII **d^hānā-* ‘(roasted) seed, grain’ > Av. *dānā-* = Ved. *dhānā-* < PIE **d^hoHnā₂-*, cf. Lith. *dúona* ‘bread’.¹³ A more general term was PII **sasyá-* ‘crop, fruit’ > Av. *hahiiia-* = Ved. *sasyá-*, derived from **sasá-* > Ved. *sasá-* ‘crops, food’, to be connected with Hittite *sēsa-* ‘fruit’, *sēsann-* ‘fruit tree’ and Brythonic **sasjo-* > Welsh *haidd*, Breton *heiz* etc. ‘barley’.¹⁴ Iranian also has another word, PIr. **ādu-* ‘grain, corn’ > Av. *ādu-*, Sogd. *δwkh*, *δwk*.¹⁵ It could be derived either from **ad-* ‘to eat’ (PIE **h₂ed-*, otherwise not well attested in Iranian), but more probably it should be compared to other words for ‘grain’ like Arm. *hat* ‘grain, seed’, Lat. *ador* ‘barley, spelt’, probably derived from PIE **h₂ad-* ‘to dry’.¹⁶

2. ‘Barley’ *Hordeum vulgare*

Most frequently the generic term PII **yáwa-* < PIE **jéwo-* is used more specifically for ‘barley’, especially in Western Iranian, Indo-Aryan and Southern Nuristanic.

11. Blažek (2017) reconstructs **iéuho-* and connects the word with a root *‘to ripen’. This is possible, although there is no concrete evidence for a laryngeal in this word. In any case, it has to be separated from PII **HyawH-* ‘to eat, consume’ and words for ‘pasture’, see Nikolaev (2014).

12. Cf. Katz (2003: 212–3, 334); Blažek (2017: 55).

13. Tocharian B *tāno*, *tām* ‘grain, seed’ < **tanā-* may be a cognate or a loan from Iranian. Cf. Hock et al. (2015: 245–6).

14. Cf. Witczak (2003: 41–2); Blažek (2005: 222). The Indo-Iranian word was probably the source of Udmurt *šežj* ‘oats’, cf. Katz (2003: 215).

15. Old Persian **ādu-* was also borrowed into Elamite as *ha-du-iš* ‘revenue, yield, increase’, cf. Henkelman (2010: 737–738).

16. Cf. Emmerick (1966); Szemerényi (1969); Hamp (1973); Blažek (2005: 219); Rossi (2010); Blažek (2017: 56).

Beside **yawa-*, Iranian has also other words. The most widespread is Clr. **kas(a)ka-* (cf. Edelman 2011: 337–338):¹⁷ NP. *kašk*, Pam. Munj. *kosk*, Šu. + *čūšj/čūšč*, Yazg. *kusk*, with a slightly deviant variant in Khot. *chaska-*. Blažek (2005: 221; 2017: 55–56), following Pachalina (1983: 115), connects these words with Common Slavic **kolsъ*, Albanian *kallë* ‘ear’ and assumes a preform **karška-/křšaka-*, but this is not compatible with *sk* in most Iranian languages.¹⁸ This connection is therefore rather improbable, and it would be much better to assume a connection to words for ‘grass, straw, millet’ going back to PII **kāća* (cf. below). In Eastern Iranian, some more words are attested. One of them is Clr. **rucā-* > Khot. *rrusā-*, Xwar. *rsy*. It was connected by Bailey (1979: 367) to some other words found in Modern Eastern Iranian, but despite a certain similarity they cannot be cognate, since these words point to something like Clr. **arpucyā-*: Pto. *orb’əša*, Pam. Išk. *urvbš* (and Yidya *yeršio?*); maybe Turkic *arpa* has to connected somehow (cf. Blažek 2017: 53)? Derivation from (European) IE **alb^hi-* (Greek *álphi* etc.) is also impossible (pace Blažek 2005: 219; Witzel 2009; 2017), since the Pto. cluster *rb* can only go back to **rp*. Northern Nuristanic uses **wriji-* which elsewhere only means ‘rice’. Northwest Indo-Aryan has also **sitiya-* in Kho. *sirī*, Kal. *šilī*, apparently derived from *sītā-* ‘furrow’.¹⁹ Khwar *blan* and some Nuristanic words point to a base **brā(k)-*.²⁰

3. ‘Wheat’ *Triticum sp.*

In this meaning, there apparently was only one word, but it shows irregular variation. The Iranian words appear to presuppose Clr. **gantūma-* or **gandūma-*, but do not agree whether the third consonant is **t* or **d* and whether the *u* was lengthened or not:²¹ Av. *gan̥tuma-* < **gantūma-*; Pto. *yan’əm* < **ganTūma-*; Parth. *gndm* < **ganTuma-*; Waxi *ǰədim* < **ganTūma-*; MP. *gnm* < **ganduma-*; Khot. *ganama-* <

17. The cluster *šj* in Šuyñi (if correctly described) cannot go back to old **sk* or **šk* which would have remained voiceless; *j* thus presupposes a syncopated vowel, ergo **kasaka-*; Persian *šk* may be regular also from post-syncopate *sk*. If *šj* were not old, a preform **kaska-* is also possible, but **s* < PIr. **(s)c* remains assured and **rš* is excluded.

18. The alleged Tocharian cognate B *klese* (A **klas* maybe as a loan in Old Chinese **k’ras*, cf. B-S 51) would also show an irregular correspondence. There is also a vaguely similar word in Uralic: **čaši* ‘barley’, attested in Mordva, Mari and Permian (cf. Zhivlov 2014: 130).

19. The similarity to Greek *sītos* ‘corn, grain’ cannot be due to inheritance, cf. Blažek 55.

20. These are connected by Blažek (2017: 60–63) to other Indo-European words with **m(V)r^o*, especially Celtic **mraki-* ‘barley, malt’ > *mraich*, Welsh *brag* and HLuw. *marwalli-*, but they do not really match. However, there may be a connection to similar words in different Asian languages, as mentioned by Blažek.

21. The possible preforms are shown in the following table:

gandūma-*; Bal. *gandūm* < **gandūma-*. Indo-Aryan and Nuristanic rather point to **gawd^hūma-* (folk-etymological ‘cow-smoke?’) > Ved.+ *godhūma-* (with variants **gadhūma-*, **gedhūma-*). A shorter stem is probably found in Yazy. *γ^wont* ‘roasted and dried wheat’ < **gantū-* (gunt-*), where the preservation of *t* is unexpected. The base **gant-/gand-* can possibly be compared to Hittite *kant-* ‘emmer, wheat’ etc., but it is unclear whether this is due to inheritance or a parallel loan. The amount of irregular phonological variation may point to parallel loans (or a *wanderwort*).

4. ‘Oats’ *Avena sativa*

There is little data on this type of grain, since oats are not much cultivated in the regions of Indo-Iranian. Khotanese *hau* ‘a sort of grain, oats?’ (Bailey 1979: 497) might continue *(*h*)*awiš-*, possibly the base for Clr. *(*h*)*aw(V)ša(ka)-* ‘ear (of corn), awn; Spica’,²² cf. Bal. *mazan-hōš* ‘with large awns’ and MP. p. *hwšk*, m. *hwšg/hōšag*/, Classical NP. *xōša*; Xwar. *wwfyk*; Kurd. *ûšî*; Pto. *w’azay*. This word may be connected to Proto-Slavic **awisa-*, Proto-Baltic **awižā-* and Latin *auēna* ‘oats’, although there is no regular correspondence; another irregular correspondence of these words may be **wis-* in Yazg. *wis* ‘avena’, Taj. Wj. *gis* (cf. Blažek 2005: 220).²³ In most Pamir languages, we find a different term derived from Clr. **dāsi-* > Pam. Šu. *δēsak* ‘oats’; Munj. *lisok*; > Išk. *δvsín*, cf. also Waxi *δosn* ‘Setaria’ (cf. Steblin-Kamenskij 1999: 165). Northern varieties of European Romani (an Indo-Aryan language) use *job*, a continuant of the old term **yawa-* (normally ‘barley’). At the Northwestern fringe of Iranian, still another, possibly old term is attested: The Jassic word list from Hungary has *zabar /saβar/* ‘oats’ which would correspond to Ossetic **sævæŕ* (Abaev 1979: 306) < Alanic **sab(a)ra-* < Clr. **sap(a)ra-* < PII **čap(a)ra-*, and this word could be compared to Germanic **habran-* < **ka/opró-* + *-n-*.²⁴ Possibly this

	nt	nd	uH	U
Avestan	+		+	+
Wakhi	+	+	+	
Sogdian; Pashto; Šuyni	+	+	+	+
Northwestern; Xwarezmian +	+	+		+
Balochi		+	+	
Khotanese		+	+	+
Middle Persian, Larestani		+		+
Yazyulami	+?		+	+

22. Cf. Rastorgueva & Edelman (2000: 269–70).

23. Starostin (1988: 121) compared some Northern Caucasian words pointing to **HVbVgV* and assumes borrowing from a substrate.

24. Cf. Blažek (2005: 222) with references.

belongs to some words for ‘green, vegetables’, cf. PIr. **capá-* ‘green plants, grass’ > Pto. *sāb’ā*, *sab’o* PL. ‘vegetables’, Yid. *səwī*; PIr. **capačV-* ‘green’ / **cāpačī-* ‘green’ > MP. p. *spz*, *sbz* /*sabz*/, NP. *sabz*; Šu. *sēpc*, Roš. *sēpc*. They can be compared to Ved. *śāpa-* ‘drifting wood’, Lith. *šāpas* (2) ‘straw, twig’ < IE **kópo-*, but details are not very clear.²⁵

5. ‘Rye’ *Secale cereale*

This northern grain was not normally cultivated, and there is no older term. According to Blažek (2005:222), **rugi-* > Clr. **rujika-* is found in Pam. Šu. Bajui *rōž*, Roš. *rūz* ‘ear (of grain)’ and can be compared to Germanic, Balto-Slavic **rug^hi-* ‘rye’ (cf. Hock et al. 2015:876–7). But this is impossible: Šu. *ō* goes back to **ā/a* vs. *u/a* < **u*, and the cluster *ž*/*wz* can only continue **rz* (cf. Pam. Šu. *vūžj*, Roš. *vūz* < **bərzu-* ‘high’) or maybe also < **zVk* but not **jVk*. These Pamir Iranian words must rather be reconstructed as **rārza-* and may then be related to **rarz-* ‘to tremble’ > Šu. *rāžj-*, as proposed by Morgenstierne (1974:67) under the assumption that the expression originally referred to stalks and ears waving in the wind.

6. ‘Rice’ *Oryza sativa*

There is a widespread term in PII **wri^hí-* > OP. **vrīzi-*; Pto. *wri^hže* F.PL. = Ved. + *wri^hi-* m.; Pa. Pk. *vīhi-* ... (→ LW Par. *rahō*), and Iranian variants with identical root are also found: **wrijana-* > Orm. *rīzan*; **wrijaka-* > Sogd. sm. *rysk*, b. *ryžkh*; **wrijuka-* > Khot. *rrīysū*. However, there also is a more distantly related variant in Clr. **bringá-* > MP. p. *blnc*, m. *brynz* /*brinz*/, Sogd. *βrync* /*vrinč*/, Arm. *brinj*; NP. *birinǰ*, Kurd. *birinc*, and further variants in Tal. *birz*, Siv. *birji*. Similar words apparently also were the source of Western loans like Greek *óryza/óryzon* etc., and it seems obvious that what we find here is an ancient *wanderwort*. In the East, we also find a different term for ‘unhusked rice’: Skt. ep.+ *śāli-* etc., Nur. km. *šāl’i*, borrowed in Eastern Iranian: Pto. *šole* PL.; Orm. *šōl*, Par. *šēl*, YM. *šālē*.

7. ‘Broomcorn millet’ *Panicum miliaceum*

Most Iranian languages have a common word here, and this is also found in parts of Nuristanic: Clr. + Nur. *(*h*)*arjaná-*: (Parth. >) MP. p. *’rzn*, NP. *arzan*, Pto. *zđən*, Wx. *yirzn* etc.; Nur. v. *üj’ü*, a. *az’ü*, w. *žzü*. In the Southwest, the word has been contaminated with **ganduma-* ‘wheat’ and yielded OP. *(*h*)*arduma-* > MP. p. *’lwm*, Baxt. *halum*. Vedic and Indo-Aryan have a word *añu-* (> Nur. *añ-*) that looks like *añú-* ‘thin’; there may be an indirect connection to the Iranian word, if it goes back

25. Cf. Mayrhofer (1996:629); Hock et al. (2015:1010). Bailey (1979:355, 407–8, 419–20) also mentions four further Khotanese words, all with irregular sound correspondences: *ysba* ‘cane, reed’ (Tib. *spa!*); *šavara-* ‘green plant’; *sapala-* ‘green stuff (in the crop of a bird)’.

to something like **arjnu-*. Nuristanic also has a different word **lāwa-*: km. kt. *řov*, w. *lav*. Another species is apparently named as ‘black’: Ved.+ *śyāmāka-* ‘Echinochloa frumentacea’. In any case, no clearly inherited terms are found.

8. ‘Foxtail millet’ *Setaria (italica)*

Again, Iranian has one common word not found anywhere else: Clr. **gāwərca-*: OP. **gāwərθa-*, MP. p. *g(ʾ)hl / gʷls*; NP. *gāl / gāwars* ... Pto. *γošt*; Khot. *gauśā* ‘*Setaria (viridis)*’. The beginning of the word reminds of **gāw-* ‘cattle’, but its formation is unclear. In contrast to that, Nuristanic has **kāca-*: km. *kačo*, kt. *kčo*, w. *kač* ‘*Setaria italica*’, comparable with MP., NP. *kāh* ‘straw’ and Ved. *kāśa-* ‘*Saccharum spontaneum (a grass)*’ < PII **kāca-*; possibly also Clr. **kasaka-* ‘barley’ (see above) may belong here.²⁶ A completely different word is used in Indo-Aryan: Ved. *priyāngu-* ‘*Setaria italica*’ etc.; Pam. Šu. *pīnʃ* etc. are loanwords from Indo-Aryan, not inherited (the similarity to Latin *pānicum* ‘*Setaria italica*’ must be accidental). Later we also find Skt. *kañku-*, *kañgu-* ‘id.’.

2.3.2 Pulses

1. ‘Bean’ *Vicia faba* (Europe); ‘Vetch’ *Vicia ervilia*, *Vicia sativa* and ‘Mung bean’ *Vigna radiata* (< Iran); ‘Black bean’ *Vigna mungo* (< India)

The most common word has the shape CII **māša-* with a clearly non-IE structure: Cf. MP. p. *mš* ‘vetch?’, NP. *māš* ‘*vigna radiata*, pulse’; (Nur. km. *mōš* LW?); Ved.+ *māša-* ‘*vigna mungo*’ (not apt for sacrifice). The word was apparently borrowed into Toch. B *māśak*, *māšikāni* and also into Arabic *māš*. Slightly similar is PIr. **mušā-*, **mušakā-*?, only found in Eastern languages, cf. Shgr.+ *maš*; Sogd. *mwwškh*, Yagn. *mušk*, possibly an independent loan.²⁷ Other terms of unclear origin are the following: Ved.+ *mudgá-* ‘*vigna radiata*’; Ved. *khálva-* ‘*vigna radiata*?’ AV+, *khárva-* MS; Ved. *garmút-* 𐎧 ‘wild beans’ YV+, and another Eastern Iranian word **sraxā-*: Xufi + *šāš*, Yazg. *šāx* (Wx. *šax* ‘pea’).

2. ‘Lentil’ *Lens culinaris*

For this meaning we find a widespread group of words with vaguely similar shape, possibly reflecting parallel loans: Clr. **nazyuka-* (?); Khot. *nīysva* PL.; NP. *nask*, Kurd. *nīsk*; Clr. **mižuka-*: MP. p. *mycwk*!, *myšwk*! /mižüg/, NP. *mižū?* (cf. **māša-*, **mušā-*?), and Pre-IA **masuHra-* > Ved.+ *masúra-*. In Western Iranian, also a quite

26. Beside **kāca-*, **kača-* there was also **čāka-* in Ved. (late) *śāka-* ‘herb, vegetables’, probably < **kéh₁ko-*, cf. Lith. *šėkas* ‘grass, hay’, Latv. *šēks* (although Prus. *schoks* is problematic), cf. Hock et al. (2015:1017).

27. Cf. Lubotsky (2001:315).

different word is attested: Clr. **winuka-* > MP. p. *wynwk* /winūg/ ‘lentil’, NP. *bunū* ‘pulse’ (Khot. LW *viṇakā* ‘chickpea?’).

3. For different types of ‘peas’, there is a variety of different terms, partly overlapping with those for other pulses
‘Chickpea’ *Cicer arietinum* and ‘Pea’ *Pisum sativum*

Clr. **nahwata-* < **nasw-ata-* > MP. p. *nhwt* /noxūd/, NP. *nuxūd* ‘chickpea, pea’; Skt.+ *caṇa(ka)-* ‘chickpea’; Clr. **sardaiči-* (?) only in Khot. *salicā* ‘pea’; Ved. (YV)+ *satínā-*; Skt. *hareṇu-*; Yid. *xur-muyo* ‘pea’.

2.3.3 Some other vegetables

1. ‘Onion’ *Allium cepa* and ‘Leek’ *Allium ampeloprasum*

The most widespread Iranian term is PIr. **piyāwa-(ka-/ča-)* ‘onion’: Khot. *pau*, Sogd. *pyk*; Yidga *pīy*, Yazg. *piyeḡ*; NP. *pīyāz*, Kurd. *pīvaz* (> Bal. *pīmāz*). Another word is only attested in the West: PIr. **cauxa-* ‘onion’: NP. *sōx*, Arm. LW *sox*; it looks similar to Turkic **soyan*. In Indo-Aryan, we find the foreign-looking word Ved. *palāṇḍu-* DhS+ ‘onion’. For ‘leek’, there is an Iranian term without an etymology: PIr. **kabardā-*: NP. *kavār*; Sogd. *kβrδh* /kəvarδ/, connected by Bailey (1979: 137) to Khot. *tcahai* ‘leek’ by irregular changes.

2. ‘Garlic’ *Allium sativum*

There is a variety of Iranian words for ‘garlic’: OP. **θigra-*, NP. *sīr* (> Kurd. *sīr*) would point to PIr. **cigra-*, but Lubotsky 2002 has explained it as borrowing from Scythian **tsigra-* < PIr. **tigra-* ‘sharp’. In the East, there is another, unclear word PIr. **bərjnā-*: Sogd. *βzny*, Pto. *uza*, Wan. *m’urža*, Yid. *wiznu* (cf. Rastorgueva & Edelman 2003: 126). An obvious and well motivated innovation is found in Par. *bīn*; Oss. D. *bodæn*, going back to Clr. **baudana-* ‘smelling’. Khot. *ysam̐basta-* is isolated.²⁸ In Indo-Aryan, the usual word is Skt. *lašuna-* etc., the origin of which is unknown.

28. Bailey (1979: 346) assumes **zamb(a)-* ‘cleft, yawn’ + possessive *-sta-*, but the preservation of *mb* is unexpected (cf. *ysimā* ‘teeth’ < **zambya-*); alternatively, it may be connected to **zam-* ‘earth’, but the formation remains unclear.

2.4 Agricultural technology

1. 'Field, cultivated plants'

PII. **hər(H)wārā-* f. > PIr. **hərwarā-*:²⁹ Av. *uruuarā-* f. 'plant, plants'; MP. p. *'wlwl*, m. *'wrwr* /urwar/; Parth. *'wrwr* /urwar/, PL. *'rwr-n*; Sogd. *(r)rwr(h)*, *rwr'* /ərwar-a/ 'healing plant' = Ved. *urvārā-* 'field, seed field, cornfield', cf. Greek *ároura*, Myc. *a-ro-u-ra* 'cornfield', Old/Middle Irish *arbor*, *arbe* 'grain, corn'.³⁰ Traditionally, this word has been derived from **h₂arh₃-* 'to plough', but the semantics are problematic: there is no connection to ploughing in Celtic (Bailey 1960: 80), nor much of it in Indo-Iranian.

2. 'To plough', 'to sow'

Proto-Indo-Iranian used a root **kar-*, more often enlarged **karš-* 'to pull' also for 'to plough'. Proto-Iranian had a suppletive paradigm with a present stem **kāraya-* 'to sow, plant, plough, till' from the simple root and a verbal adjective (later > past stem) **k²ršta-* from the longer form; in Vedic, only the enlarged form is used: present *kṛśá-* 'to plough' (~ *kárša-* 'to pull'). Cf. the nominal derivative **kṛśi-* 'ploughing, furrow' > Av. *karšāii-* = Ved. *kṛśi-* (both in *figura etymologica*) with its derivatives Av. *karšiuuant-* ≈ Ved. **kṛšivan-* 'ploughing person, farmer'. Derivatives of **karš-* can also mean 'border, land', cf. Clr. **kəršwar* 'region'. Cf. Greek *télson* 'furrow' < **k^wéls-o-*. A verb from the root **h₂arh₃-* 'to plough' normally used in European branches of Indo-European is not attested in Indo-Iranian. One general term for 'seed' was PII **bíja-* > Clr. **bīza-* in Sogdian *βyžk* = Ved. *bīja-*, with a foreign-looking structure (cf. Mayrhofer 1996: 227).

3. 'Plough' and its parts

We can reconstruct a Proto-Indo-Iranian word for 'ploughshare', namely **spāra-*³¹ > PIr. **spāra*: MP. **spār* > NP. *sipār*, *supār* = Ved. **spāra-*, **spāla-* > Pa. Pk. *phāla-* (→ Ved. *phāla*), enlarged in PIr. **spārana-* (?) > Šu. *sipōrn* and probably in Wx. **spārna-* > *spundr*? For the plough itself no Proto-Indo-Iranian word can be assured. Ved. *śīra-* n. (since RV10, Pa.+), probably connected to *śītā-* 'furrow', may be old

29. The vocalism of Greek points to original **h₂-* and thus PII **h-*; this is not contradicted by the Iranian evidence, since the Persian word (without *h-/x-*) may easily be influenced by Avestan and/or Parthian (cf. also the general merger of *u-* and **hu-* in Old Persian).

30. Blažek (2005: 220) follows Witczak 2003 in deriving this Irish word from a putative **arg^wh-r/n-* 'millet' (with aspirate because of Greek *orphimē*) they also see in Iranian, but it is both semantically and formally better to connect it with the Greek and Indo-Iranian words. Irish <*b*> can only go back to **b* or **w*, but **g* would yield **g* <*g*>.

31. Probably a loanword, see Lubotsky (2001: 312).

but has no Iranian cognates. A clearly borrowed term is attested in Ved. *lāṅgala-* n. (since RV), Pa. *naṅgala-*, Pk. *laṅgala-*, *ṅaṅgara-* (see Mayrhofer 1996:477). PII **hayš(a)-* ~ **hišá-* (maybe originally **háyHš-h-* ~ **hiHš-áh-*) ‘pole, shaft’ was also used to design the plough-pole or the plough in Iranian, cf. **hayš(a)-* in Av. *aēša-* dual V. 14,10 ‘shaft, plough-pole?’; MP. p. *’yš /ēš-* ‘plough?’, m. *hyš /hēš-* ‘plough-share’, NP. *xēš* ‘yoke, plough’; Wx. *yīšək* ‘handle of plough’, also found as a loanword in Mordva and Permian (cf. Katz 1983; 2003:252); the other stem-variant in Ved. *īšá-* ‘pole, shaft’ appears not to have been used for ploughs. Cf. Hitt. *hissa-* ‘carriage pole, thill’ and Slavic **ajes-* > Slov. *oje(s)-* ‘thill’, PIE **h₂ajH-(e)s-* ~ **h₂iH-s-*³² (also borrowed from an unidentified Indo-European language into Finnic, cf. Finnish *aisa* ‘shaft’). The specialization to a plough term seems to be an Iranian innovation.

3. Conclusions

The pastoral terminology of Indo-Iranian is clearly inherited; most often we find regular correspondences within and outside of Indo-Iranian. In contrast to that, plant cultivation terminology most often shows irregular correspondences, pointing to early or later loans (sometimes *wanderwörter*). Only for some few grain terms is inheritance probable: it is assured for **yáwa-* ‘corn, barley’ and **d^hānā-* ‘(roasted) grains’ and probably also Iranian **ādu-* ‘grain’, and possible for two only marginally attested terms for ‘oats’, **(h)awiš-* and **ćapar-*. The terms for ‘wheat’ looks like loanwords somehow connected to Hittite *kant-*. All the other terms are most probably or even certainly loanwords with no clear connections to Western languages. Also the terms for agricultural technology are rather different from those found in Europe. Taken together, this situation speaks for a mainly pastoralist rather than agricultural economy at the time of Proto-Indo-Iranian. This agrees with the picture found in the earliest Indo-Iranian texts.

32. The exact reconstruction is not very clear; cf. also another derivative in Greek *oāξ* ‘helm, handle of rudder’ which has been taken as an argument for initial **h₃*.

Abbreviations

Arm.	Armenian	Par.	Parachi
Av.	Avestan	Parth.	Parthian
Bactr.	Bactrian	(P)IE	(Proto-)Indo-European
Bal.	Balochi	(P)II	(Proto-)Indo-Iranian
Baxt.	Baxtiari (Bakhtiari)	Plr.	Proto-Iranian
CII	Common Indo-Iranian	Pk.	Prakrit
Clr.	Common Iranian	PL.	plural
f.	feminine	Pto.	Pashto
Finn.	Finnish	Roš.	Rošani
Hitt.	Hittite	Shgr.	Shughni-Group
HLuw.	Hieroglyphic Luwian	Siv.	Sivendi
IA	Indo-Aryan	Skt.	Sanskrit
Išk.	Iškašimi (Ishkashimi)	Skt. ep.	Epic Sanskrit
Kal.	Kalasha	Slov.	Slovene
Kho.	Khovar	Sogd.	Sogdian
Khot.	Khotanese	Sogd. b.	Sogdian (Buddhist)
Kurd.	Kurdish	Sogd. sm.	= s.m. (Sogdian script), (Manichaean)
Lat.	Latin	Šu.	Šuyni (Shughni)
Lith.	Lithuanian	Taj.	Tajik
LW	loanword	Tib.	Tibetan
m.	masculine	Tal.	Talyshi
MP. m.	Middle Persian (Manichaean)	Toch.	Tocharian
MP. p.	Middle Persian (Pahlavi)	Ved.	Vedic
Munj.	Munji	Ved. AV	Atharvaveda
Myc.	Mycenaean Greek	Ved. DhS	Dharmas
NP.	New Persian	Ved. YV	Yajurveda
Nur. a.	Ashkun	Wan.	Wanetsi
Nur. km.	Kamviri	Wj.	Wanji
Nur. kt.	Kataviri	Wx.	Waxi
Nur. v.	Vasi-vari	Xwar.	Xwarezmian
Nur. w.	Waigali	Yagn.	Yaghnobi
OP.	Old Persian	Yazg.	Yazyulami (Yazghulami)
Orm.	Ormuri	Yid.	Yidgha
Oss. D.	Ossetic (Digor)	YM.	Yidgha-Munji
Pa.	Pāli		
Pam.	Pamir		

References

- Abaev, Vladimir Iljič. 1979. *Istoriko-Ėtimologičeskij Slovar' Osetinskogo Jazyka*, Vol. 3. Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Adams, Douglas Q. 1999. *A Dictionary of Tocharian B* [Leiden Studies in Indo-European 10]. Amsterdam: Rodopi.
- Bailey, Harold Walter. 1960. *Indagatio Indo-Iranica*. *TPS* 59: 62–86.
- Bailey, Harold Walter. 1979. *Dictionary of Khotan Saka*. Cambridge: CUP.
- Bartholomae, Christian. 1904. *Altiranisches Wörterbuch*. Straßburg: Trübner.
- Blažek, Václav. 2005. Review of Witczak 2003. *Sborník Prací Filozofické Fakulty Brněnské Univerzity* [Studia Minora Facultatis Philosophicae Universitatis Brunensis A 53]. *Linguistica Brunensia*, 219–225.
- Blažek, Václav. 2017. On Indo-European 'barley'. In *Etymology and the European lexicon*. In *Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft*, 17–22 September 2012, Copenhagen, Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds), 53–67. Wiesbaden: Reichert.
- Edelman, Joy = Ėdel'man, Džoj Iosifovna. 2011. *Ėtimologičeskij slovar' iranskix jazykov*. Tom 4, i-k. Moscow: Vostočnaja Literatura.
- Emmerick, Ronald E. 1966. Some reinterpretations in the Avesta. *TPS* 65: 1–23.
- Hamp, Eric P. 1973. Once again on Iranian *ādu-. *TPS* 72: 137.
- Henkelman, Wouter. 2010. "Consumed before the King." The Table of Darius, that of Irdabama and Irtaštuna, and that of his Satrap, Karkiš. In *Der Achämenidenhof / The Achaemenid Court. Akten des 2. Internationalen Kolloquiums zum Thema "Vorderasien im Spannungsfeld klassischer und altorientalischer Überlieferungen"*, Landgut Castelen bei Basel, 23–25. Mai 2007, Bruno Jacobs, Robert Rollinger (ed.), Wiesbaden: Harrassowitz.
- Hock, Wolfgang et al. 2015. *Alltitauesches etymologisches Wörterbuch*, unter der Leitung von W. H. und der Mitarbeit von Elvira-Julia Bukevi und Christiane Schiller bearbeitet von Rainer Fecht, Anna Helene Feulner, Eugen Hill und Dagmar S. Wodtko, 2 Vols. Hamburg: Baar.
- Hyllested, Adam. 2017. Again on pigs in ancient Europe: The Fennic connection. In *Etymology and the European lexicon. Proceedings of the 14th Fachtagung der Indogermanischen Gesellschaft, 17–22 September 2012, Copenhagen*, Bjarne Simmelkjær Sandgaard Hansen, Benedicte Nielsen Whitehead, Thomas Olander & Birgit Anette Olsen (eds), 183–196. Wiesbaden: Reichert.
- Katz, Hartmut. 1983. Hethitisch *hišša*- und Zubehör. *Orientalia* 52: 116–122.
- Katz, Hartmut. 2003. *Studien zu den älteren indoiranischen Lehnwörtern in den uralischen Sprachen*, aus dem Nachlaß hrsg. von Paul Widmer, Anna Widmer und Gerson Klumpp. Heidelberg: Winter.
- Kümmel, Martin Joachim. 2016. Is ancient old and modern new? Fallacies of attestation and reconstruction (with special focus on Indo-Iranian). In *Proceedings of the 27th Annual UCLA Indo-European Conference, Los Angeles, October 23rd and 24th, 2015*, David M. Goldstein, Stephanie W. Jamison, Brent Vine (eds), 79–96. Bremen: Hempen.
- Kuz'mina, Elena E. 2007. *The origins of the Indo-Iranians*, J. P. Mallory (ed.). Leiden: Brill.
- Lubotsky, Alexander M. 2001. The Indo-Iranian substratum. In *Early contacts between Uralic and Indo-European. Linguistic and archaeological considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki, 8–10 January 1999*, Christian Carpelan, Asko Parpola, Petteri Koskikallio (eds), 301–317. Helsinki: Suomalais-Ugrilainen Seura.

- Lubotsky, Alexander M. 2002. Scythian Elements in Old Iranian. In *Indo-Iranian Languages and Peoples* [Proceedings of the British Academy 116], Nicholas Sims-Williams (ed.), 189–202. Oxford: OUP.
- Mayrhofer, Manfred. 1992, 1996, 2001. *Etymologisches Wörterbuch des Altindoirischen*, 3 Vols. Heidelberg: Winter.
- Morgenstierne, Georg. 1929. *Indo-Iranian Frontier Languages, Vol. I: Parači and Ormuri*. Oslo: Aschehoug.
- Morgenstierne, Georg. 1938. *Indo-Iranian Frontier Languages, Vol. II: Iranian Pamir Languages (Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi)*. Oslo: Aschehoug.
- Morgenstierne, Georg. 1974. *Etymological Vocabulary of the Shughni Group*. Wiesbaden: Reichert.
- Morgenstierne, Georg. 2003. *A New Etymological Vocabulary of Pašto*, compiled and edited by J. Elfenbein, D. N. MacKenzie and Nicholas Sims-Williams. Wiesbaden: Reichert.
- Nikolaev, Alexander S. 2014. Greek εἰαμενή, Vedic *yáva*-. *Münchener Studien zur Sprachwissenschaft* 68: 127–140.
- Pachalina, Taťjana N. 1983. *Issledovanie po sravnitel'no-istoričkoj fonetike pamirskich jazykov*. Moscow: "Nauka".
- Parpola, Asko. 2012. The Dāsas of the R̥gveda as Proto-Sakas of the Yaz I-related cultures: With a revised model for the protohistory of Indo-Iranian speakers. In *Archaeology and Language: Indo-European Studies Presented to James P. Mallory*, Martin E. Huld, Karlene Jones-Bley, Dean Miller (eds), 221–264. Washington DC: Institute for the Study of Man.
- Pinault, Georges-Jean. 1997. Terminologie de petit bétail en Tokharien. *Studia Etymologica Cracoviensia* 2: 175–218.
- Rastorgueva, Vera S. & Edelman, Joy I. = Èdel'man, Džoj Iosifovna. 2000, 2003, 2007. *Ètimologičeskij slovar' iranskix jazykov*. Tom 1, a-ä; Tom 2, b-d; Tom 3 f-h. Moscow: Vostočnaja Literatura.
- Rédei, Károly. 1986. *Zu den indogermanisch-uralischen Sprachkontakten*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften
- Rossi, Adriano. 2010. Possiamo continuare a conettere ir. *ādu- a ie. 'mangiare'? In *Scritti in onore di Eric Pratt Hamp per il suo 90. compleanno*, G. Bellusco, A. Mendicino (eds), 331–344. Cosenza: Università della Calabria, Centro Editoriale e Librario.
- Starostin, Sergei. 1988. Indoeuropejsko-servernokavkazskije izoglossy. In *Drevnij Vostok: Ètnokul'turnye svjazi LXXX*, G. M. Bongrad-Levin, V. G. Arzinba (ed.), 112–163. Moscow: Nauka.
- Steblin-Kamenskij, Ivan M. 1999. *Ètimologičeskij slovar' vachanskogo jazyka*. St. Petersburg: Peterburgskoe Vostokovedenie.
- Szemerényi, Oswald. 1969. Etyma latina II (7–18). In *Studi linguistici in onore di Vittore Pisani*, Vol. II, 963–994. Brescia: Paideia.
- Vine, Brent. 2009. A yearly problem. In *East and West, Papers in Indo-European Studies*, Kazuhiko Yoshida & Brent Vine (eds), 205–224. Bremen: Hempen.
- Witczak, Krzysztof Tomasz. 2003. *Indoeuropejskie nazwy zbóż*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Witzel, Michael. 2009. The linguistic history of some Indian domestic plants. *Journal of BioSciences* 34 (6): 829–833. doi:10.1007/s12038-009-0096-1
- Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship / Voprosy jazykovogo rodstva* 12: 113–148.

Milk and the Indo-Europeans

Romain Garnier, Laurent Sagart and Benoît Sagot

Université de Limoges and Institut Universitaire de France / Centre National de la Recherche Scientifique / Institut National de Recherche en Informatique et en Automatique

Recent evidence from archaeology and ancient DNA converge to indicate that the Yamnaya culture, often regarded as the bearer of the Proto-Indo-European language, underwent a strong population expansion in the late 4th and early 3rd millennia BCE. It suggests that the underlying reason for that expansion might be the then unique capacity to digest animal milk in adulthood. We examine the early Indo-European milk-related vocabulary to confirm the special role of animal milk in Indo-European expansions. We show that Proto-Indo-European did not have a specialized root for ‘to milk’ and argue that the IE root **h₂melǵ-* ‘to milk’ is secondary and post-Anatolian. We take this innovation as an indication of the novelty of animal milking in early Indo-European society. Together with a detailed study of language-specific innovations in this semantic field, we conclude that the ability to digest milk played an important role in boosting Proto-Indo-European demography.

Keywords: Indo-European, etymology, DNA, archaeology, Yamnaya culture

Introduction

The Indo-European hypothesis is well over two hundred years old. A strong consensus exists among linguists on the existence of an Indo-European proto-language. There is no disagreement on which languages are Indo-European and which are not. There is also a broad consensus that the first split in the family separated the Anatolian branch, whose main representative is Hittite, from the rest, which we refer to as “Core Indo-European” in this paper. There exists a healthy range of opinions on issues of reconstruction. Beyond these, areas under discussion relate to the location of the homeland, the time depth of the ancestral language and the subsistence of the original community.

Regarding these issues, mainly two theories are in presence: the Pontic Steppe theory, supported by a majority of scholars, and the Anatolian theory. The first identifies the ancestral group with Yamnaya culture, in the steppes north of the Black Sea around 4000 BCE or slightly later, and argues that Proto-Indo-European speakers were hunter-gatherers or pastoralists. Archaeologists working under the Pontic Steppes hypothesis (Mallory 1989; Anthony 2007) have presented detailed accounts of the interface between archaeology and linguistics. An unresolved issue is that the Pontic Steppes hypothesis has not so far provided a principled answer to the question of why the Indo-European languages have replaced the languages of the farmers over much of Europe and South Asia, despite the presumably more favorable demography of farmers. The Anatolian theory arose as an attempt to answer this question. Renfrew (1987) proposed that the first Indo-Europeans were western Eurasia's first farmers, who domesticated barley and wheats in Anatolia 10,000 years ago; and that the success of their languages was the direct result of the success of agriculture. According to him, the languages of the farmers are now spoken over large tracts of western and southern Eurasia because the demography of farmers is generally more favorable than that of hunter-gatherers or pastoralists. Accordingly, the Anatolian theory currently places Proto-Indo-European speakers in Anatolia ca. 6500 BCE. The Anatolian theory of Indo-European origins is one of the models for the more general Farming/Language Dispersal Theory.

More than the mismatch between the alleged early date of Proto-Indo-European and the reconstructability of wheeled transport vocabulary – which absolutely cannot be as old as agriculture – it is the absence at the highest node in the Indo-European tree of a clear, diversified Proto-Indo-European agricultural vocabulary (Uhlenbeck 1895, 1897; Kortlandt 2009), which reveals the basic problem of the Anatolian theory. Indo-European cereal-related vocabulary exists, but is either regional, semantically too vague to permit the inference of farming, or unrelated to agriculture. Thus Lat. *hordĕum*, -ī n. 'barley', often seen as a direct cognate of Germ. **gerstō*- f. 'barley', is a late formation from *horridus* 'shaggy, bristly' > **horrid-ium* n. 'ear of barley', regularly syncopated in **hórdium* > Vulg. Lat. *hórdĕum*. The source is PIE *ǵ^h*ers-* 'to be bristly'; the Germanic word is perhaps independently derived from the same source: ears of barley are indeed strikingly bristly. The cognate set Lat. *Cērēs* f. 'goddess of vegetal growth', Hitt. *karaš* n. 'cereal plant', MoGerm. *Hirse* m. 'millet' (< Com. Germ. **hersija(n)-*), does not allow the name of a specific cereal to be reconstructed: rather, it goes back to PIE **kerh₃-* 'satiated' (cf. Lith. *šerti* 'feed', Gr. κορέννυμι 'satiated') with 'nourishing substance, kernel' as intermediate notion.

Another well-known name for 'grain' was PIE **ǵéu-o-* (NIL: 407–410), cf. Ved. *yáva-* m. 'barley, wheat, grain' (= YAv. *yauua-*), Hitt. *ewa-* (*ewan-*) 'name of a cereal', Gr. ζειαί f.PL. 'wheat', Lith. *jāvas* m. 'wheat', PL. *javaĩ* 'wheat grains', OArm. *յօվ* 'sprout'. The possibility of a final laryngeal (PIE **ǵéu(h₁)-o-*) was assumed because

of a wrong etymological connection with Ved. *gáv-yū-ti-* F. ‘pasture’ which is unrelated according to Nikolaev (2014: 131). According to Ivanov (2003: 195 ff.), we are dealing with the PIE root **ǵieǵ-* ‘to bind, mix’ (*LIV*²: 314) reflected by AVed. *yaúti* ‘to unite, bind’ and Lith. *yaūti* ‘to mix’, on which an adjective **ǵieǵ-ó-* ‘mixed’ was built (on the same pattern as Gr. λευκός ‘white’). We may assume that the barytonesis is a marker of nominalization (PIE **ǵieǵ-o-* M.PL. ‘mixed grains’). The original meaning was probably *‘mixed fodder for cattle’.

Words like ‘grain’, ‘awn’ in themselves do not necessarily indicate agriculture: knowledge of such notions is consistent with the collecting of wild cereals, as are words for grinding. Conspicuously lacking in the earliest Proto-Indo-European vocabulary are words for notions that unequivocally indicate agriculture: sowing, weeding, harvesting, fields, seeds for sowing, as well as stable names for domesticated cereals. The Austronesian family, the other model for the Farming/Language Theory (Bellwood 1985), has a much stronger claim of having arisen at least partly as a result of a shift to agriculture: Austronesian vocabulary reconstructable at the highest level includes all the notions (‘to sow broadcast’, ‘to weed’, ‘to harvest’, ‘field’, ‘seeds for sowing’) that are missing in Proto-Indo-European, plus the names of three domesticated cereals: foxtail millet, broomcorn millet and rice (Sagart et al. in press). Proto-Indo-European therefore cannot have been the language of a group of farmers, whether in Anatolia or elsewhere. Instead, Proto-Indo-European vocabulary at the highest level (i.e. including Anatolian) is animal-oriented, with stable names for bovines and ovines, animal fodder, and cattle-drawn carts, at least.

While we think the Anatolian theory is in all likelihood incorrect, we regard the idea that the formation of a language family normally implies demographic expansion as a precious insight of the Farming/Language theory. In this paper, we propose that a demographic mechanism explains part of the success of the Indo-European languages and the demise of the languages that preceded them, although the mechanism we have in mind is different from Renfrew’s. In the first part, we report on recent strands of research in archaeology, human genetics and the early history of dairying. These give additional support to the Pontic Steppe hypothesis by showing that the speakers of Proto-Indo-European were the first in Eurasia among whom the ability to drink milk into adulthood developed, and that this ability became dominant in western Eurasia as a result of Indo-European expansions.

In the second part, we examine the Indo-European dairy vocabulary, especially the verb ‘to milk’ and the noun ‘milk’, and describe historical changes in this vocabulary that testify to the rise of milking activities and the growing importance of animal milk in the early Indo-European diet. Second, we reproduce ancient textual evidence associating adult milk drinking with Indo-European, especially Indo-Iranian, speakers. Finally, we document the earliest evidence for adult milk drinking based on parallel expressions from the ritual Indo-Iranian literature.

In conclusion, we argue that lactose tolerance provided the early Indo-Europeans with a demographic edge and possibly with an increase in physical stature, both leading to military advantage over preexisting farming communities that were economically successful but lacking in the political means to mount a coordinated resistance. Elite dominance of Indo-European speakers led to widespread language shift towards Indo-European dialects on the part of farmers, explaining the success of Indo-European languages over those of their European farming predecessors.

1. The archaeological and genetic background

Recent archaeological and genetic work has provided decisive evidence for the Pontic Steppe theory. Haak et al. (2015) showed that a massive migration of Yamnaya hunter-gatherers out of the Pontic steppes into the Corded Ware culture of NW Europe ca. 4500 years ago established a new population component there, distinct from both palaeolithic hunter-gatherers and early European farmers who had previously spread from Anatolia. Further, in a study of ancient DNA from 101 Bronze Age Europeans, Allentoft et al. (2015) showed that the highest levels of a gene allowing adults to digest lactose and consume raw milk are found in the burials of Yamnaya culture and its offshoots the Corded Ware and Afanasievo cultures. They state that by 3000 BCE Yamnaya culture had replaced Neolithic farmers from Hungary to the Urals: they regard the Corded Ware culture of northwestern Europe as possibly derived from Yamnaya, but also including Neolithic farmers. They date its establishment at 2800 BCE. Despite the differences in dates, both Haak et al. and Allentoft et al. link the westward Yamnaya migration with the spread of Indo-European languages in Europe; Allentoft et al. further argue that the spread of lactose tolerance in Europe is due to Indo-European expansions.

Different strands of recent work on dairying in Neolithic Europe provide useful background on the development of lactose tolerance in Europe. As recently as 7000 years ago all human populations were lactose-intolerant (Leonardi et al. 2012): adults lacked the enzyme lactase and could not digest the sugar lactose contained in milk. Lactose tolerance arose independently in several of the world's populations, both in Africa and Eurasia. As for Eurasia, the areas of maximum lactase persistence, as mapped by Leonardi et al., broadly coincide with the Corded Ware culture in NW Europe and with a zone centered on coastal Pakistan, extending into southeastern Iran and northwestern India. This is consistent with a link between lactose tolerance and the spread of Indo-European speakers.

The invention of cheese, a milk derivative poor in lactose, by early farmers in Northwest Anatolia ca. 8500 BP (Evershed et al. 2008) for the first time allowed

humans to turn animal milk into a stable source of food. This presumably contributed to the positive demography of early farming populations. As they spread over Europe, the farmers brought cheese-making with them (Salque et al. 2013). However, they were themselves largely lactose-intolerant (Burger et al. 2007; Allentoft et al. 2015): the capacity to directly drink animal milk results from a genetic mutation allowing the enzyme lactase to persist in adults, a mutation which only arose a few millennia later. We follow Burger (oral remarks cited in Owen 2010) in supposing that contact with cheese-making farmers revealed the lactase persistence gene in certain hunter-gatherer individuals from the Pontic steppes, and that this beneficial gene was subsequently strongly selected for. Presumably, the incidence of the gene rapidly increased in the Yamnaya population, fostering population growth; increased reliance on animal milk required more pasture lands; these became scarce in the homeland area, leading to migrations and territorial expansions – towards Afanasievo culture in the Minusinsk basin before 3000 BCE (perhaps ancestral to the Tocharians); towards northern Europe and towards the Andronovo culture (perhaps ancestral to Indo-Aryan) around the Sea of Aral in the early/mid-second millennium BCE.

In the next section we examine the linguistic and philological evidence on the place of milk in the early Indo-European diet.

2. Linguistic and philological evidence on the place of milk among Indo-Europeans

In this section, we examine the Proto-Indo-European word for ‘to milk’ (2.1), starting with Hittite (2.1.1), then moving to the Core IE root $*h_2melĝ-$ ‘to milk’ (2.1.2). In an excursus in Section 2.1.3 we discuss the Indo-Iranian root $*d^hauĝ-$, both ‘to milk’ and ‘to give milk’. We next move on to the noun ‘milk’: we first examine languages where both ‘to milk’ and ‘milk’ are from $*h_2melĝ-$ and those where only ‘to milk’ is from $*h_2melĝ-$ (2.2). We discuss, and reject, the widely accepted equation between Gr. γάλα N. ‘milk’ and OArm. *kat’n* ‘id’, proposing a new etymology for Gr. γάλα (2.2.1). Our new etymology for Lat. *lac*, *lactis* N. ‘milk’ (2.2.2) tentatively places it under the root IE $*h_2melĝ-$ ‘to milk’. We then show that the Core IE root $*h_2melĝ-$ ‘to milk’ (2.2.3) is secondary, suggesting it originates in a Core IE compound $*h_2mĥH-léĝ-$, GEN. $*h_2mĥH-léĝ-ós$ ‘he who collects ($*lēĝ-$) liquids/milk’. In Section 2.3, we scan Greek and Latin texts for evidence of milk-drinking among “barbarian” adults: Homer and Homeric scholia (2.3.1), Hesiodus (2.3.2), Hippocrates (2.3.3), Herodotus (2.3.4) and Pliny the Elder (2.3.5), showing that all such references point to speakers of Indo-European languages. In a conclusion to Section 2 (2.3) we note the involvement of milk with Indo-Iranian ritual, pointing

out that it prescribes milk drinking by adults. Finally, in our general conclusion we describe the demographic and biological mechanisms through which milk-drinking promoted the spread of Indo-European languages and the demise of the languages of early European farmers.

2.1 Indo-European words for ‘to milk’

2.1.1 Hittite

Hittite does not have a specialized verb ‘to milk’. Milking was practiced but the texts either use the Hittite root *lā-* ‘to let, make flow’ (< PIE **leh₁-* ‘to let’), for instance GA *lāttat* ‘he let the milk flow = he milked’ (KBo III 8 III 30–31), or the locution GA *ḫamikta* ‘he pressed the milk, he milked’ (KBo III 8 III 12–13), where GA, the sumerogram for ‘milk’, is more probably an accusative of product or result than an accusative of direct object. The verb *ḫamikta* ‘he pressed’ is from the nasal-infixes present stem *ḫamink-* ‘to tie together, press together’ (< PIE **h₂emǵ^h-* ‘to squeeze; narrow’). A third expression occurs in Hittite texts: *ḫūratiššan ḫamikta* ‘he squeezed the udder’ (KBo III 8 III 12–13). This makes it likely that like Anatolian, its primary branch, Proto-Indo-European lacked a specialized root for the verb ‘to milk’. However, this is only an *argumentum ex silentio*.

The Hittite name for ‘milk’ cannot be recovered due to generalized use of the sumerogram GA: consequently the Proto-Indo-European word for ‘(to) milk’ cannot be known either. Only a Proto-Indo-European root for ‘to suck mother’s milk’ is known: **d^heh₁-* (*LIV*²: 138 ‘Muttermilch saugen’). The same root (with **-i-* extension) is well attested in Anatolian (cf. Hitt. *tēdan* ‘teat’ < **d^héh₁-i-tom*).¹ The archaic reduplicated neuter stem PIE **d^héd^hh₁-i* ‘mother’s milk’ (Ved. *dādhi*, *dadhnás* n. ‘thick sour milk’) underwent a sporadic shift to a generic name for ‘milk’, as is clear from OPr. *dadan* n. ‘milk’.

2.1.2 The Core IE root **h₂melǵ-* ‘to milk’

This root is widespread among Indo-European languages outside of Anatolian: Lat. *mulgeō*, *ēre* ‘to milk’ (< PIE iterative stem **h₂molǵ-éǵ-e/o-*), reflected by Rom. *mulge*, It. *mungere*, OFr. *moudre* (< Vulg. Lat. **mulǵērē*); Gr. *ἀμέλω* ‘to milk’ (< PIE root present **h₂mélǵ-e/o-*), whence MoGr. *ἀλέω* (Vulg. *ἀρμέγω*); Lith. *milžti* (present stem *mélžu*) ‘to milk’; OCS *mlěšti* ‘id.’; Com. Germ. **mel(u)kan^{an}* (OE *melcan*, MoGerm. *melken* and *melchen*). Albanian agrees with the reflex of a *e*-grade present stem as well (Alb. *mjell* ‘to milk’). Common Celtic is unique among Indo-European

1. According to Kloekhorst (2008: 877), the lenition is triggered by the preceding accented diphthong.

languages of Europe in reflecting a zero-grade thematic root present: Com. Celt. **mlig-e/o-* ‘to milk’ (< PIE **h₂mlǵ-é/ó-*), whence OIr. *bligim* ‘id.’ and Gallo-Rom. **blig-āre* ‘id.’ (< Gaul. **blig-*) reflected by OFr. *blechier* ‘to milk’, mostly famous for its designation of a French cheese: *Roblochon* (or *Re-*), which is made from milk of a second milking (cf. OFr. *re-blechier* ‘to milk a second time’). Root **h₂melǵ-* ‘to milk’ is also found in the very far east of the Indo-European domain: Toch. B *malḡwer n.* ‘milk’ and Toch. A *malke* ‘id.’. These Tocharian nominal stems are not likely to be directly inherited from the *Ursprache*: rather, they point to an unattested verb Com. Toch. **mälk-* ‘to milk’ (< Core IE **h₂mlǵ-*).

It is noteworthy that there is no evidence at all for root **h₂melǵ-* ‘to milk’ in Indo-Iranian, not even in the modern dialects. The Vedic Narten present *márj-mi* ‘to rub’, sometimes presented as related to **h₂melǵ-* (e.g. Mayrhofer *EWAia* II: 325), must in fact relate to a distinct root, namely **h₂merǵ-* ‘to wipe clean, cleanse, purify, remove completely’.² The two roots have largely non-overlapping semantics, although the derived meaning ‘to pluck’ in Gr. *ἀμέργω* ‘pluck’ (always applied to plant products) could be construed as similar to the action of milking.³ The fact that Vedic *márj-mi* ‘to rub’ and other Indo-Iranian forms under the Vedic root *MRJ-* ‘to wipe, brush’ seem to regularly reflect **h₂melǵ-* ‘to milk’ is the result of the merger of PIE **l* and **r* in Indo-Iranian: Ir. **marj-* can reflect both **h₂melǵ-* and **h₂merǵ-*. In addition, the initial laryngeal **h₂* in **h₂merǵ-* is problematic: there is no reflex of it in Vedic or in Avestan – Mayrhofer’s reluctance to assume an Indo-Iranian etymon *(*H*)*marj-* is understandable (*ibid.*).⁴ The proposed **h₂-* relies exclusively on initial *ǎ-* in Gr. *ἀμέργω*: but the alternation with initial *ó-* in the related form *ὀμόργνυμι* ‘to dry’ (< *‘to rub, wipe out’) is not consistent with **h₂-*.⁵ It is more probable that initial *ǎ-* in *ἀμέργω* is the fruit of contamination from the phonetically and semantically similar, but etymology distinct verb *ἀμέρδω* ‘to deprive, take away’.⁶ The by-form *ὀμόργνυμι* ‘to dry’ itself is analyzable as an old preverbed zero-grade stem **h₂o-mrǵ-néu-*, with a dialectal reflex of **r*. As a result, **h₂merǵ-* should be emended to **merǵ-*, without a laryngeal initial, removing it further away from **h₂melǵ-* ‘to milk’.

2. Cf. Late Av. *ni-marəzišta-* ‘best cleanser’ (of Ahura Mazda).

3. E.g. in *καρπὸν ἀμέργουσιν πεποτημέναι* ‘they pluck the fruit on their wings’, of bees (AP 1. 882).

4. As pointed out by a reviewer, the lengthened reduplication of the Vedic perfect *māmri-* may be considered an argument for **Hmarj-* but it must be admitted that it is not very strong.

5. *Mid.*: ‘to dry oneself’ (most often tears), # *δάκρυ ὀμορξαμένην* ‘drying her tears’ (λ 530).

6. PIE **h₂merd-* ‘to harm, mistreat’ (*LIV*²: 280, s.v. **h₂merd-* ‘ein Leid antun, mißhandeln’). Note the confusion between *ἀμέρσᾶς* and **ἀμέρξᾶς* in AP 7.657.7, pointed out in *BDAG*: 2015, 107, s.v. *ἀμέρδω*.

This in fact suggests a likely explanation for the lack of Indo-Iranian reflexes of $*h_2melĝ-$ ‘to milk’: homonymic clash with a verb ‘to rub, wipe etc.’ reflecting $*merĝ-$ may have caused Indo-Iranian speakers to replace $*h_2melĝ-$ ‘to milk’ with an innovated form, in this case $*d^h aug^h-$ (discussed in Section 2.1.3. below). Homophony between a verb ‘to rub’ and a verb ‘to milk’ would have been particularly undesirable, since rubbing a cow’s udders during milking is painful to the animal, causing it to balk, as is well known to those who practise milking. A homophonic clash of $*h_2melĝ-$ ‘to milk’ and $*h_2merĝ-$ ‘to rub, etc.’ occurs only in Indo-Iranian because only Indo-Iranian does lose the distinction between $*l$ and $*r$. In Section 2.2.5., we will propose a new etymology for $*h_2melĝ-$ ‘to milk’.

To sum up, Vedic *mārj-mi* and other forms under the Vedic root *MRJ-* ‘to wipe, brush’ may be connected to Gr. *ἀμέργω* ‘pluck’ (earlier **μέργω*): both are from a root $*merĝ-$ without a laryngeal, and without any significant connection to $*h_2melĝ-$ ‘to milk’.

2.1.3 Excursus: Indo-Iranian $*d^h aug^h-$ ‘to milk; to give milk (of a cow)’

It is generally assumed that the Iir. root $*d^h aug^h-$ ‘to milk; to give milk’ directly reflects PIE $*d^h eug^h-$ ‘to be efficient’ (Mayrhofer *EWAia* I: 747–8), making it a very ancient root and raising the possibility that Proto-Indo-European may have had another verb ‘to milk’ competing with $*h_2melĝ-$. Indeed, the Vedic verb exhibits a very archaic conjugation pattern, associating an athematic root active present in PIE 3SG. $*-ti$, 3PL. $*-énti$: *dōgdhi* 3SG., *duh-ánti* 3PL. ‘to milk (a cow), extract (soma)’ (< PIE $*d^h éug^h-ti$, $*d^h ug^h-énti$) with a middle present in PIE 3SG. $*-óĵ$, 3PL. $*-róĵ$: *duh-é* 3SG., *duh-ré* 3PL. (< PIE $*d^h ug^h-óĵ$, $*d^h ug^h-róĵ$). This supports the Indo-Iranian verb’s Proto-Indo-European antiquity and is consistent with a link to the PIE root $*d^h eug^h-$, at least on a phonological plane.

At the same time, in the languages (outside of Indo-Iranian) where it is attested, the root $*d^h eug^h-$ is unrelated to milk: Gr. *τεύχω*, ‘to do, make, prepare, build’, Com. Germ. $*dugan^{an}$ (intr.) ‘to be fit, avail’ ~ $*daug-$ (o-grade) ‘id.’ (Go. *daug* 3SG.prf-prs. ‘id.’, G. *taugen* ‘id.’). In addition, there are no expressions using the PIE root $*d^h eug^h-$ and meaning ‘to produce milk’, whether in Greek, Germanic or Indo-Iranian. Moreover, a semantic shift from ‘to produce’ to ‘to milk’ strikes us as unmotivated. These points seem to argue that the Iir. root $*d^h aug^h-$ ‘to milk; to give milk’ acquired its connections to milk no earlier than Indo-Iranian, and not as a result of a straightforward semantic shift.

Based on an old suggestion of Szemerényi, we attempt a new solution to this conundrum. Almost sixty years ago, Szemerényi (1958: 171, fn. 3) suggested that the Iir. root $*d^h aug^h-$ ‘to milk’ is a back-formation from the Indo-Iranian name for ‘daughter’ (Iir. $*d^h ug^{(h)}-H-tár-$), which he thought had originally meant ‘suckling child’ or the like. Szemerényi’s proposal has against it the fact that a back-formation

in Indo-Iranian times from ‘daughter’ could not have possessed the archaic conjugation pattern of IIr. $*d^h\acute{aug}^h-$. His hypothesis has met with a great deal of resistance among scholars. Yet it can be adapted as follows. We assume an unattested action noun PIE $*d^h\acute{e}ug-h_2-e/os-$ N. ($< *d^h\acute{e}(h_1)-u-g-h_2-e/os$) ‘action of sucking mother’s milk’, ultimately based on the PIE root $*d^h\acute{e}h_1-$ ‘to suck mother’s milk’, whose *u*-stem PIE $*d^h\acute{e}(h_1)-u-$ ADJ. ‘female, breastfeeding’ had a velar enlargement $*d^h\acute{e}(h_1)-u-g-$ with a concrete meaning ‘teat (*vel sim.*)’. This secondary derivative served as the basis for an amphidynamic abstract noun PIE $*d^h\acute{e}(h_1)-u-g-h_2$ (GEN.SG. $*d^h-u-g-\acute{e}h_2-s$) ‘femininity’. From this hypothetical form the Proto-Indo-European name for ‘daughter’, containing an athematic variant of the “characterizing” suffix $*-ter-o-$ (Pinault 2007) can be derived: $*d^h(h_1)-u-g-h_2-tér-$. Semantically a daughter would then be a ‘suckling [female] child’, or, perhaps more convincingly, a person giving suckle, assuming the term first designated daughters of child-bearing age. Because PIE $*gh_2$ and $*g^h$ merge as $*g^h$ in Indo-Iranian – and nowhere else – the secondary derivative PIE $*d^h\acute{e}ug-h_2-e/os-$ N. would have resulted in IIr. $*d^h\acute{aug}^{(h)}-H-as$, $*d^h\acute{auj}^{(h)}-H-as-$ N. ‘sucking’ (whence also ‘milking’). This term is in fact attested as Ved. $dóh-as-$ ‘milking’. There is another possibility: a thematic secondary derivative PIE $*d^h\acute{oug}-h_2-o-$ M. ‘id.’ reflected by Ved. $dógham$ ‘milking’ (hap. leg.) and by Pašto *lway* ‘id.’ ($<$ Com. Ir. $*daug-a-$).⁷ As a result of the phonological merger, to Indo-Iranian speakers, $*d^h\acute{aug}^{(h)}-H-as$ N. ‘sucking’ or $*d^h\acute{aug}^{(h)} H-a-$ M. ‘id.’ would have seemed to contain the homophonous – but unrelated – primary IIr. root $*d^h\acute{aug}^h-$ ‘to be efficient, produce’. This would have resulted in the appearance of a hybrid verb, combining the archaic conjugation pattern of root $*d^h\acute{e}ug^h-$ and the milk-related semantics of the action noun PIE $*d^h\acute{e}ug-h_2-e/os-$ (or its thematic by-form $*d^h\acute{oug}-h_2-o-$).

2.2 Indo-European words for ‘milk’ derived from ‘to milk’

We have argued that the Core IE root $*h_2melǵ-$ ‘to milk’ is an innovative form, since Hittite has no specialized root for ‘to milk’. Two sets of languages may be distinguished with respect to this root: (1) those where both the verb ‘to milk’ and the noun ‘milk’ are from $*h_2melǵ-$ (Table 1), and (2) those where only the verb ‘to milk’ is from $*h_2melǵ-$ (Table 2). The situation in Tocharian is more complex: the nouns for ‘milk’ in the two dialects: Toch. A *malke* ‘milk’, B *malkwer* ‘id.’, have different Common Tocharian etymologies: *malke* is from Com. Toch. $*mælk-æy$ ($<$ IE $*h_2molǵ-óǵ-$), a secondary derivative built on the (isolated) IE action noun

7. The Pašto *lway* ‘to milk’, from Com. Ir. $*daušš-aǵa-$ ‘id.’, may rather reflect IIr. $*dauk-$ ‘to milk’, from IE $*deuk-$ ‘to draw’ (e.g. Ossetic *doc-*, Waxi *dic-*). According to Cheung (2007: 66f.), the reconstruction $*daušš-$ is not secure, since most verbs in question can also be explained from Com. Ir. $*daušya-$ $<$ IIr. $*daučja-$ which is required for Ossetic anyway.

* h_2 *molǵ-í-* F. ‘milking’, whereas Toch. B *malk-wer* is a secondary derivative built on a verbal stem Com. Toch. **mälk-* ‘to milk’ (< IE * h_2 *mǵǵ-*).

The fact that the languages where the verb is derived from * h_2 *melǵ-* are a subset of those where the noun from * h_2 *melǵ-* argues in favor of the hypothesis that the nouns are derived from the verb. However, there is clear evidence that we are not dealing with a single innovation: in each language where both the noun and the verb reflect * h_2 *melǵ-*, the noun has the same vocalic grade as the verb: therefore, terms for ‘milk’ must have been derived independently in the daughter languages of Core Indo-European.

Table 1. Languages where both ‘to milk’ and the name for ‘milk’ are from * h_2 *melǵ-*

Com. Germ.	* <i>melk-a-</i> ‘to milk’	* <i>melk-</i> F. ‘milk’ (root noun)
Com. Celt.	* <i>mliġ-e/o-</i> ‘to milk’	* <i>mliχtos</i> M. ‘milk’
Com. It.	* <i>molġ-ĕj-e/o-</i> ‘to milk’	* <i>mlókto-</i> M. ‘milking’ * <i>mlaktá</i> F. ‘milk flow’ (2.2.4)

Table 2. Languages where only the verb ‘to milk’ is from * h_2 *melǵ-*

Greek	<i>ἄμῆλυ</i> ‘to milk’	<i>γάλα, γάλακτος</i> N. ‘milk’
Lithuanian	<i>mélžu</i> ‘to milk’	<i>pienas</i> M. ‘milk’ (Latv. <i>piēns</i>)
Albanian	<i>mjel</i> ‘to milk’	<i>dhallë / dhalltë</i> F. ‘buttermilk’
Com. Slav.	* <i>melz-ti</i> ‘to milk’	(* <i>melkò</i> ‘milk’ < Com. Germ.) ⁸

Because the nouns for ‘milk’ in Table 1 are derived from the verb ‘to milk’, their original referent must have been ‘animal milk’ rather than ‘mother’s milk’. Other innovative Indo-European words for ‘milk’ are not derived from the verb ‘to milk’. Lith. *pienas* M. ‘milk’ and Latv. *piēns* ‘id.’ reflect an IE masculine stem * $póǵH-no-$ ‘thick fluid, *mother’s milk’ (the acute intonation of Lith. *pienas* for expected ***piēnas* according to Saussure’s effect is analogical to the Lith. verb *pýti* ‘to have milk’).

From the underlying IE root * $pēǵH-/pēǵH-$ ‘to be thick’ (*LIV*²: 464 ‘anschwellen’), a neuter stem * $pēǵH-mǵ$ ‘thick fluid’ was also built. This is reflected by OAv. *paēman* N. ‘mother’s milk’, MidPers. *pēm* ‘milk’. This term was also borrowed by Fin. *piimä* ‘sour milk’. The IIr. etymon * $páǵH-as-$ N. (Ved. *páyas-* ‘Lebenskraft’, OAv. *paiiah-* ‘milk’) reflects IE * $pēǵH-e/os-$ N. ‘thick fluid’. On the zero grade of $pēǵH-/pēǵH-$ ‘to be thick’, an adjective * $piH-ǵú-$ ‘thick’ was built, whence the abstract noun * $piH-ǵú-h_2$ F. ‘thickness’. That word became the starting point for a secondary derivative * $piH-ǵu-h_2-s-ó-$ ‘thick fluid’ (cf. Ved. *pīyúṣa-* M.N. ‘colostrum, the milk of

8. A borrowing from Germanic (Derksen 2008: 307). Ru. *молозиво* N. ‘colostrum, beestings’ is inherited.

a cow during the first seven days after calving, biestings, any thick fluid'). According to Garnier (2016b: 1.8), the primary root is PIE $^{*}(s)peh_1-$ 'to swell, get fat, fatten, thrive', with an acrostatic neuter PIE $^{*}(s)póh_1-i$ 'fat'. The adjective PIE $^{*}(s)pəh_1-i-tó-$ 'full of fat, having corpulence' underwent the regular metathesis of laryngeals: PIE $^{*}(s)pih_1-tó-$ 'fattened, fat' and PIE $^{*}pih_1-nó-$ 'swollen', reinterpreted as participles of a secondary root IE $^{*}(s)pejh_1-$ 'to be fat, be thick'.

2.2.1 Other innovative forms for 'milk': Gr. γάλα N. and OArm. kat'n.

An etymological link is widely assumed between Gr. γάλα N. 'milk' and OArm. kat'n 'id.' (Dial. Arm. kaxc'). A link cannot be taken for granted. According to Martirosyan (2010: 345–6), the Armenian forms reflect a proto-Arm. paradigm NOM.SG. *kac' (< *kalc' < PIE $^{*}g_lk-t-s$), acc.SG. *katt'n (< PIE $^{*}g_lk-t-m$), levelled to *kac', *kat'n; this in OArm. kat'n, used as both nominative and accusative while the dialects exhibit the symmetrical levelling *kalc', *katt'n with analogically re-introduced velar *l*, and extension of the old nominative *kalc' (Dial. ModArm. kaxc') throughout the whole paradigm. This theory relates the Armenian forms to an IE etymon $^{*}g_lk-t-s$ and appears to provide a viable link to Gr. γάλα, γάλακτος. However, it stumbles upon three obstacles. First, the animate gender of the PIE etymon $^{*}g_lk-t-s$, acc. $^{*}g_lk-t-m$, does not match the neuter gender of Gr. γάλα. Second, from a Greek point of view, the unexpected disyllabism of the stem γάλακτ- is hardly compatible with an original stem $^{*}g_lk-t-$. Third, as recently demonstrated by Kümmel (2017: 445f.), the inner-Armenian connection of kat'n with kit' 'milking; harvest' and kowt' 'harvest' is no longer compatible with a reconstruction $^{*}g_lk-t-$. An alternative and perhaps preferable explanation is to posit an etymological link between Gr. γάλα and Alb. dhallë / dhalltë F. 'butter milk', reflecting a Proto-Alb. *dzalā- 'id.' (whence also the Rom. loanword zară 'id.'), where *d regularly reflects a PIE palatal $^{*}g$, not a pure velar $^{*}g$. On the basis of the Homeric formula γάλα λευκόν 'white milk' (Δ 434, E 902), we propose an origin of the Greek and Albanian forms in a color adjective Gr. γάλαξ, -ακος * 'white'; this form is actually attested, with the meaning 'a kind of a shell, prob. *Mactra lactea*' (Aristot. HA 528a 23). *Mactra lactea* is white in color. This adjective, reflecting a PIE stem $^{*}g_lh_2-n-k-$ 'bright, white' from PIE $^{*}g_elh_2-$ 'to shine', could have resulted in a substantivized neuter Gr. γάλα. The dental stem of GEN.SG. γάλακ-τος would be secondary. As a typological parallel, we may mention MoAr. laban M. 'milk, whey' with root LBN- 'to be white'.

2.2.2 Other innovative forms for ‘milk’: Lat. *lac*, *lactis* N.

Contrary to a tenacious legend,⁹ this Latin word has nothing to do with Gr. γάλα, γάλακτος N. ‘milk’. Garnier (2016: 306–7) proposes that Lat. *lac*, *lact-is* N. ‘milk’ is a back-formation.¹⁰ The stem **lact-* would be from an unattested verb **amb-lactāre* ‘to milk’¹¹ resulting from **ambī-blactāre* ‘to milk with both hands’ through haplology; **amb-lactāre* itself underwent depreverbatation to *lactāre* ‘to milk’, and a stem **lact-* ‘milk’ was extracted through back-formation. The underlying Proto-Indo-European root must have been IE **h₂melǵ-* ‘to milk’ (cf. Gr. ἀμέλω, Lat. *mulgĕō*). We may envision an action noun IE **h₂mólǵ-to-* M. ‘the milking’, regularly metathesizing to Common Italic **mlók-to* M.; this then further affected with Italic collective suffix *-ā* of concrete meaning (< IE **-eh₂*), giving ‘milk flow’; affixation of *-ā* in turn required change to zero degree still in Italic. One would have expected **molk-tá* (< **młk-tá*) but due to analogy with the strong stem **mlók-to*, resyllabification resulted in **mlāk-tá* F. ‘milk flow’, coexisting with **mlók-to* at Common Italic level.¹² In turn, **mlāk-tá* regularly evolved to unattested Lat. **blactā-*, out of which **ambī-blactāre* ‘to milk with both hands’ was formed.

2.2.3 A new etymology for IE **h₂melǵ-* ‘to milk’

The IE root **h₂melǵ-* ‘to milk’ is phonologically too complex to be primary. Its meaning is both highly specialized and remarkably stable across languages, despite widespread attestation, suggestive of a relatively recent formation. Benveniste (1935: 157) assumed a primary root **h₂em-* ‘to collect liquid’ (cf. Gr. ἄμη F. ‘bucket’) with nominal enlargement PIE †*h₂m-el-g-* in his notation. As a parallel to the proposed *-el-g-* enlargement, he cited Ved. *s^uvargá-* ADJ. ‘heavenly’ which he took to be from PIE †*su-él-g-*. Kümmel (*LIV*²: 265) reconstructs the same root with a final laryngeal: PIE **h₂emH-* ‘to pour’. Reflexes are Com. Celt. **ande=am-je/o-* ‘to pour [water] upon’ (cf. OIr. *and.aim* ‘to wash’) and the doublet Com. Celt. **ad=am-je/o-* ‘id.’ (cf. v.-irl. *ad.aim*), supported by Matasović (2009: 31). The fact that **h₂emH-* was

9. Szemerényi (1991: 1117) and Leumann (1977: 187) assume for Lat. *lac*, *lactis* a PIE etymon **ǵlakt-* N. ‘milk’ supposedly also explaining the Greek forms. This etymology is maintained by Weiss (2011: 147, fn. 82).

10. Archaic nominative *lact* in Varro (*Men.* 26), deemed incorrect by Julius Caesar according to Pompeius Grammaticus (*GLK* 5: 199). Vulgar form *lactĕ* in Plautus (*Bacch.* 6), prefiguring the Romance evolution (cf. It. *latte*).

11. The by-form *lactĕ* (pl.) would be a back-formation from a vulgar doublet **lactiāre* ‘to milk’.

12. Such a resyllabation may be paralleled by OHG *nusta* F. ‘Verbindung’ (< Com. Germ. **nustōⁿ*), analogous to the strong stem Com. Germ. **násta^z* M. ‘binding’ (< PIE **Hnód^h-to-*), instead of phonetically expected *unstōⁿ* according to Griepentrog (1995: 457).

the verb used for pouring or collecting milk is clear from textual evidence: Hom. ἀμάομαι ‘to draw milk, collect’ (mid. ἀμάομαι) is said of curdled milk in ι 247:

αὐτίκα δ’ ἤμισυ μὲν θρέψας λευκοῖο γάλακτος
 πλεκτοῖς ἐν ταλάροισιν ἀμησάμενος κατέθηκεν,
 ἤμισυ δ’ αὐτ’ ἔστησεν ἐν ἄγγεσιν, ὄφρα οἱ εἴη
 πίνειν αἰνυμένῳ καὶ οἱ ποτιδόρπιον εἴη.

He curdled half the white milk and *collected* it in wicker strainers, but the other half he poured into bowls so that he might drink it for his supper.

The secondary derivative Gr. ἄμης, -ητος m. ‘milk cake’ (Aristoph., *Ploutos* 999) is perhaps from an unattested masculine or neuter o-stem *ἄμος ‘milk left to curdle in a bucket, curdled milk’. The Greek word ἄμη f. ‘bucket’ is considered a back-formation (Dieu 2016: 112). We propose that IE * $h_2mel\tilde{g}$ - ‘to milk’ is a secondary root based on a compound * $h_2m\tilde{H}l\tilde{g}$ -, GEN. * $h_2m\tilde{H}l\tilde{g}$ -ós ‘one who collects (* $le\tilde{g}$ -) liquids/milk’. We assume this compound dates back to a period preceding the formation of Core Indo-European since its derivatives meaning ‘to milk’ and ‘milk’ are widespread in the daughter branches of Core Indo-European, including Tocharian. Phonologically, the old GEN.SG. * $h_2m\tilde{H}l\tilde{g}$ -ós resulted in * $h_2m\tilde{.}l\tilde{g}$ -ós with hiatus, whence resyllabation as * $h_2m\tilde{.}l\tilde{g}$ -ós. For a parallel, cf. the resyllabation in the Proto-Indo-European name for ‘wind’: PIE * h_2ueh_1nt -ó- > * $h_2ue.nt$ -ó- > * h_2uent -ó- m. ‘wind’ (cf. Go. *winds*, Lat. *uentus*), a derivative of appurtenance (‘the fast one’) built on the PIE *nt*-stem * h_2uh_1 -ónt-, *-nt-és ‘running’ (Garnier 2014: 63). Finally, through back-formation IE * $h_2ml\tilde{g}$ - ‘milker’ would have triggered the creation of the secondary root * $h_2mel\tilde{g}$ - ‘to milk’, out of which several words for ‘milk’, described above, would later be derived: pre-Core IE * $h_2ml\tilde{g}$ - ‘milker’ → Core IE * $h_2mel\tilde{g}$ - ‘to milk’ → Post-Core IE names for ‘milk’.

2.3 Greek and Latin textual evidence for milk-drinking among Indo-European “barbarians”

2.3.1 *Homer and Homeric scholia*

Homer’s *Iliad* already alludes to milk-drinking among a legendary people of pastoral nomads referred to as “the lordly Hippemolgi”; Herodotus mentions the (Indo-Iranian) Scythians, who drink mare’s milk. Let us start with the very beginning: Homer’s *Iliad*, in which dairy culture was first depicted.

αὐτὸς δὲ πάλιν τρέπεν ὄσσε φαεινῶ
 νόσφιν ἐφ' ἰπποπόλων Θρηϊκῶν καθορώμενος αἶαν
 Μυσῶν¹³ τ' ἀγγεμάχων καὶ ἀγαυῶν Ἰππημολγῶν
 γλακτοφάγων Ἀβίων τε δικαιοτάτων ἀνθρώπων

(N 4–6)

Now Zeus turned away his bright eyes, and looked afar, upon the land of the Thracian horsemen, and of the Mysians that fight in close combat, and of the lordly Hippemolgi who are cheese-eaters, and of the Abii, the most righteous of men.

Those words became enigmatic to the ancients themselves; this very passage was widely commented¹⁴ by antique scholiasts, who identified the Abii either with the Scythians or the (equally Indo-Iranian) Sarmatians:

1. γλακτοφάγων Ἀβίων τε δικαιοτάτων ἀνθρώπων· †λακτινες† ἔθνος, οἱ γαλακτοπόται. Τινὲς τούτους Σαρμάτας φασιν. (Il. xiii.5) “dairy (?) people, who are milk-drinkers. Some also call them *Sarmatians*.”
2. Ἀβίων· πάντων Σκυθῶν ὑποκυψάντων Ἀλεξάνδρῳ μόνους Ἀβίους φασὶν οὐχ ὑπεῖξαι “The *Abii*: amongst all Scythians who have bowed to Alexander the Great, it is said that only the *Abii* didn’t surrender.”
3. οὐς δικαιοτάτους φησὶ διὰ τὸ ἀνεπίμικτον “(Homer) says they are the most righteous among men for their people is unmixed.”
4. Ἀβίων· τῶν νομάδων Σκυθῶν “*Abii*: the nomad Scythians.”
5. τινὲς δὲ τούτους Σαρμάτας φασὶν “some others call them *Sarmatians*”

Modern scholars¹⁵ identify the Abii either with the legendary Hyperboreans, or with the Gabii mentioned by Aeschylus in a fragment of *Prometheus Unbound* (fr. 186). Aristarchus himself endeavoured without success to distinguish between epithet and ethnonym in Homer’s *Iliad* (N 5–6). The word ἰππημολγός could be understood as an epitheton meaning ‘mare-milkers’ (cf. Gr. ἵππος m.f. ‘horse, mare’), associated to Hom. ἀγαυός ‘noble’ and to metrically syncopated # γλακτο-φάγος ‘cheese-eaters’ (here standing for †# γαλακτο-φάγος). Even the word ἄβιος could be understood as an epitheton: ‘without (fixed) subsistence’, whence ‘nomad’ (Gr. βίος m. ‘life’ means also ‘means of life, resources, sustenance’).¹⁶ The legendary Abii who have occasioned so much discussion may eventually be nothing else but

13. Those are the Mysians living on the shore of the Danuvius, not the Mysians from Asia.

14. Text of the scholia by Erbse (1974: 392–396).

15. For instance, Deforge (1986: 194–195), Janko (1992: 42–43), Reece (2001).

16. See for instance βίον πορίζειν τινί “to furnish s.o. the means of sustenance” (Aristoph. *Ve*. 706).

a ghost-word;¹⁷ but the fact that a dairy nomad tribe living on milk and “without fixed subsistence” (that is to say without agriculture) is referred to in this passage is beyond doubt.

2.3.2 *Hesiodus*

Γλακτοφάγων ἐς γαῖαν ἀπίννας οἰκί' ἐχόντων (fr. 54)¹⁸ [probably Scythians]

To the land of Cheese-Eaters, whose houses are chariots.

2.3.3 *Hippocrates*

In his famous treatise *Airs, Waters, Places*, 18, Hippocrates depicts the Scythians' milk-based diet: *Αὐτοὶ δ' ἐσθίουσι κρέα ἐφθὰ καὶ πίνουσι γάλα ἵππων. Καὶ ἵππάκην τρώγουσι· τοῦτο δ' ἐστὶ τυρὸς ἵππων.* “They themselves eat boiled meats and drink mares' milk.¹⁹ They have a sweet-meat *hippake*,²⁰ which is a cheese from the milk of mares.”

2.3.4 *Herodotus: The Massagetae and the Scythians*

2.3.4.1 *The Massagetae*

γαλακτοπόται δ' εἰσὶ (Hdt. i.216) “the Massagetae are milk-drinkers”.

2.3.4.2 *The Scythians (Hdt. iv.2.1–2)*

(1) Τοὺς δὲ δούλους οἱ Σκύθαι πάντας τυφλοῦσι τοῦ γάλακτος εἵνεκεν τοῦ πίνουσι ποιεῦντες ὧδε. Ἐπεὰν φυσητῆρας λάβωσι ὀστεῖνους αὐλοῖσι προσημφερεστάτους, τούτους ἐσθέντες ἐς τῶν θηλέων ἵππων τὰ ἄρθρα φυσῶσι τοῖσι στόμασι, ἄλλοι δὲ ἄλλων φυσώντων ἀμέλγουσι. Φασὶ δὲ τοῦδε εἵνεκα τοῦτο ποιέειν· τὰς φλέβας τε πίμπλασθαι φυσωμένας τῆς ἵππου καὶ τὸ οὔθαρ κατίεσθαι. (2) Ἐπεὰν δὲ ἀμέλξωσι τὸ γάλα, ἐσχέαντες ἐς ξύλινα ἀγγήια κοῖλα καὶ περιστίζαντες κατὰ τὰ ἀγγήια τοὺς τυφλοὺς δονέουσι τὸ γάλα, καὶ τὸ μὲν αὐτοῦ ἐπιστάμενον ἀπαρύσαντες ἠγεῦνται εἶναι τιμώτερον, τὸ δ' ὑπιστάμενον ἦσσαν τοῦ ἐτέρου.

17. In this case, we may amend the *textus traditus* as follows: Μυσῶν τ' ἀγχεμάχων καὶ ἀγαυῶν *ἵππημολγῶν # γλακτοφάγων *ἀβίων τε δικαιοσάτων *τ' ἀνθρώπων (N 4–6) “and upon the land of the illustrious Mysians that fight in close combat, mare-milkers, cheese-eaters, without (fixed) subsistence, the most righteous of men.”

18. Rzach's edition (1902: 145).

19. Note OPr. *aswinan* [*dadan*] N. ‘mare's milk’.

20. This Greek word could be a calque of an Iranian word, **aspa-kā* or the like.

“(1) Now the Scythians put out the eyes of all their slaves because of the milk which they drink; and they do as follows: they take blow-pipes of bone just like flutes, and these they insert into the vagina of the mare and blow with their mouths, and others milk while they blow: and they say that they do this because the veins of the mare are thus filled, being blown out, and so the udder is let down. (2) When they have drawn the milk they pour it into wooden vessels hollowed out, and they set the blind slaves in order about the vessels and agitate the milk. Then that which comes to the top they skim off, considering it the more valuable part, whereas they esteem that which settles down to be less good than the other. For this reason the Scythians put out the eyes of all whom they catch.”

2.3.5 *Pliny the Elder*

Mīrum barbarās gentēs, quā lacte uīuant, ignōrāre aut spernere tot saeculīs cāseī dōtem, densantēs id alioquī in acōrem iucundum. (HN xi.96.3),

“It is a remarkable circumstance, that the barbarous nations which subsist on milk have been for many ages ignorant of the merits of cheese, or else have totally disregarded it; and yet they understand how to thicken milk and form therefrom an acrid kind of liquid with a pleasant flavour.”

2.4 Concluding remarks

The Post-Anatolian innovation points to the creation of a “secondary” root $*h_2melǵ-$ ‘to collect liquid (in a bucket), to milk’. A major part of the Indo-European languages (including Tocharian) used this specialized root to build new names for ‘milk’: Com. Germ. $*mel(u)k-a^n$ N. ‘milk’, Com. Celt. $*mliχtos$ M. ‘id.’, Toch. A *malke*, B *malower* ‘id.’, and (maybe) Lat. *lac*, *lact-is* N. ‘id.’. The highly innovative Balkanic area, although using $*h_2melǵ-$ as a verbal root (Alb. *mjell* ‘to milk’, Gr. ἀμέλω ‘id.’), shows a lexical renewal exemplified by Gr. γάλα N. ‘milk’ (< PIE $*ǵʰh_2-n̥-k-$ ‘white’) and Alb. *dhallë* / *dhalltë* F. ‘buttermilk’, which could reflect Proto-Alb. $*dzalā-$ (< PIE $*ǵʰh_2-éh_2$ F. ‘whiteness’).²¹ Such “modern” designations point to an innovative name for ‘milk’ as consumed by both infants and adults.

21. As already mentioned, there is evidence for a similar lexical renewal in Semitic, where the root ‘to milk’ is $\sqrt{HLB-}$, whereas several languages created a new name for ‘milk’ based on $\sqrt{LBN-}$ ‘to be white’.

The Indo-Iranian data are particularly complex: we may admit that the Core IE root **h₂melǵ-* ‘to collect liquid (in a bucket), to milk’ was lost, because of its homophonic confusion with the unrelated root PIE **merǵ-* (cf. Ved. \sqrt{MRJ} - ‘to wipe, brush’). Besides, the Indo-Iranian tribes seem to have been quite significantly living on milk, at the time when they were still nomads: the first mention ever of milk-drinking by lactose-tolerant adults appears to be in the old Indo-Iranian formula **sáumas iás gáuā* ‘soma-juice²² mixed with (cow’s) milk’, reflected by Late Av. **haomō.yō gauua* ‘soma-juice mixed with milk’ (Yt 3.18 ff.).²³ References to milk in Early Vedic texts are ubiquitous²⁴ and the posterior Ayurvedic literature emphatically states that milk can be consumed by all healthy individuals.²⁵

Common Indo-Iranian has an etymologically totally obscure generic word for ‘milk’: Ved. *kṣīrá-* n. ‘milk’ (also Classical Skr.), and Com. Ir. **xšīra-* ‘id.’, reflected by MoPers. *šīr* ‘id.’. The Iranian languages also have a protean word for ‘milk’, not reflected in Indo-Aryan: OAv. *xšuuīpta-* ‘milk’, Pašto *šaudə* ‘id.’, Khot. *švīda* ‘id.’ from Com. Ir. **xšwīfta-*, of unknown origin (not from Com. Ir. **xšwīd-*, pace Mayrhofer *EWAia* I: 453).

3. Conclusion

Some of the findings in Section 2 are directly interpretable in terms of the genetic and archaeological findings on lactase persistence described in Section 1. First, we have shown that after the separation of the Anatolian branch, and before the breakup of Core Indo-European (dated to ca. 2800 BCE by Chang et al. 2015), a specialized root for ‘milker’ came into existence, out of which a specialized verb ‘to milk’ **h₂melǵ-* was formed. We take the appearance of this term as signalling the new status of animal milking as a well-identified social activity in early Indo-European society. While Proto-Indo-European must have had a word for human milk – not recoverable due to the specificities of Hittite script, we have shown that

22. Whatever soma-juice may have been, it certainly referred to a strong intoxicating liquor – definitely not a beverage for suckling infants.

23. *Lectio* supported by de Vaan (2003: 370) for the *textus traditus*, which reads here †*haomaiō*.

24. If we may say so, the whole *Rig-Veda* is crawling with mentions of milk and sperm.

25. See for instance the gnomic stanza: *kṣīraṃ sarveṣāṃ dehināṃ cānuṣete kṣīraṃ pibanti ca na roga eti || kṣīrāt paraṃ nānyadhāsti vṛṣyaṃ kṣīrāt paraṃ nāsti ca jīvanīyam ||90||* (*Ka.Ka.* 7.90) “Milk is beneficial for healthy individuals; by drinking milk one does not get diseases (*roga-*); hence there is no better aphrodisiac (*vṛṣya-*) than milk; there is no better life-prolonger (*jīvanīyam*) than milk.” Note also: *pravaraṃ jīvanīyānāṃ kṣīraṃ uktaṃ rasāyanam ||218||* (*Cāraka Saṃhitā Sūtrasthāna* 27.218) “Milk is said to be a life-elixir per excellence.”

new Indo-European words for ‘milk’ were formed independently from the verb **h₂melǵ-* ‘to milk’ in the Germanic, Celtic, Italic, Slavic and Tocharian branches: these terms must have designated animal milk for human consumption. This probably indicates a widespread social need for distinct specialized names for the two notions, animal milk and mother’s milk.

By themselves, these linguistic data are silent on whether adult speakers drank animal milk or whether the milk was used to make cheese, or both; though the greater prominence of terms for ‘milk’ compared to those for ‘cheese’ does suggest that milk was directly consumed by adults. Unequivocal linguistic evidence of milk consumption by Indo-European-speaking adults can be found at a later date, in prescribed ritual drinking of *soma* mixed with milk by Proto-Indo-Iranian adults: lactose-persistence in the Proto-Indo-Iranian population (slightly after 2000 BCE according to Chang et al. 2015) must therefore have reached very high levels. Adult milk drinking by Indo-Iranian peoples is further confirmed by descriptions by Roman and Greek authors. Despite being linguistically Indo-European, Roman and Greek authors considered adult milk-drinking a barbarian custom, perhaps because Roman and Greek populations included a large pre-Indo-European farmer component.

To conclude, we suggest that the ability to drink milk in adulthood played an important role first in boosting Proto-Indo-European demography. A larger population in turn required more milk: the need for more pasture lands is probably one strong motivation behind Indo-European territorial expansion. In confrontations with preexisting farming populations, increased population numbers allowed Indo-European groups to prevail militarily over small, or even not-so-small, farming communities which had until then been secure. As a result Indo-European speakers were able to establish themselves durably as a ruling elite over sedentary farming communities speaking non-Indo-European languages. Like horseback riding, teenage and adult milk consumption may also have amplified the military might of Indo-European raider groups by conferring higher bodily stature to Indo-European individuals with the lactase persistence phenotype (Okada 2004).

Abbreviations

ADJ.	adjective	mid.	middle
Alb.	Albanian	MoAr.	Modern Arabic
AP	Anthology Palatine	MoGerm.	Modern German
AVed.	Atharva-Vedic	MoPers.	Modern Persian
Fin.	Finnish	MPers.	Middle Persian
GEN.	genitive	n.	neuter
Com. Celt.	Common Celtic	OArm.	Old Armenian
Com. Germ.	Common Germanic	OAv.	Old Avestan
Com. Ir.	Common Iranian	OCS	Old Church Slavonic
Com. It.	Common Italic	OE	Old English
Com. Slav.	Common Slavic	OFr.	Old French
Gallo-Rom.	Gallo-Romance	OIr.	Old Irish
Gaul.	Gaulish	OPr.	Old Prussian
Go.	Gothic	PL.	plural
Gr.	Greek	PIE	Proto-Indo-European
Hitt.	Hittite	Proto-Alb.	Proto-Albanian
Hom.	Homeric	Rom.	Romanian
IE	Indo-European	sg.	singular
Iir.	Indo-Iranian	Skr.	Sanskrit
It.	Italian	Toch.	Tocharian
Khot.	Khotanese	Ved.	Vedic Sanskrit
Late Av.	Late Avestan	Vulg.	Vulgar
Lith.	Lithuanian	YAv.	Younger Avestan
m.	masculine		

References

- Allentoft, Morten E. *et al.* 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–172. doi:10.1038/nature14507
- Anthony, David. 2007. *The Horse, the Wheel and Language*. Princeton NJ: Princeton University Press.
- Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian Archipelago*. Sydney: Academic.
- Benveniste, Émile. 1935. *Origines de la formation des noms en indo-européen*. Paris: Maisonneuve.
- Burger, Joachim *et al.* 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences* 104(10): 3736–3741. doi:10.1073/pnas.0607187104
- Chang, Will, Cathcart, Chundra, Hall, David & Garrett, Andrew. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1): 194–244. doi:10.1353/lan.2015.0005

- Cheung, Johnny. 2007. *Etymological Dictionary of the Iranian Verb* [Leiden Indo-European Etymological Dictionary Series, Vol. 2]. Leiden: Brill.
- Deforge, Bernard. 1986. *Eschyle, poète cosmique*. Paris: Les Belles Lettres.
- Derksen, Rick. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon* [Leiden Indo-European Etymological Dictionary Series 4], Alexander Lubotsky (ed.). Leiden: Brill.
- Dieu, Éric. 2016. *L'accentuation des noms en *-ā (*-eh₂) en grec ancien et dans les langues indo-européennes*. Innsbruck: IBS 156.
- Erbse, Harmut. 1974. *Scholia Græca in Homeri Iliadem (scholia uetera). Recensuit Hartmut Erbse. Volumen tertium scholia ad libros K-Ξ continens*. Berolini: apud Walter de Gruyter et socios, MCMLXXIV.
- Evershed, Richard P. et al. 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455: 528–531. doi:10.1038/nature07180
- Garnier, Romain. 2014. Sur l'étymologie du latin *uirgō* 'vierge'. *Studia Etymologica Cracoviensia* 19: 59–70.
- Garnier, Romain. 2016. *La dérivation inverse en latin*. Innsbruck: IBS 157.
- Garnier, Romain. 2016b. Secondary roots in PIE. Handout, 15. *Fachtagung der Indogermanischen Gesellschaft*. Vienna, September 2016.
- GLK: *Grammatici Latini ex recensione Henrici Keil* 7: 1873–1923. Leipzig: Teubner.
- Griepentrog, Wolfgang. 1995. *Die Wurzelnomina der Germanischen und ihre Vorgeschichte*. Innsbruck: IBS 82.
- Haak, Wolfgang et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207–211. doi:10.1038/nature14317
- Ivanov, Vyacheslav. V. 2003. On the origins of Tocharian terms for GRAIN. In *Language in Time and Space: A Festschrift for Werner Winter on the Occasion of his 80th Birthday*, Brigitte L. M. Bauer & Georges-Jean Pinault (eds), 189–210. Berlin: Mouton de Gruyter. doi:10.1515/9783110897722.189
- Janko, Richard. 1992. *The Iliad: A Commentary*, IV: 13–16. Cambridge: CUP.
- Kloekhorst, Alwin. 2008. *Etymological Dictionary of the Hittite Inherited Lexicon* [Leiden Indo-European Etymological Dictionary Series 5], Alexander Lubotsky (ed.). Leiden: Brill.
- Kortlandt, Frederik. 2009. C.C. Uhlenbeck on Indo-European, Uralic and Caucasian. *Historische Sprachforschung* 122: 39–47.
- Kümmel, Martin. 2017. Even more traces of the accent in Armenian? The development of tenues after sonorants. In *Usque ad radices. Indo-European Studies in Honour of Birgit Anette Olsen* [Copenhagen Studies in Indo-European 8], Bjarne Simmelkjær et al. (eds), 439–452. Copenhagen: Museum Tusulanum.
- Leonardi, Michela, Gerbault, Pascale, Thomas, Mark G. & Burger, Joachim. 2012. The evolution of lactase persistence in Europe. A synthesis of archaeological and genetic evidence. *International Dairy Journal* 22: 88–97. doi:10.1016/j.idairyj.2011.10.010
- Leumann, Manu. 1977. *Lateinische Laut- und Formen-Lehre*. München: C.H. Beck'sche Verlagsbuchhandlung.
- LIV²: Rix, Helmut et al. 1998[2001]. *Lexikon der Indogermanischen Verben: Die Wurzeln und Ihre Primärstammbildungen*. Unter Leitung von H. Rix, bearbeitet von Martin Kümmel, Thomas Zehnder, Reiner Lipp, Brigitte Schirmer. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Mallory, James Patrick. 1989. In *Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames & Hudson.

- Martirosyan, Hrach K. 2010. *Etymological Dictionary of the Armenian Inherited Lexicon* [Leiden Indo-European Etymological Dictionary Series 10], Alexander Lubotsky (ed.). Leiden: Brill.
- Matasović, Ranko. 2009. *Etymological Dictionary of Proto-Celtic* [Leiden Indo-European Etymological Dictionary Series 9], Alexander Lubotsky (ed.). Leiden: Brill.
- Mayrhofer, Manfred. 1986–2001. *EWAia (I to III): Etymologisches Wörterbuch des Altindoarischen*, III Bände. Heidelberg: Universitätsverlag Winter.
- Nikolaev, Alexander. 2014. Greek εἰαμενή, Vedic yávasa-. *Münchener Studien zur Sprachwissenschaft* 68(1): 127–139.
- NIL: Wodtko, Dagmar S., Irlinger, Britta & Schneider, Carolin. 2008. *Nomina im indogermanischen Lexikon*. Heidelberg: Winter.
- Okada, Tomoo. 2004. Effect of cow milk consumption on longitudinal height gain in children. *American Journal of Clinical Nutrition* 80(4): 1088–1089.
- Owen, James. 2010. Stone Age milk use began 2,000 years earlier. *National Geographic News*, October 28.
- Pinault, Georges-Jean. 2007. A star is born: a ‘new’ PIE *-ter- suffix. In *Verba Docenti, Studies in Historical and Indo-European Linguistics Presented to Jay H. Jasanoff by Students, Colleagues, and Friends*, Alan J. Nussbaum (ed.), 271–279. Ann Arbor MI: Beech Stave Press.
- Reece, Steve. 2001. The ἄθιοι and the γάθιοι: An Aeschylean solution to a Homeric problem. *AJPh*, 122(4): 465–470.
- Renfrew, Colin. 1987. *Archaeology and Language*. London: Jonathan Cape.
- Rzach, Aloisius. 1902. *Hesiodi carmina. Recensuit Aloisius Rzach*. Lipsiae: In aedibus B. G. Teubneri.
- Sagart, Laurent, Hsu, Tze-fu, Tsai, Yuan-Ching & Hsing, Yue-Te. In press. Austronesian and Chinese words for the millets. *Language Dynamics and Change*.
- Salque, Melanie *et al.* 2013. Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature* 493: 522–525. doi:10.1038/nature11698
- Szemerényi, Oswald. 1958. Greek γάλα and the Indo-European term for ‘milk’ with an excursus on Avesta xšvid-. *Zeitschrift für Vergleichende Sprachforschung* 75: 170–190.
- Szemerényi, Oswald. 1991. *Scripta minora: selected essays in Indo-European, Greek, and Latin*, Vol. 3. Innsbruck: IBS 53.
- Uhlenbeck, Christian Cornelius. 1895. Waar werd de Indogermaansche stamtaal gesproken? *Tijdschrift voor Nederlandsche Taal- en Letterkunde* 14: 69–74.
- Uhlenbeck, Christian Cornelius. 1897. De voorgeschiedenis der Indogermaansche volken. *Taal en Letteren* 7: 1–25.
- de Vaan, Michiel. 2003. *The Avestan Vowels*. Amsterdam: Rodopi.
- Weiss, Michael. 2011[2009]. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor MI: Beech Stave Press.

Language index

A

Afroasiatic 3–4, 7, 16, 22, 218,
228–233

see also Proto-Afroasiatic

Abia 167

Abkhaz-Adyghe 216

Abui 167

Aeka 167

Afaqina 167

Agardabi 167

Ainu 71, 100, 119

Ainuic 96

Aisi 168

Akkadian 227

Akuna 167

Alanic 276, 282

Alaskan Inuit 70

Albanian 271, 281, 296, 300–301

see also Proto-Albanian

Alekano 167

Aleut 10, 14, 47–57, 61–62, 64,
67, 69, 218

Aleutian 47

Alor-Pantar 162, 167–169, 180

Altaic 2, 12, 15–16, 93–94,

97–100, 132, 219, 225, 229

see also Proto-Altaic,

Transeurasian

Alutiiq 14–15, 18, 47–51, 62–70

Amami 104

Amuric 1, 96

Anatolian 270, 272, 292–293,
296, 306–307

Aneme Wake 168

Angal Heneng 165, 172

Angan 173

Anim 161, 171, 178, 181

Anjam-Lalok 168

Apali 168

Arabic 227–228, 284, 309

Arawakan 2

Armenian 261–262, 266,

268–269, 274, 276, 301,
310–311

see also Old Armenian

Aslian 195, 197, 202

Assamese 184, 193

Atharva-Vedic 309

Atkan (Unangam Tunuu) 50,
55, 60, 66

Attuan (Unangam Tunuu) 50

Australian 21, 156, 158, 175,
178–179, 181

Austro-Tai 200–203, 212

Austroasiatic 1, 3–4, 7, 9, 16,
19–20, 183, 187, 192, 200–
208, 217

see also Proto-Austroasiatic

Austronesian 1, 7, 9, 16, 21, 156,
158–159, 168–169, 175, 178,
180, 217, 222, 293

see also Proto-Austronesian

Avestan 265, 271–272, 276, 279,
282, 286, 297, 311

see also Old Avestan, Late

Avestan

Aymara 4, 19–20

see also Proto-Aymara

Aymaran 2, 25–35, 37, 41–42

Azerbaijani 94, 97

B

Bactrian 276, 280

Bahnaric 197

Bái 188

Baimak 168

Bajui 283

Balkar 118

Balochi 282

Balto-Slavic 283

Bantu 1, 3–4, 11, 17–20, 22, 158,
235–258

see also Central-Western

Bantu, Eastern Bantu,

Kwilu Bantu, North-

Western Bantu, Proto-

Bantu, South-Western

Bantu, West-Coastal Bantu

Bao'an 105, 107, 113–114, 118

Barikiwa 167

Baruga 167

Baruya 173, 176

Bashkir 94, 97

Basque 16, 215, 222–224, 229,
231, 233

see also Proto-Basque

Bauwaki 168

Beja 227

Bembe 244

Benabena 167

Bengali 184

Berber 226, 228–230

Biangi 167

Binahari 168

Binandere 167

Biyom 167

Black Lahu 188

Bodo-Koch 188, 207

Bomberai 162, 180

see also Proto-Bomberai

Bosavi 173

see also Proto-Bosavi

Brahmaputran 201

Breton 280

Brythonic 280

Bulgar 127, 129–131, 133–134, 137,
142, 153

Bunaq 167–168

Buriat 105, 107, 113, 118

Burmo-Qiangic 201

Burum 167

Burushaski 222

C

Celtic 110, 262–263, 281, 286,
296, 308–309, 311

see also Common Celtic

- Central-Western Bantu 242,
246, 248, 253
- Chadic 225–230, 233
see also Proto-Chadic
- Chinese 14, 134, 137, 201
see also Middle Chinese, Old
Chinese
- Chukchee-Kamchatkan 218
- Chuvash 118, 126, 129–130,
134, 141
- Cokwe 244, 247–248
- Common Celtic 296, 309
- Common Germanic 309
- Common Indo-Iranian 271,
307
- Common Iranian 278, 309
- Common Italic 302, 309
- Common Slavic 281, 309
- Common Tocharian 299
- Common Turkic 127, 129,
133–134, 136, 141
- Core Indo-European 291, 300,
303, 307
- Cushitic 228
see also Proto-Cushitic
- Czech 261
- D**
- Dadibi 167
- Daga 167
- Dagur 105, 107, 113, 118, 120
- Dani 165, 177
- Dano 167
- Danube Bulgar 130
- Dem 167, 289
- Dena'ina 51, 68, 70, 72
- Dene-Caucasian 222, 231
- Dibiyaso 168
- Dimasa 188
- Dolgan 113
- Dongxiang 114, 118
- Doriri 167
- Doromu 167
- Dravidian 2, 21, 215–216, 230,
232
see also Proto-Dravidian
- Dugeme 167
- Duma 249–250
- E**
- Eastern Bantu 244, 248, 253
- Eastern Iranian 128, 131, 134,
281, 283–284
- Eastern Timor 168
- Edolo 173
- Egyptian 226, 228
- Eipomek 172, 177
- Elamite 230, 280
- Engan 165, 171–172
- English 18, 118, 120, 184, 262,
266, 268, 274, 280
see also Old English
- Eskimo 10, 14–15, 18–20,
47–62, 64–69
see also Proto-Eskimo
- Eskimo-Aleut 2–4, 14–15,
19–20, 47–57, 61–62, 64,
69–70, 218
see also Proto-Eskimo-Aleut
- Ethiosemitic 227–228
- European Romani 282
- Euskaro-Caucasian 16, 223–
224, 228–229
- Evenki 107, 114, 118
- Ewage-Notu 167
- Eyak 51, 70, 72
- F**
- Fataluku 167–168
- Finisterre-Huon 167, 171, 173,
180
- Finnic 287
- Finnish 266, 287, 309
- Finn-Ugric 218
- Folopa 167
- Fore 167, 180
- Formosan 119
- Fwe 248
- G**
- Gagauz 118
- Gaina 167
- Gallo-Romance 309
- Garo 188
- Garus 168
- Gaulish 309
- Gebi 167
- Germanic 110, 216, 232, 263,
266, 282–283, 292, 298,
300, 308
see also Common Germanic
- Gimi 167
- Goilalan 167, 171
- Goldi 118
- Gongduk 201
- Goroka 167, 171
- Gothic 110, 263, 266, 309
- Greek 261–263, 267–270, 281,
283, 286–287, 290, 295, 298,
300–303, 305, 308
see also Homeric Greek,
Mycenaean Greek
- Guhu-Samane 167
- Gwahatike 167
- Gwedena 167
- H**
- Hamitic 226
- Helleno-Armenian 262
- Hezhe 118
- Himalayan-Yangtzean 201
- Hirara (Miyako) 113, 115
- Hittite 261–262, 264, 268–269,
271, 273, 280, 282, 287, 291,
295–296, 299, 307
- Hmong-Mien 3–4, 7, 9, 14,
16, 20, 183, 186–190, 195,
200–203, 205–207
see also Proto-Hmong-Mien
- Hmongic 187–188
- Ho Nte 187
- Homeric Greek 264, 271
- Hungan 237–238, 244, 247,
250, 256
- Hungarian 129–131, 143
- Hunjara-Kaina Ke 167
- I**
- Ibero-Caucasian 223
- Igarashi (Yaeyama) 113
- Iha 167–168
- Indo-Aryan 275–277, 280–285,
295, 307
see also Middle Indo-Aryan,
Old Indo-Aryan
- Indo-European 1–4, 7, 9, 12,
14, 16–22, 72, 90, 110, 131,
134–135, 159, 176, 179,
215–218, 224–225, 229–233,
259, 275, 279–281, 286–290,
306, 310–311
see also Core Indo-European,
Proto-Indo-European

- Indo-Iranian 3–4, 12, 14, 17,
271–272, 275–280, 282,
286–290, 293, 295, 297–
299, 303–304, 307–309
see also Common Indo-
Iranian, Proto-Indo-
Iranian
- Indo-Uralic 219, 232
see also Proto-Indo-Uralic
- Indonesian 156, 165, 180
- Inland Gulf 161
- Inoke-Yate 167
- Inuit 47, 50, 53, 55, 59–61, 65, 71
see also Alaskan Inuit, Proto-
Inuit
- Iranian 3–4, 12, 14, 17, 146,
271–272, 275–289, 293, 295,
297–299, 303–305, 307–310
see also Common Iranian,
Eastern Iranian, Middle
Iranian, Old East Iranian,
Old Iranian, Proto-
Iranian, Western Iranian
- Irish 262
see also Middle Irish, Old
Irish
- Ishigaki (Yaeyama) 104, 115
- Italian 199, 309
- Italic 262–263, 272, 308
see also Common Italic
- Italo-Celtic 262–263
- Iyaa 245, 257
- J**
- Japanese 15, 83–92, 97, 101, 103,
106–107, 110, 116, 119–120,
142–143, 205, 207–208, 232
see also Middle Japanese, Old
Japanese, Proto-Japanese
- Japanic 96
- Japano-Koreanic 3–4, 12,
19–20, 75–78, 80, 83, 86,
90–91, 99–100, 103–104,
124–125, 135
- Japonic 15, 76, 87–88, 90,
93–95, 100–102, 104–105, 107,
111–116, 118, 125, 141, 218
- Jassic 282
- Jimajima 167
- Jinuo 188
- Jurchen 108, 118
- K**
- Kabira 115
- Kaera 167–168
- Kalmuck 105, 107, 113, 118, 141
- Kaluli 168
- Kamang 168
- Kamano 167
- Kambaata 227
- Kanyok 244, 256
- Karaim 113–115, 146
- Karakalpak 113
- Karakhanid (Old Turkic) 104,
115
- Kare 168
- Karluk 134
- Kartvelian 216, 218, 230
- Kaser 167
- Kâte 167
- Katuic 197
- Kazakh 118, 129, 141
- Khakas 97
- Khalaj 104, 113–115
- Khalkha 105, 113, 118
- Khasi-Aslian 195, 197, 202
- Khasi-Pakanic 195, 202
- Khasian 197
- Khmeric 197
- Khmuic 197
- Khotan Saka 128, 131, 141, 289
- Khotanese 276, 279, 282–283,
309
- Khowar 281
- Kikongo Language Cluster 238,
243–245, 248–250, 253, 255
- Kikongoid 238
- Kimbundu 243–244, 247–248
- Kiranti 201
- Kirghiz 118
- Klon 167–168
- Koguryoan 87–88
- Korafe 167
- Korean 11, 15, 86–90, 97, 100–
103, 105, 110, 114, 119–120,
141–143, 205, 207, 218, 232
see also Middle Korean, Old
Korean, Proto-Korean
- Kradai 9, 16, 189
- Ku Waru 167
- Kuman 167
- Kunimaipa 167
- Kur-Urmi (Nanai) 106
- Kwalean 167
- Kwamashi 248
- Kwilu Bantu 238
- Kypchak 130
- L**
- Laadi 250
- Laali 250
- Lak 10
- Lake Murray 161
- Larestani 282
- Latege 245, 256
- Latin 110, 223, 261–264, 267–
269, 282, 284, 302–303
- Lezghian 10
see also Proto-Lezghian
- Lingala 243, 258
- Lithuanian 110, 261, 269, 300,
309
- Lower Fly River 161
- Lucazi 243
- Lwel 243
- M**
- Madang 167, 177
see also Proto-Madang
- Mâgobineng 167
- Maiagolo 167
- Mailu 168
- Maiwa 167
- Makalero 167–168
- Makasae 167–168
- Malayo-Polynesian 201
- Manchu 106–108, 114–115,
118–120, 134, 137–138, 143,
152, 221
- Mape 167
- Mapena 167
- Maranomu 167
- Mari 131, 147, 153, 281
- Maria 143, 167
- Marind 161, 171, 178
- Marind-Yaqay 161
- Maring 167
- Mawan 168
- Mbaham 167–168
- Mbam-Bubi 246–247
- Mbede 247
- Mbuun 237–238, 244, 256
- Mek 109, 172
- Melpa 167

- Middle Chinese 91, 188
 Middle Indo-Aryan 276
 Middle Iranian 276
 Middle Irish 286
 Middle Japanese 76, 79, 118
 Middle Korean 12, 76–78,
 80–81, 83–86, 88, 92, 108–109,
 118
 Middle Mongolian 105, 118
 Middle Persian 276, 282, 309
 Middle Turkic 118
 Mienic 187–188
 see also Proto-Mienic
 Migabac 167
 Miyako 104, 113
 Moghol 105, 118
 Moikodi 168
 Momare 167
 Mon-Khmer 195, 202, 205
 Mongo 247
 Mongolic 12, 15, 18, 93–97,
 104–105, 107, 109, 111–116,
 123–126, 132, 134–140, 146,
 148, 151, 218, 220–221
 see also Proto-Mongolic
 Monguor 104–105, 107, 113–115,
 118
 Monic 197
 Mor 167–168
 Mordva 281, 287
 Mosimo 168
 Mouwase 167
 Mpiem 247, 258
 Mpiin 237–238, 244, 248–250,
 256
 Mpur 243
 Mulaha 167
 Mum 168
 Munda 7, 195, 197, 202
 Munji 290
 Murupi 168
 Mycenaean Greek 288
- N**
 Na-Dene 222, 233
 Na-e 187
 Naga 79, 167
 Nake 168
 Nakh 10, 216
 see also Proto-Nakh
 Nakh-Daghestanian 216, 223
 see also Proto-Nakh-
 Daghestanian
 Nanai 106–107, 114, 118, 138, 152
 Nawaru 168
 Naze (Amami) 104
 Ndumu 249
 Nedebang 167–168
 Negidal 106–107, 114, 118
 Nend 168
 Nepali 184
 Ngong 237–238, 244–245, 250,
 256
 Ngwi 243
 Nicobarese 197
 Nivkh 1, 100
 Nogai 97, 118
 North Caucasian 10–11, 13,
 215–216, 221, 226, 230, 233
 see also Proto-North
 Caucasian
 North-Western Bantu 246–247,
 253
 Northwest Indo-Aryan 281
 Nostratic 3–4, 12, 16, 217–221,
 225, 229, 231–232
 see also Proto-Nostratic
 Nsong 237–238, 243–245, 247,
 249–250, 256
 Ntandu 243–245, 248–250, 254
 Nuristanic 275, 280–284
 Nusu 188
 Nzadi 243, 254
 Nzebi 239, 243–244, 248–250
 Nzebi-Mbete-Teke 239, 243–
 244, 248–250
- O**
 Oghuz 129, 148–149
 Oirat 115
 Oirata 167–168
 Oiso 167
 Ok 161, 171, 180–181
 Ok-Awyu 161
 Okinawa 87, 104, 113
 Olcha 106–107, 113–115
 Old Armenian 309
 Old Avestan 309
 Old Chinese 89, 91, 127–128,
 135, 143, 187–189, 195, 205, 208,
 211, 222, 281
 Old Church Slavonic 110, 309
 Old East Iranian 127
 Old English 110, 261, 269, 309
 Old French 309
 Old High German 110, 269
 Old Icelandic 266
 Old Indo-Aryan 276
 Old Iranian 276, 290
 Old Irish 110, 261, 268, 309
 Old Japanese 75–82, 84, 86, 88,
 109, 114, 118
 Old Korean 86–87
 Old Norse 264
 Old Persian 276, 280, 286
 Old Prussian 110, 309
 Old Swedish 266
 Old Turkic 118, 126, 143
 see also Karakhanid (Old
 Turkic), Old Uyghur
 Old Uyghur 134
 Old Welsh 268
 Omotic 225–226, 228, 233
 Onabasulu 168
 Ono 167
 Ordos 105, 111, 113, 115, 127, 141
 Ormuri 290
 Oroch 107–108, 112–115
 Orok 106–108, 112–115
 Orokaiva 167
 Ossetic 282, 299
 Otomanguean 21–22
 see also Proto-Otomanguean
 Oura (Miyako) 104
- P**
 Pacific Coast Yupik 47–49
 Pakanic 195, 202
 Palaungic
 Pali 276
 Pamir 282–283, 290
 Papuan 16, 155–156, 158–160,
 168–169, 171, 174, 176–177,
 179–181
 Parthian 276, 280, 286
 Pashto 271, 282
 Pearic 195, 209
 Permian 281, 287
 Pero 228
 Persian 278, 281, 309
 see also Middle Persian, Old
 Persian

- Proto-Afroasiatic 215, 225–227, 229–231
 Proto-Albanian 309
 Proto-Altaic 98, 123–126, 134–141, 145, 150, 152–153, 218, 220–221, 223, 231
 Proto-Andi 223
 Proto-Austroasiatic 195–198, 206
 Proto-Austronesian 116
 Proto-Avar-Andi 224
 Proto-Aymara 14, 25–39, 41–42, 44
 Proto-Baltic 282
 Proto-Bantu 240–244, 246–247, 251, 254
 Proto-Basque 223
 Proto-Benue-Congo 244
 Proto-Binandere 167
 Proto-Bomberai 167–168
 Proto-Bosavi 168
 Proto-Chadic 225
 Proto-Chimbu-Waghi 167
 Proto-Cushitic 225–227
 Proto-Dagan 167
 Proto-Dravidian 218
 Proto-Duna-Bogaia 168
 Proto-Eskimo 53–54, 58, 70–71
 Proto-Eskimo-Aleut 53, 71
 Proto-Finisterre-Huon 167
 Proto-Goilalan 167
 Proto-Gorokan 167
 Proto-Hmong-Mien 2, 187
 Proto-Hmongic 187–188
 Proto-Huon 167
 Proto-Indo-European 11, 18, 42, 44, 141, 220, 223, 230, 232, 260–274, 276–277, 291–293, 295–296, 298–299, 302–303, 307–309
 Proto-Indo-Hittite 273
 Proto-Indo-Iranian 3, 275–278, 286–287, 308
 Proto-Indo-Uralic 220
 Proto-Inuit 70
 Proto-Iranian 141, 278, 286
 Proto-Japanese 78–81, 85, 220–221
 Proto-Japonic 87, 90, 100–102, 114, 118, 141
 Proto-Japono-Koreanic 77, 98, 104, 125
 see also Proto-Korean-Japanese
 Proto-Kainantu 167
 Proto-Kainantu-Goroka 167
 Proto-Korean 75, 77–78, 82–85, 87–91, 221
 see also Proto-Koreanic
 Proto-Korean-Japanese 75, 77, 82, 84–85, 87–91
 see also Proto-Japono-Koreanic
 Proto-Koreanic 100, 114, 118, 141
 see also Proto-Korean
 Proto-Lezghian 224
 Proto-Mabuso 167
 Proto-Madang 168
 Proto-Mailuan 168
 Proto-Manubaran 167
 Proto-Mienic 187–188
 Proto-Mongolic 13, 118, 125, 137, 141, 220–221
 Proto-Mongolo-Turkic 98
 Proto-Nakh 223–224
 Proto-Nakh-Daghestanian 224
 Proto-North Caucasian 16, 222–224, 229
 Proto-Nostratic 215, 218–219
 Proto-Otomanguean 2
 Proto-Quechua 11, 14, 25–42, 44
 Proto-Rai Coast 167
 Proto-Ryukyuan 118
 Proto-Samoyedic 127
 Proto-Semitic 225–228, 266
 Proto-Sinitic 188
 Proto-Sino-Caucasian 215, 222
 Proto-Sino-Tibetan 125, 208, 222
 Proto-Slavic 261, 282
 Proto-Timor-Alor-Pantar 162, 167–168
 Proto-Tocharian 141
 Proto-Trans-New Guinea 10, 157, 160–161, 163, 166–169, 173–174
 Proto-Transeurasian 15, 93–96, 98–99, 102–105, 109–110, 116–118, 123–125, 139, 232
 Proto-Tsez 223
 Proto-Tungusic 112, 115, 118, 125, 141, 221
 Proto-Turama-Kikororian 167
 Proto-Turkic 11–13, 15, 110, 118, 123–141, 143–144, 220–221
 Proto-Turko-Mongolic 125, 136
 Proto-Ugric 127
 Proto-Uralic 219–220, 230, 232
 Proto-West-Coastal Bantu 17, 235, 237, 239, 241–252
 Proto-Yareban 167–168
 Proto-Yeniseian 127
 Proto-Yupik 62, 70
 Punu 244, 250
- Q**
 Quechua 4, 19–20, 25–44
 see also Proto-Quechua
 Quechua-Aymara 14, 27–30, 44
 Quechuan 2, 25–35, 37, 40–43
 Quechumara 29, 44
- R**
 Raiapu Enga 171
 Raptling 168
 Rempi 168
 Romance 29, 110, 223, 257, 302, 309
 Romanian 309
 Russian 53, 62, 66, 68, 99–100, 111, 119, 142, 215, 233
 Ryukyuan 87, 90, 119
 see also Proto-Ryukyuan
- S**
 Saka 128, 131, 141, 276, 289
 Salt-Yui 167
 Sam 168, 254
 Samba 244
 Samosa 168
 Samoyed 218
 see also Proto-Samoyedic
 Sanskrit 110, 263, 269–271, 274, 276
 see also Vedic Sanskrit
 Santa 118
 Saruga 168
 Sawila 167
 Sayan 127, 129
 Scythian 285, 290

- Semitic 1, 7, 21, 215–216, 229–231, 306
see also Proto-Semitic
- Sene 167
- Serbian 261
- Sgaw Karen 188
- Shira-Yughur 113–114
- Shodon 115
- Shona 248, 255
- Shor 104
- Shuri (Okinawa) 104, 113–116
- Sialum 167
- Siane 167, 171
- Sibe 106–108, 110, 113
- Sillan 87
- Silopi Bagupi 168
- Sinitic 2, 14, 96, 187, 201, 207
see also Proto-Sinitic
- Sino-Bodic 188
- Sino-Caucasian 4, 10, 16, 23, 221–223, 225, 231, 233
see also Proto-Sino-Caucasian
- Sino-Tibetan 1, 9–10, 16, 22, 158, 230, 232–233
see also Proto-Sino-Tibetan
- Sirva 168
- Slavic 110, 262, 283, 287, 308, 310
see also Common Slavic, Proto-Slavic
- Sogdian 276, 282, 286
- Solon 108, 114, 138, 152
- South-Western Bantu 242–244, 246, 253
- Suena 167
- Shughni 281–282, 290
- Suku 243, 254
- Sumerian 218, 230, 269
- T
- Tai-Kadai 1, 158
- Tanahmerah 168
- Tangkhul 188
- Tatar 113, 118
- Tauya 167
- Teberan 167
- Tehit 169
- Teiwa 168
- Teke 129, 136, 152, 239, 243–244, 248–250, 255
- Telefol 171
- Tetela 248
- Tibetan 1, 9–10, 16, 22, 125, 158, 194, 205, 207–209, 211, 214, 222, 230, 232–233
- Timor-Alor-Pantar 169
see also Proto-Timor-Alor-Pantar
- Tocharian 127, 132, 278, 281, 297, 303, 306, 308–310
see also Common Tocharian, Proto-Tocharian, Tocharian B
- Tocharian B 271, 280, 289
- Tofa 149
- Tok Pisin 165
- Tokano 167
- Trans-Himalayan 188, 193, 200–203, 205–207, 209
- Trans-New Guinea 2–4, 7, 9, 12–13, 16, 19–20, 155, 158–169, 178
see also Proto-Trans-New Guinea
- Transeurasian 2–4, 7–9, 12, 16, 18–20, 97, 100, 119–120, 126, 135–137, 140, 142–143, 219
see also Proto-Transeurasian
- Tsogo 247
- Tumshuqese 276
- Tungus 118
- Tungusic 12, 15, 93–97, 100, 102, 105, 107, 109–116, 120, 123–126, 132, 135–138, 140–143, 148, 151, 218–219, 232
see also Proto-Tungusic
- Tupian 2
- Turama-Kikori 168
- Turkic 2, 18–19, 93–98, 104–105, 113–116, 120, 142–153, 218, 232, 279, 281, 285
see also Common Turkic, Middle Turkic, Proto-Turkic
- Turkish 110, 118, 129
- Turkmenian 118, 129, 141
- Turko-Mongolic 124, 221
see also Proto-Turko-Mongolic
- Turumbu 247
- Tuva 114
- U
- Udehe 107, 114, 118
- Udmurt 280
- Uduri 167
- Ulcha 112, 138, 152
- Unangam Tunuu 14–15, 18, 47–50, 53–57, 60–71
- Unangan 47–48, 50–54, 56, 60–61, 64, 67–70
- Uralic 12, 218, 222, 278–281, 289–290, 310
see also Proto-Uralic
- Utu 168
- Uighur 118, 129, 134, 141
- Uzbek 110, 118
- V
- Vaantura 167
- Vedic 272–273, 276, 283, 286, 290, 297–298, 307, 311
see also Vedic Sanskrit
- Vedic Sanskrit 264–265, 268, 309
- Vietic 197
- Vili 244
- W
- Waffa 167
- Wagi 168
- Wamas 168
- Wamorà 167
- Warkay-Bipim 161
- Watiwa 168
- Wakhi 281–282, 290, 299
- Welsh 110, 280–281
see also Old Welsh
- Weri 167, 171
- Wersing 167–168
- West Caucasian 10
- West Greenlandic 70
- West Makian 169
- West Pantar 167–168
- West-Coastal Bantu 3–4, 11, 19–20, 238, 253
see also Proto-West-Coastal Bantu
- Western Iranian 280, 284
- Wiru 172
- Wongbe (Sam) 168
- Written Mongolian 105, 118, 139, 141

X

Xufi 284
Xwarezmian 276, 282

Y

Yaeyama 87, 104, 113
Yagaria 167
Yakut 113–114, 129, 134
Yali 165, 178
Yangtzean 187, 205, 207
Yans 243, 249
Yanzi 238, 243–245, 248–250

Yareba 167–168
Yareban 169
 see also Proto-Yareban 167–
 168
Yaweyuha 167
Yazghulami 282
Yegha 167
Yeniseian 72, 128, 144, 221–222,
 233
 see also Proto-Yeniseian
Yidgha 290
Yidgha-Munji 290

Yoidik 168
Yombe 249–250
Yonaguni 104, 113, 115
Yopno 171, 173
Yupik 14, 50–51, 53, 55, 61,
 65–68
 see also Pacific Coast Yupik,
 Proto-Yupik

Z

Zia 167

Subject index

A

- Anatolian
 - hypothesis 2
 - theory 22, 232, 292–293
- animal
 - domestication 240
 - husbandry 136, 256, 270, 275, 277
- agricultural
 - package 10, 101, 155, 157, 166, 173–174
 - technology 275, 286–287
- agropastoralism 25–26, 28, 40–42
- ahu 183, 191, 193, 206–207
 - rice 184–185, 190
- arboriculture 251

B

- Bambara groundnut 235, 239–241, 245, 250, 252
- banana 10–12, 16, 155–157, 163, 166–171, 173–176, 179, 239, 242–243, 251, 256
- barley 133–135, 139–140, 144, 221–222, 227, 262–263, 266, 269–272, 280–282, 287, 292
- basic vocabulary 31, 75–77, 94, 96, 121, 124–125, 176
- Bayesian 8, 95, 97–98, 100, 125, 143
 - phylolinguistics 96, 120
 - inference 96, 117
- betel nut 155
- breadfruit 156, 162, 181
- broomcorn millet 98, 105–106, 108, 110, 134–135, 138, 146, 149, 153, 192, 283, 293
- buckwheat 15, 75–77, 84–87, 133, 153

C

- calibration 96
- cassava 156, 176, 239, 252
- cattle 17, 128, 130–131, 135, 138, 152–153, 272, 275, 277, 284, 293, 310
- cheese 11, 131–132, 147, 150, 294–295, 297, 304–306, 308, 311
- chicken 155, 176, 180
- climate 17, 47, 51, 68–69, 99, 177, 242
 - change 18, 100, 117, 139, 180, 236, 251, 254
- Comparative Method 30, 91, 97, 159, 161–162, 166, 174, 179, 272–273
- contact linguistics 6, 12–13, 92, 127
 - see also* language contact
- corn 34, 37, 40, 133–134, 138, 146, 152, 262, 269–270, 280–282, 286–287
- cowpea 235, 240–241, 245, 250, 252
- cultural reconstruction 6, 9–10, 15, 17, 21, 103, 110, 117, 123–126
 - see also* linguistic archaeology, linguistic ethnobiology, linguistic paleontology, paleobiolinguistics, Wörter und Sachen

D

- dairying 17–19, 293–294
- demographic
 - dispersal 6, 8–9
 - transition 6, 9, 93, 95
- demography 4, 6, 9, 15, 19, 102, 117, 291–292, 295, 308
- Demography/Subsistence Model 18–19

- Densi Tree 97
- directionality 14, 31, 33, 37, 63, 66–67
- disease 7, 18
- distribution of cognates 10, 47–48, 54, 57, 60–61, 266
- Diversity Hotspot Principle 6–8, 15, 95
- DNA 73, 100, 142, 157, 164, 178–180, 199, 211, 291, 294
- dog 54, 60, 108, 150
- domestic 137, 143, 179, 258, 290
 - see also* domesticated, domesticate, domestication
 - animals 38, 131, 135–136, 138, 276–277
 - donkeys 137
 - pigs 136, 155
 - production 39
 - ungulates 128
- domesticate 38, 241
 - see also* domestic, domesticated, domestication
 - animals 26, 33, 37–38, 163, 218, 240
- barley 292
- camelid 38
- cereals 293
- crops 11, 241, 245, 251
- pearl millet 239
- plants 26, 235, 237, 251
- reindeer 138
- rice 183–186, 189–196
- taro 164–166
- wheats 292
- yams 240

- domestication 17, 19–20, 98–99,
125, 173, 207, 217, 239–240,
243, 245, 251
see also domestic,
domesticated, domesticate
center 157, 164
history 63, 169
of banana 170
of barnyard grass 110
of millet 102
of rice 16, 183–186, 189–196,
198, 206
of sugarcane 169
of taro 164–166
process 183
plant 45, 155–157, 163–164,
173, 175
- E**
ecological 14, 33, 35, 42, 117
change 18
complementarity 25, 43
zones 25, 27, 38, 43
ecology 27, 92, 175, 196, 207, 210
ecosystem 50, 72, 252
Elite Dominance Model 18–19,
294
ethnography 17, 52, 70, 72–73,
177, 258
- F**
farming package
see agricultural package
female terminology 70
field rice 15, 77, 87, 89, 103
fish 50, 55, 58, 60, 63, 65, 69, 116
foraging 17, 19–20, 101, 119, 174,
227–228, 232, 235
foxtail millet 82, 98, 134–135,
138, 149, 189, 192, 194, 208,
212, 284, 293
- G**
genealogical relatedness 15,
93–94, 96, 116, 158–159, 161
gender 10, 51, 55, 61–62, 67, 70,
120, 200, 202, 301
glottochronological dating 127,
216, 223–225
goat 17, 129, 131–132, 135–137,
147–148, 151–152, 197, 275,
277–278
- groundnut 228
see also Bambara groundnut
guava 34–35, 42
- H**
haplogroup 199, 200–207, 209,
211, 213–214
herding 14, 25–27, 32–33, 38–39,
42, 126, 226, 310
Historical-Comparative Method
see Comparative Method
horse 17, 128, 130–131, 137, 144,
146, 152–154, 275, 277,
278–279, 304
meat 127
pastoralism 13, 19, 138
riding 18, 126, 308
hunter-gatherer 1–2, 11, 41,
52, 295
husbandry *see* animal
husbandry
- I**
immunity 19
instability 165–166
indica rice 183–186, 190–191,
194, 198, 206–207, 209
inflection 268
- J**
japonica rice 2, 92, 183–186,
190–191, 194, 198, 206–209
- L**
lactase persistence 18–19,
294–295, 307–310
language
contact 18, 22–23, 25–26, 28,
30–31, 43, 47–49, 51, 66,
71–72, 82, 124
maintenance 18
replacement 6, 15, 19, 70
shift 18–19, 27, 294
lexical
analysis 259–260, 267–268,
271, 273
domains 26, 42, 70
recycling 10–11, 15, 75, 77,
84–85, 89–91, 123, 139
lexicostatistics 8, 127, 219
linguistic
archaeology 9, 23, 103
see also cultural
reconstruction,
linguistic ethnobiology,
linguistic paleontology,
paleobiolinguistics,
Wörter und Sachen
ethnobiology 9, 22
see also cultural
reconstruction,
linguistic archaeology,
linguistic paleontology,
paleobiolinguistics,
Wörter und Sachen
paleontology 9, 103, 215, 219
see also cultural
reconstruction,
linguistic archaeology,
linguistic ethnobiology,
paleobiolinguistics,
Wörter und Sachen
long-range comparison 4, 215
- M**
macrofamily 2, 10, 16, 215–218,
222, 225, 231
maize 21, 26, 28, 33–35, 37, 42,
45, 133–134, 149, 156, 175, 227,
239, 242, 252
Mapping Demographic
Dispersal on Linguistic
Phylogeny 6, 8–9
millet 10, 15, 75–77, 81, 84–87,
89–91, 107, 109, 115, 133,
136, 139–140, 144, 187–188,
205, 220–222, 227, 281,
286, 292
see also broomcorn millet,
domestication of millet,
foxtail millet, pearl millet
agriculture 98–102, 117,
consumption 128
cultivation 94, 98–99,
102–103, 117, 138
farming 125, 128
milk 9, 11, 18, 131–132, 136–138,
150–151, 154, 272–273, 279, 291,
293–308, 310–311
mobility 18–19, 173
morphological complexity 14,
55–56, 64, 105, 108, 137, 140
mythology 259, 271–272

- N**
- Neolithic 20, 93–94, 100, 111, 115–116, 210, 212, 214, 219, 225
 culture 99, 125
 farmer 294
 site 110
 transition 95
 vocabulary 275, 277
- nomadic pastoralism
see pastoralism
- O**
- okra 227, 235, 241, 243–245, 250, 252
- Oryza sativa 88–89, 155, 183–186, 193, 196, 198, 206, 210–211, 239, 283
- P**
- paper mulberry 155, 175, 178
- paddy rice 15, 77, 91, 103
- paleobiolinguistics 10, 21
see also cultural reconstruction, linguistic archaeology, linguistic ethnobiology, linguistic paleontology, Wörter und Sachen
- Panicum milliaceum 98, 106, 119, 134–135, 192, 283
see also broomcorn millet
- paradigm 45, 71, 160, 217, 286, 301
- pastoralism 12, 17, 31, 38, 95, 123–124, 126, 128, 132, 140, 220, 277
see also agropastoralism, horse pastoralism
- camelid 39
- nomadic 95, 138
- related vocabulary 15, 131, 137, 139
- pearl millet 235, 237, 239–242, 245, 250–252
- phylogeography 175, 177, 183, 199–200, 211, 257
- phylolinguistics 6, 8, 15, 96, 120
- phylogeny 6–9, 15, 22, 102, 178, 187, 195, 197, 201, 213, 232, 237
- phytolith 239, 242, 251
- pig 38, 125, 131, 136, 138–139, 142, 146–147, 153, 178, 277, 279
- plague 19
- plantain 235, 239, 241–242, 245, 250
- plant domestication
see domestication
- Pontic Steppes 292, 294–295
- population
 genetics 186, 209
 growth 1, 295
 pressure 18
- potato 34, 35–37, 156, 162, 165, 171, 180, 239, 252
- pottery 54, 100, 103, 116–117, 156, 241
- preservation bias 183, 186, 190–192
- prestige 15, 27, 69, 72
- pronoun 49, 160, 179
- pronominal paradigm 160, 180
- R**
- religion 17, 75, 259, 271–272
- rice 10–11, 14, 19, 76, 78–79, 81, 86, 98, 102, 106, 153, 188, 202, 244, 281, 293
see also ahu, ahu rice, domesticated rice, domestication of rice, field rice, indica rice, japonica rice, Oryza sativa, paddy rice, wet rice
- agriculture 3, 15, 77, 82–84, 87, 89–90, 101, 103–104, 187, 190–192, 194, 204–206
- vocabulary 2, 12, 15, 75, 77, 80, 82–83, 85, 89, 90
- ritual 75, 106, 177, 180, 248, 252, 259, 271–272, 293, 295, 308
- S**
- semantic domain 10, 12, 54–55, 66
- semantic stability 11, 13
- Setaria italica 82, 84, 98, 119, 134–135, 192, 212, 282, 284
see also foxtail millet
- sewing 54, 57–58, 61, 63, 67, 69, 103, 113–114, 113
- sheep 17, 128–129, 131, 137, 149, 151, 275, 277–278
- small-scale cultivator 15
- social status 70
- sorghum 133–134, 251
- stability 165, 167, 174, 180
see also semantic stability
- Steppe hypothesis 2, 18, 42, 44, 292–293, 309
- Subsistence/Demography Model
see Demography/Subsistence Model
- subsistence
 -driven language spread 2, 15, 17
 strategy 2, 4–6, 8, 19–20, 68, 116–117, 119, 123–124, 131, 246, 252
 terminology 15, 47–48, 52–53, 57, 61–62, 67, 69–70
- substrate 16, 191–192, 216, 225, 229, 232, 266, 282, 289
- substratum *see* substrate
- sugarcane 10–13, 16, 155–157, 162–163, 166–175, 177
- system collapse 18
- T**
- taro 10, 12–13, 16, 155–157, 162–163, 171, 174–178, 180–181, 239, 242
see also domestication of taro
- tuber 33–35, 40, 42, 165–166, 177, 239
- V**
- vegeculture 155, 163
- vegetables 17, 53, 133, 162, 275, 283–285
- volcanic activity 18, 50–51, 68
- W**
- warfare 15, 19, 47–48, 51, 68–69, 72, 200
- wheat 53, 84, 86, 133–135, 139–140, 145, 205, 222, 224, 226–227, 262, 281–283, 287
see also domesticated wheats 292
- weaving 33, 39, 54, 57, 60, 63, 66–67, 69, 103, 112–114, 117
- wet rice 82, 89, 91, 101
- word formation 120

Wörter und Sachen 9, 103, 220,
231, 260
see also cultural
reconstruction,
linguistic archaeology,
linguistic ethnobiology,

linguistic paleontology,
paleobiolinguistics

Yamnaya culture 9, 291–292,
294

Y

yam 156, 162, 165–166, 176, 178,
227, 239, 242, 252
see also domesticated yams
240

Why do some languages wither and die, while others prosper and spread? Around the turn of the millennium a number of archaeologists such as Colin Renfrew and Peter Bellwood made the controversial claim that many of the world's major language families owe their dispersal to the adoption of agriculture by their early speakers. In this volume, their proposal is reassessed by linguists, investigating to what extent the economic dependence on plant cultivation really impacted language spread in various parts of the world. Special attention is paid to “tricky” language families such as Eskimo-Aleut, Quechua, Aymara, Bantu, Indo-European, Transeurasian, Turkic, Japonic, Hmong-Mien and Trans-New Guinea, that cannot unequivocally be regarded as instances of Farming/ Language Dispersal, even if subsistence played a role in their expansion.

ISBN 978 90 272 1255 9



9 789027 212559

John Benjamins Publishing Company