Linguistic Atlas of Asia and Africa

I

Edited by
Hiroyuki Suzuki, Mika Fukazawa,
Akiko Yokoyama, and Mitsuaki Endo

Geolinguistic Society of Japan

Linguistic Atlas of Asia and Africa

I

EDITED BY
HIROYUKI SUZUKI,
MIKA FUKAZAWA,
AKIKO YOKOYAMA,
AND MITSUAKI ENDO

Cover photo: A mouse from Hakodate © 2001 Hitoshi Suzuki

Preface

This volume, *Linguistic Atlas of Asia and Africa*, Volume I, is a sequel to *Linguistic Atlas of Asia*, Hituzi Syobo, Tokyo, 2021. This is the direct outcome of a joint research project at the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies titled "Studies in Asian and African Geolinguistics" from the academic year 2020 to 2022, in collaboration with Grant-in-Aid for Scientific Research on Innovative Areas "Deciphering the History of Yaponesians through Comparison and Analyses between Japanese and the Other Concerned Languages" Project Number: JP18H05510 sponsored by MEXT, Japan, and other grants.

The most remarkable new characteristic of this volume is the full coverage of Africa: Niger-Congo or Bantu is the in-charge of SHINAGAWA Daisuke and KOMORI Junko; the languages in the Kalahari Basin Area are handled by NAKAGAWA Hirosi and KIMURA Kimihiko, and Nilo-Saharan by NAKAO Shuichiro. New specialists in some language families in Asia have joined as well: ONO Chikako for Chukotko-Kamchatkan, KODAMA Nozomi for Dravidian, IWASAKI Takamasa for Iranian, and TOMITA Aika and HIRANO Ayaka for Kra-Dai. The section on Caucasian languages by SUZUKI Hiroyuki is another innovation.

In Volume I, animal terms for 'rat/mouse, chicken, horse, dog (wolf, optional), and bear' are addressed. The criterion for the selection is those that seem to have a close relationship with human life. However, since some animals do not exist in some areas of Asia and Africa or are rarely described, it was impossible to draw maps for them; therefore, they are not included in this volume. They are as follows: 'horse' in Chukotko-Kamchatkan and the languages in the Kalahari Basin Area; 'wolf' in Sinitic, Hmong-Mien, Austroasiatic, Austronesian, Dravidian, Nilo-Saharan, Niger-Congo, and the languages in the Kalahari Basin Area; 'bear' in Nilo-Saharan, Niger-Congo, and the languages in the Kalahari Basin Area.

We researched the geolinguistic distribution in the AA area as DNA information from Asia is also available. In our meeting held on 4 September, 2021, the following presentations by geneticists were made: "Human impacts on the evolution of rats and mice" by SUZUKI Hitoshi (Hokkaido University); "Phylogeography of brown bears in the northern hemisphere" by MASUDA Ryuichi (Hokkaido University); "The evolutionary process of dogs domesticated from gray wolves" by TERAI Yohey (The Graduate

University of Advanced Studies, SOKENDAI); "Genetic diversity and relationships among European, Asian and Japanese horse breeds" by TOZAKI Teruaki (Genetic Analysis Department, Laboratory of Racing Chemistry, Japan); and "Origin and history of Japanese native chickens as inferred from mitochondrial DNA analysis" by YONEZAWA Takahiro (Tokyo University of Agriculture). Thanks are due to these scholars and especially Professor SUZUKI Hitoshi for cohosting and introducing them.

This series will be followed by Volume II, including crop terms, and Volume III, including stop series, grammatical relations, the system of 'sibling' terms, and numeral systems soon. A complete bibliography for primary data used for mapping will appear at the end of Volume III. Only references to cited works appear in this volume.

ENDO Mitsuaki

❧

TABLE OF CONTENTS

Chapter IV   DOG

Chapter V   WOLF

Chapter VI   BEAR

References

❦

Contributors

EBIHARA Shiho	Tokyo University of Foreign Studies
ENDO Mitsuaki	Aoyama Gakuin University
FUKAZAWA Mika	National Ainu Museum
FUKUI Rei	The University of Tokyo
HIRANO Ayaka	Osaka University
IWASA Kazue	Nagoya University of Foreign Studies
IWASAKI Takamasa	Kyoto University
KIMURA Kimihiko	Tokyo University of Foreign Studies
KODAMA Nozomi	Kumamoto University
KOMORI Junko	Osaka University
KURABE Keita	Tokyo University of Foreign Studies
MATSUMOTO Ryo	Kobe City University of Foreign Studies
MINEGISHI Makoto	Tokyo University of Foreign Studies
NAGATO Youichi	J. F. Oberlin University Graduate Division
NAKAGAWA Hirosi	Tokyo University of Foreign Studies
NAKAO Shuichiro	Osaka Univeristy
NAKAZAWA Kohei	Shinshu University
ONO Chikako	Hokkai-Gakuen University
SAITÔ Yoshio	Takushoku University
SHIMIZU Masaaki	Osaka University
SHINAGAWA Daisuke	Tokyo University of Foreign Studies
SHIRAI Satoko	The University of Tokyo
SUZUKI Hiroyuki	Kyoto University
TAGUCHI Yoshihisa	Chiba Univeristy
TOMITA Aika	Osaka Shoin Women's Univeristy
UTSUMI Atsuko	Meisei University
YAGI Kenji	Kokushikan University
YOKOYAMA Akiko	Tokyo University of Foreign Studies
YOSHIOKA Noboru	National Museum of Ethnology

❧

Acknowledgements

❧

Abbreviations

#	heuristically reconstructed (Nilo-Saharan) root
[±ATR]	the 'advanced tongue root' feature
Akk	Akkadian
CDIAL	*A comparative dictionary of Indo-Aryan languages* (Turner 1985)
DEDR	*A Dravidian etymological dictionary* (Burrow & Emeneau 1984)
ES	Eastern Sudanic
IA	Indo-Aryan
LB	Lolo-Burman
n	the northern branch of Eastern Sudanic
PIE	Proto-Indo-European
PIr	Proto-Iranian
PKC	Proto-Kuki-Chin
PLB	Proto-Lolo-Burman
PTB	Proto-Tibeto-Burman
s	the southern branch of Eastern Sudanic
STEDT	*The Sino-Tibetan etymological dictionary and thesaurus*
TB	Tibeto-Burman
WrT	Written Tibetan

Subgrouping of languages

Subgrouping of Paleoasian Languages

"Paleoasian" is not a genealogical grouping but an aerial one. The languages that belong to the Paleoasian group are Chukotko-Kamchatkan and Nivkh, and Yukagir and Ket have also been considered as group members. In recent years, it has been suggested that Ket could be a cognate with Na-Dene languages and that Yukagir and Uralic languages have a genealogical relationship.

The language data mapped in this volume are those of Chukchi, Alyutor, Koryak, Itelmen, and Nivkh.

・Chukotko-Kamchatkan
    Northern
        ——Chukchi
        ——Alyutor
        ——Koryak
    Southern
        ——Itelmen
          Northern dialect
          Southern dialect
・Nivkh
    Sakhalin dialect
    Amur dialect

(ONO Chikako)

- ■ Chukchi
- ▲ Koryak
- ◆ Alyutor
- ⊗ Itelmen
- ★ Nivkh
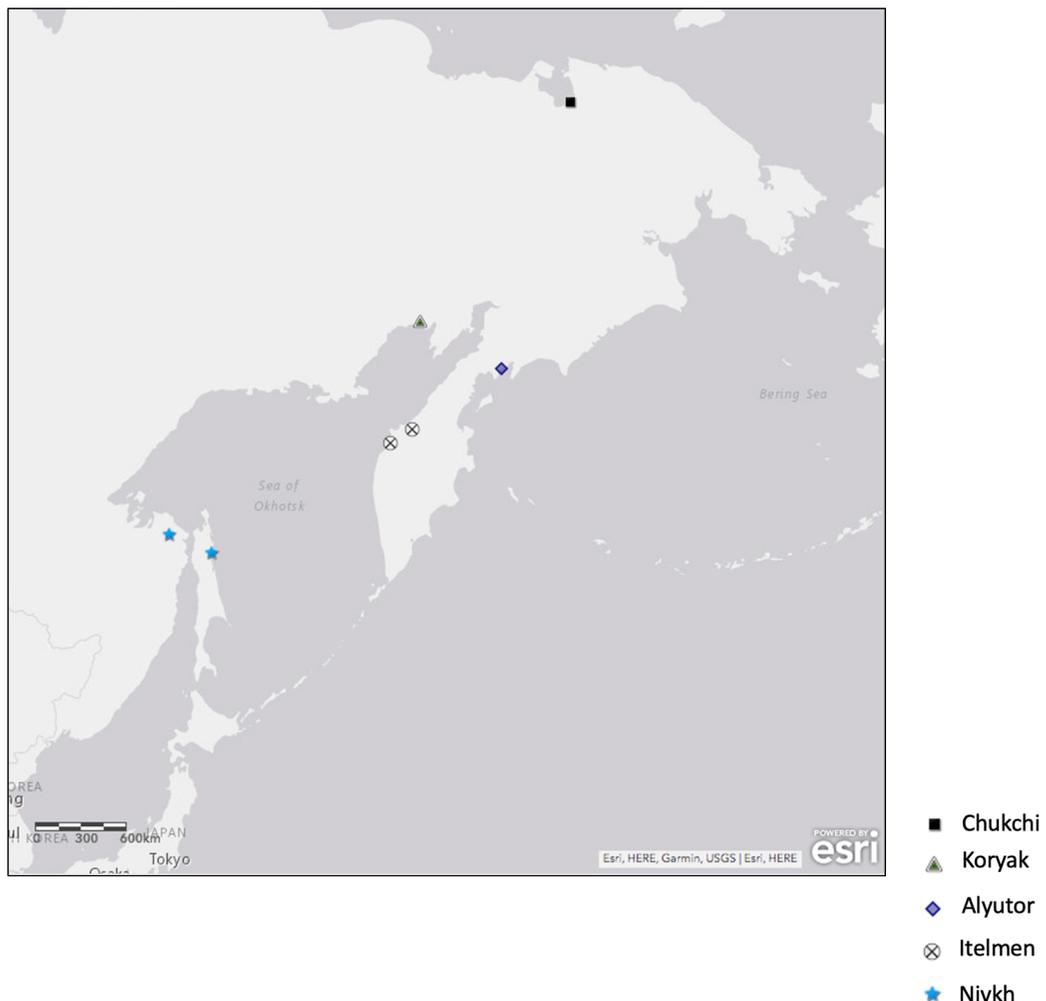
Subgrouping of Ainu

The major subgrouping of Ainu is into the three groups of the Sakhalin, Hokkaido, and northern Kuril dialects, generally accepted in previous studies (Hattori and Chiri 1960, Asai 1974, Tamura 2000). The Hokkaido dialect can be divided into the eastern and western dialectal groups. The southern Kuril dialect can be included in the eastern Hokkaido dialect (Hayashi 1973 [1940]).

We will not deal with further subgroupings in Ainu here apart from the following brief note. The dialects in and around Saru and Chitose in western Hokkaido area often show special patterns in vocabulary, including functional words, that may be similar to those of the Sakhalin dialect. Hattori and Chiri (1960) and Asai (1974) suggested the minor subgrouping of the northernmost dialect of Soya and the southernmost dialect of Samani in Hokkaido.

(FUKAZAWA Mika)

・Hokkaido dialect
— Western Hokkaido dialect ●
— Eastern Hokkaido dialect ○
・Sakhalin dialect □
・Northern Kuril dialect ☆

Subgrouping of Japonic

Although there are various hypotheses about how to divide Japonic languages, we can broadly classify them into Japanese and Ryukyuan. Japanese is divided into Eastern Japanese (EJ), Western Japanese (WJ), and Kyūshū Japanese (KJ). Ryukyuan languages are divided into Northern Ryukyuan (NR, including Amami) and Southern Ryukyuan (SR). The criteria for classification are as shown in the table: forms for 'be' (LAJ 53), suffixes for 'purpose of motion' (GAJ 21), forms for the interrogative 'what' (Pellard 2015), and forms for 'say' (cf. SR *�à̃ĩz- < *ani+ip- 'say so,' *ip- 'scold' < 'say').

We include Hachijō dialect in Eastern Japanese since it shares innovations with the Eastern Japanese dialects (Igarashi 2021).

There are more narrow divisions than this, and there are many differences depending on the researcher.

It is difficult to draw a phylogenetic tree because it is uncertain which forms are innovative or retained.

Table 1: Criteria for classification of Japonic

criterion / branch	be	(go) for	what	say
EJ	*wi-	*-ni	*nani	*ip-
WJ	*wor-	*-ni	*nani	*ip-
KJ	*wor-	*-ga	*nani	*ip-
NR	*wor-	*-ga	*nawo	*ip-
SR	*wor-	*-ga	*nawo	*ə̃ïz-

(NAKAZAWA Kohei and YOKOYAMA Akiko)

A proposal for the phylogenetic tree of Japonic languages
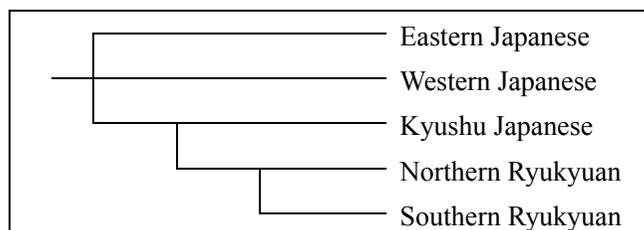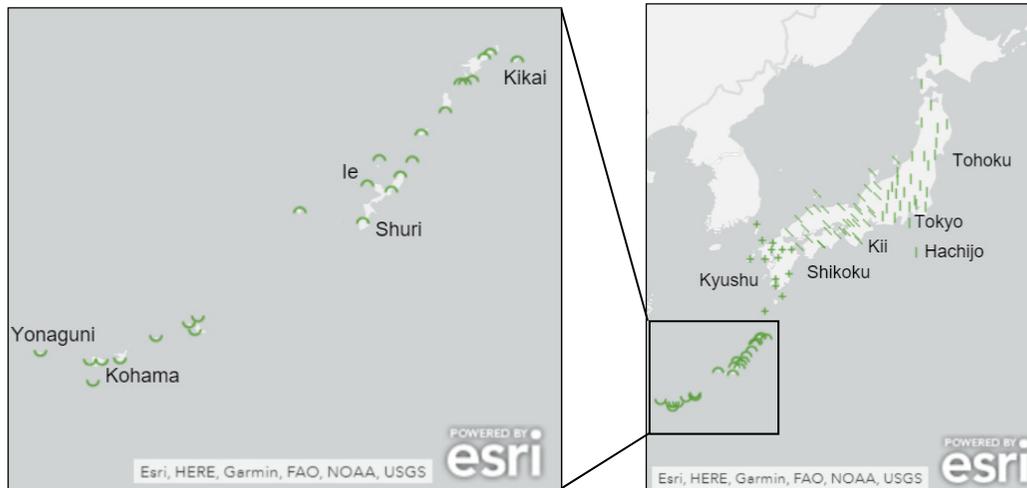
- | Eastern Japanese
- \ Western Japanese
- + Kyushu Japanese
- ⌒ Northern Ryukyuan
- ⌣ Southern Ryukyuan

Subgrouping in Korean

Current standard way of subgrouping Korean dialects is the following:

North-eastern dialects: dialects spoken in the Hamgyŏng province
North-western dialects: dialects spoken in the Phyŏng'an province
Central dialects: dialects spoken in Hwanghae, Kyŏnggi, Ch'ungch'ŏng, and Kangawŏn provinces

South-eastern dialects: dialects spoken in the Kyŏngsang province
South-western dialects: dialects spoken in the Chŏlla province
Cheju dialects: dialects spoken in the Cheju province

In the Map, only a few representative cities are marked for each subgroup.

(FUKUI Rei)
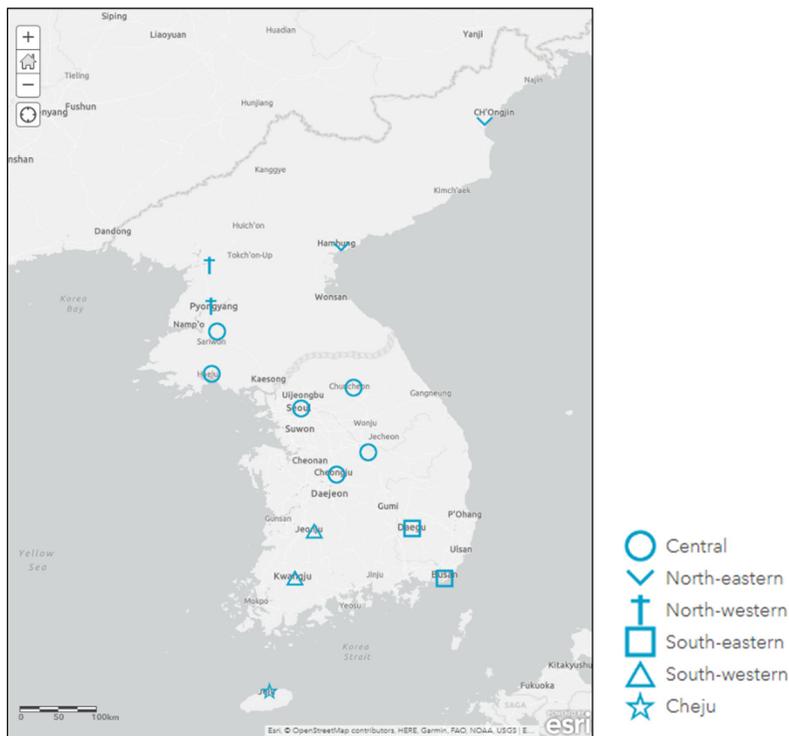
Subgrouping of Sinitic

We basically adopt the subgrouping in Sinitic proposed in Wurm et al. 1987 (Data are from Zhan et al. 2017, Hou 2002, Qian 2010). 1. Mandarin, 2. Jin, 3. Wu, 4. Xiang, 5. Gan, 6. Kejia, 7. Yue, 8. Min, 9. Hui, 10. Ping / Tu hua. Mandarin is further divided into 8 subgroups. 1a. Beijing, 1b. Dongbei, 1c. Jilu, 1d. Jianghuai, 1e. Jiaoliao, 1f. Lanyin, 1g. Xinan, 1h. Zhongyuan.

This subgrouping is a kind of the traditional dialect classification in China, and is said to have some consistent with some phonological changes from middle Chinese, such as developments of voiced initials or entering tone. However, at this stage, it is difficult to create a phylogenetic tree because this classification also takes into account non-linguistic backgrounds such as social and cultural backgrounds or geographical distribution.

(YAGI Kenji)
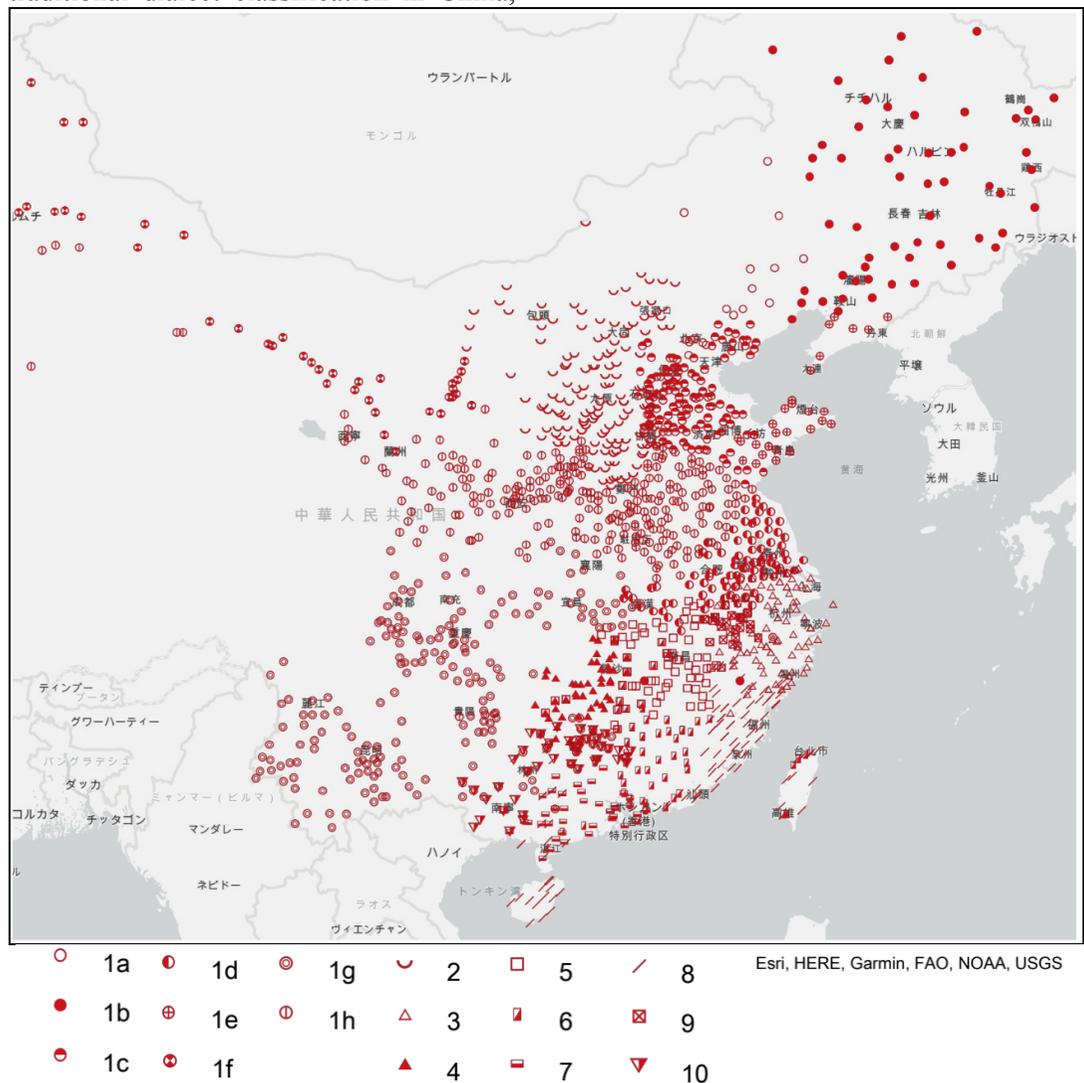
Subgrouping of Hmong-Mien

The subgrouping indicated by the following tree diagram is based on the phylogenetic study that the author conducted using lexical data. The tree indicates that the languages family comprises two branches: Hmongic and Mienic. It shows the internal structure of the Hmongic branch because it has more diversity inside than Mienic. West Hmongic and Pu-Nu constitute a clade, which might be called West Hmongic as a whole, but we here use traditional terms to denote each group. Some phonological evidence might suggest a tree with a higher resolution, which places North Hmongic and Pa-Hng in higher nodes than other Hmongic languages. Here, we rather conservatively place these two languages in a parallel fashion with other Hmongic languages.

(TAGUCHI Yoshihisa)
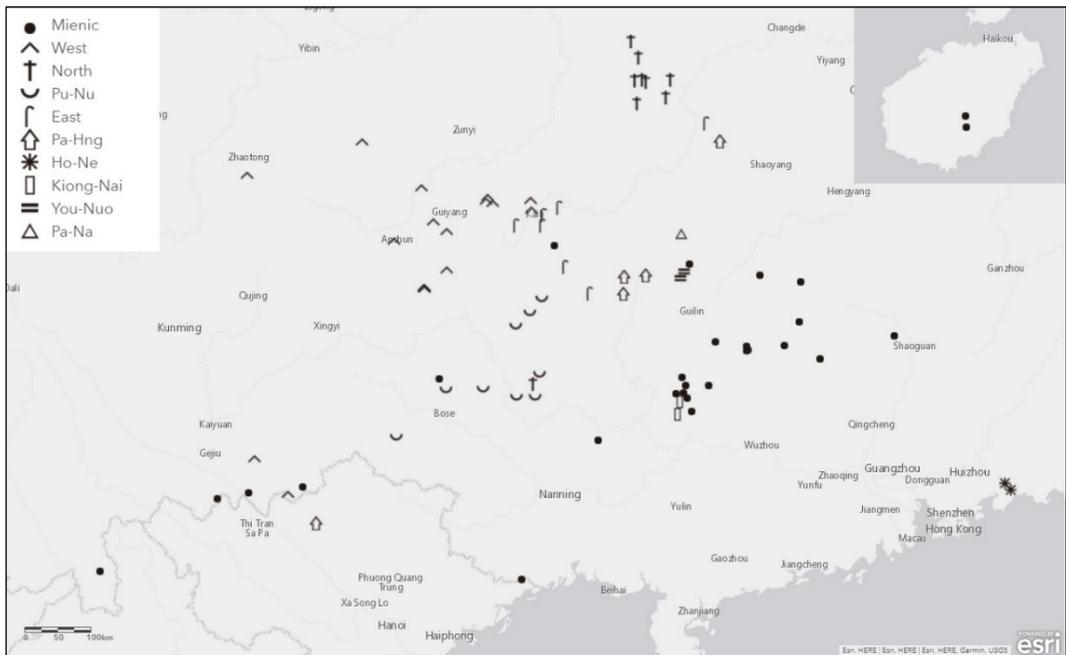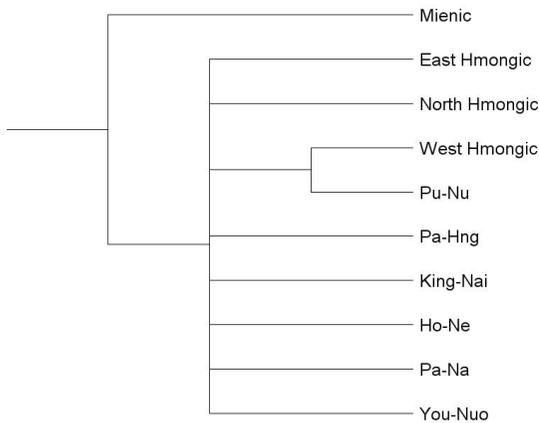
Subgrouping of Kra-Dai

We adopt the subgrouping and its hierarchy in Kra-Dai as proposed by Liang and Zhang (1996:13) to denote a whole. The established classification by Li (1977) is adopted for the sub-branches of the Tai branch.

Kra-Dai
- Kra
- Lingao
- SW Tai
- C Tai
- N Tai
- Li
- Dong-Shui
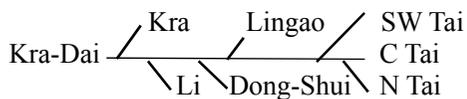
Kra is the most conservative branch, while Li ranks second. They preserve common vocabulary with Austronesian, for example, numerals, and so on. Northern Tai is divided on the basis of a phonological criterion that no distinction of aspiration exists.

(ENDO Mitsuaki)
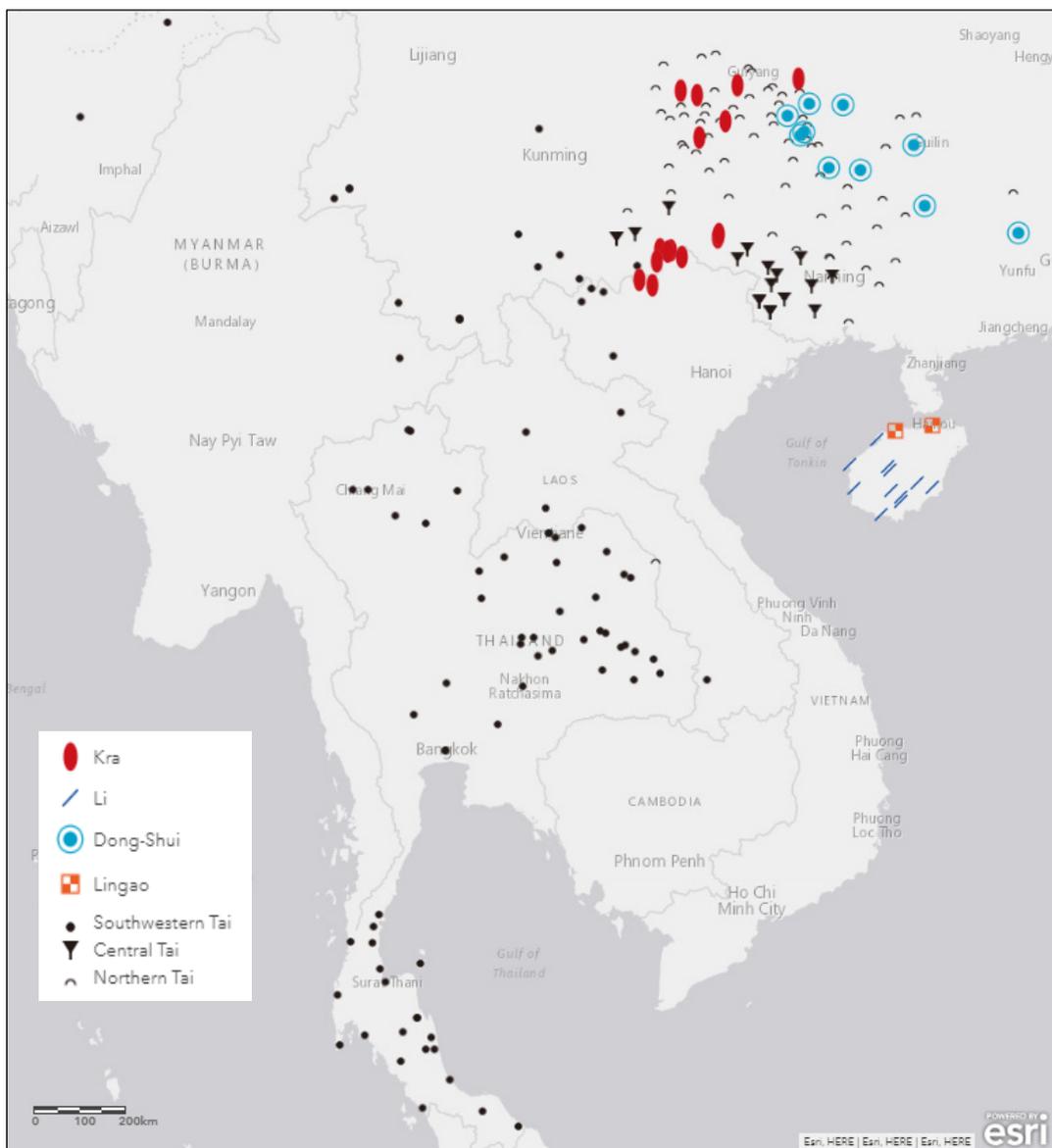
Subgrouping of Tibeto-Burman

There have been varying suggestions for the subgrouping of Tibeto-Burman (TB) (van Driem 2015; Matisoff 2015; Thurgood 2017; Zhang et al. 2019; Sagart et al. 2019; Zhang et al. 2020). Here, the model following STEDT (Matisoff 2015) with some updates is referred to, with the TB language hierarchy shown in Figure 1. There are also one unclassified TB language and two Sinitic-Tibetic mixed languages.

Abbreviations: NE IAG: North-eastern Indian areal group; TQ: Tangut-Qiang; LBN: Lolo-Burmese-Naxi; 'NA': 'North Assam'; KC: Kuki-Chin; 'N'AG: 'Naga' areal group; TK: Tibeto-Kannauri; KMC: Kham-Magar-Chepang; LB: Lolo-Burmese.
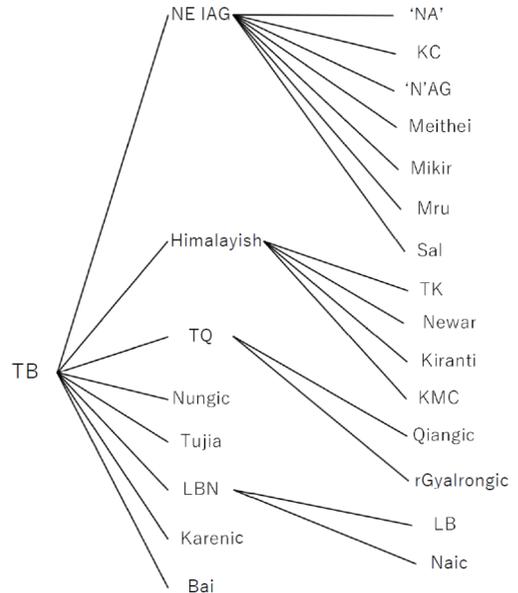
Fig.1: Subgrouping of TB

(SUZUKI Hiroyuki, EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko)
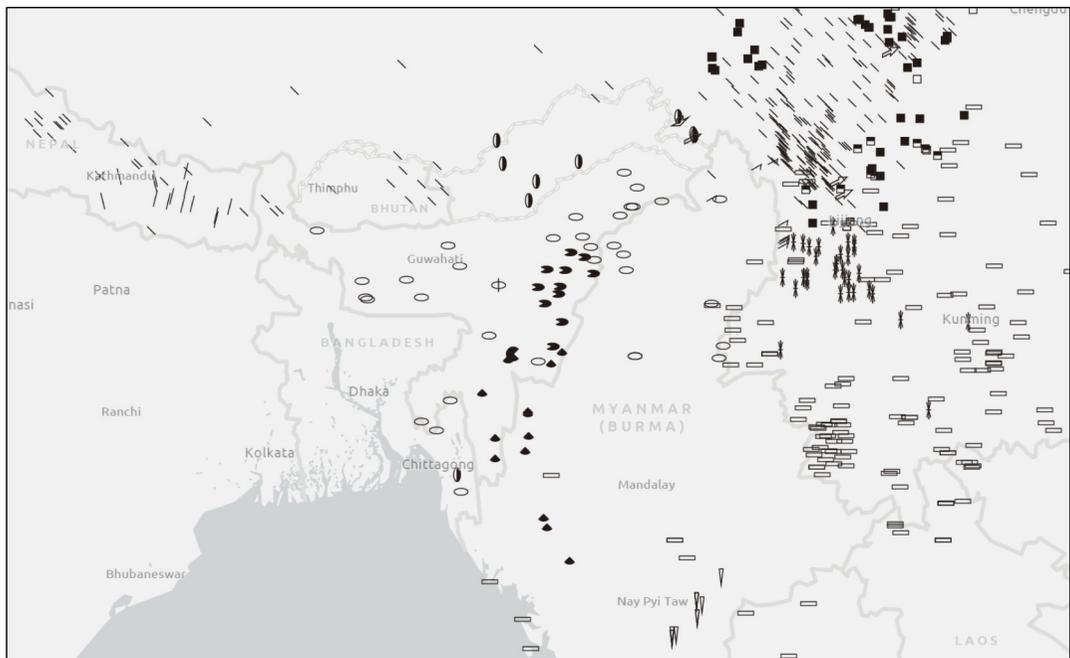
Esri, HERE, Garmin, FAO, NOAA, USGS

Figure 2: Distribution of Tibeto-Burman subgroups (enlarged).

◖	North Assam	**/**	Kiranti	**⍌**	Karenic
▲	Kuki-Chin	**✕**	Kham-Magar-Chepang	**✶**	Bai
🝜	Naga Areal group	**■**	Qiangic	**⬋**	mixed
◖	Meithei	**◻**	rGyalrongic	**⬋**	unidentified
⬦	Mikir	**◹**	Nungic		
◯	Sal	**◺**	Tujia		
＼	Tibeto-Kannauri	**▭**	Lolo-Burmese		
＼	Newar	**▬**	Naic		

Figure 3: Distribution of Tibeto-Burman subgroups (whole).

Subgrouping of Austroasiatic

Austroasiatic is first divided into Munda and Mon-Khmer. Regarding Mon-Khmer, we adopt the subgrouping of Austroasiatic by Diffloth & Zide (1992) whose subgrouping is given below as Figure 1. Sidwell (2014), after describing the history of Austroasiatic classification proposals since the middle of the 19th century, offers 'provisional' classification. His tree is based on 'lexical, lexicostatistical, computational phylogenetic, and phonological studies', and is characterized as strongly branching: with eleven primary subgrouping nodes, among which only two nodes have secondary branching; one is Khasian and Palaungic, and the other, Aslian and Nicobarese.
(MINEGISHI Makoto & SHIMIZU Masaaki)
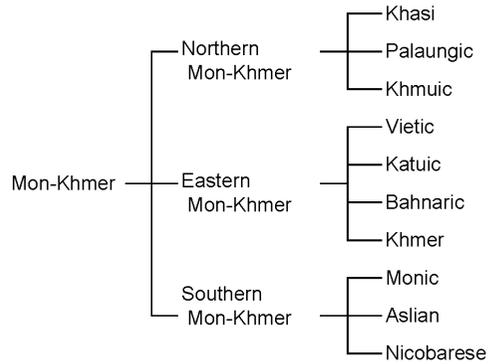
Figure 1: Mon-Khmer subgrouping.
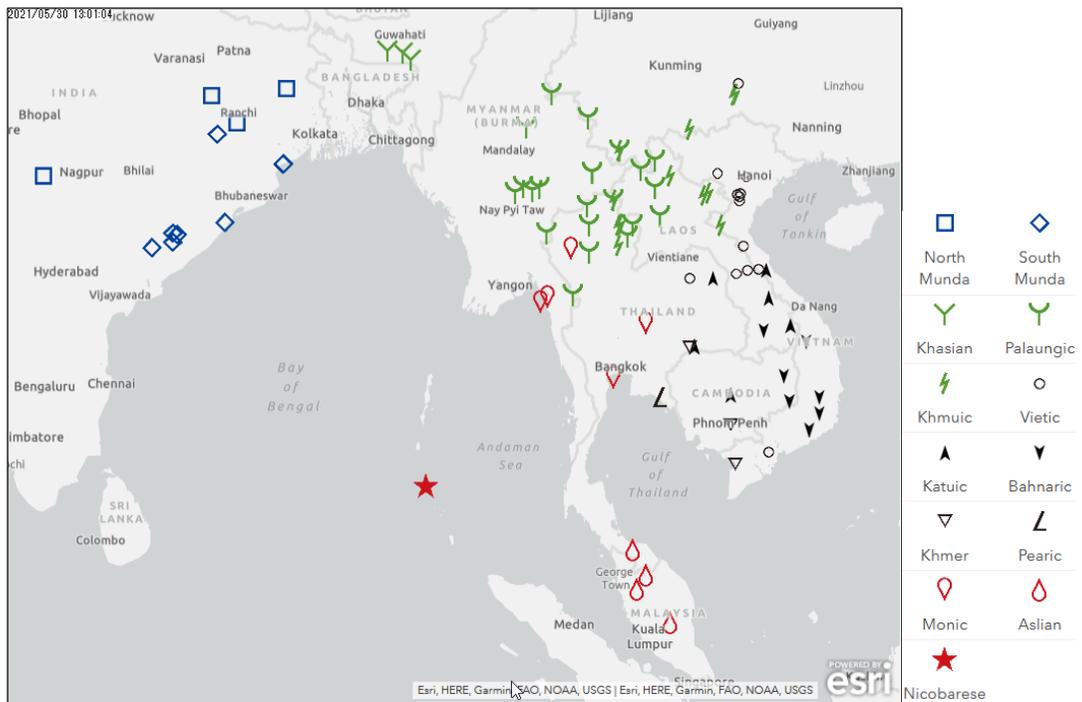
Subgrouping of Austronesian

We adopt the subgrouping and its hierarchy in Austronesian Languages proposed by Blust 1980 and Blust 1999. The Formosan languages, or the Austronesian languages of Taiwan belong to nine primary branches of the Austronesian family. They are "generally believed to be the most diverse in the entire Austronesian language family" (Li 2008). They do not form a subgroup linguistically, but for the purpose of this geolinguistic study, they are grouped together and referred to Formosan languages (FRM).

All of the non-Formosan languages belong to a tenth primary branch, which is Malayo-Polynesian (MP). MP split into West Malayo-Polynesian(WMP) and Central-East-Malay-Polynesian (CEMP), the latter of which split into Central-Malayo-Polynesian (CMP) and East Malayo-Polynesian (EMP). EMP are grouped into South-Halmahera-West-New-Guinea languages (SHWNG) and Oceanic languages.

The geological perspective as well as actual geolinguistic characteristics are considered for the subgrouping of non-Formosan languages. They are grouped into WMP, Oceanic, and the rest which will be referred as CEMP (i.e., CEMP languages except for Oceanic languages). WMP languages are frequently divided into Philippine languages and Indonesian languages when they show remarkable difference within WMP.

Oceanic Austronesian

FRM
WMP
MP
EMP
CEMP   SHWNG
CMP
(UTSUMI Atsuko)

FRM ◢
WMP ▭
CEMP ⊤
Oceanic ●
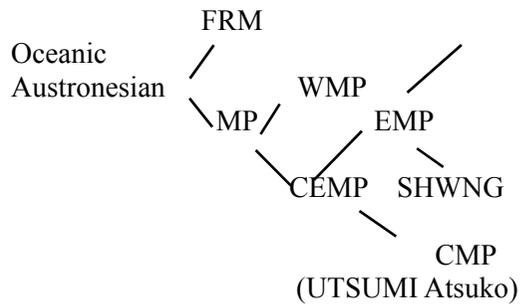
Subgrouping of Tungusic

According to Ikegami (1989), Tungusic languages are divided into four groups:

⬤ (Group I) Evenki, Ewen, Negidal, Solon (Evenki in China)
☐ (Group II) Udehe, Orochi

◯ (Group III) Nanay, Ulcha, Uilta
★ (Group IV) Sibe

(MATSUMOTO Ryo)
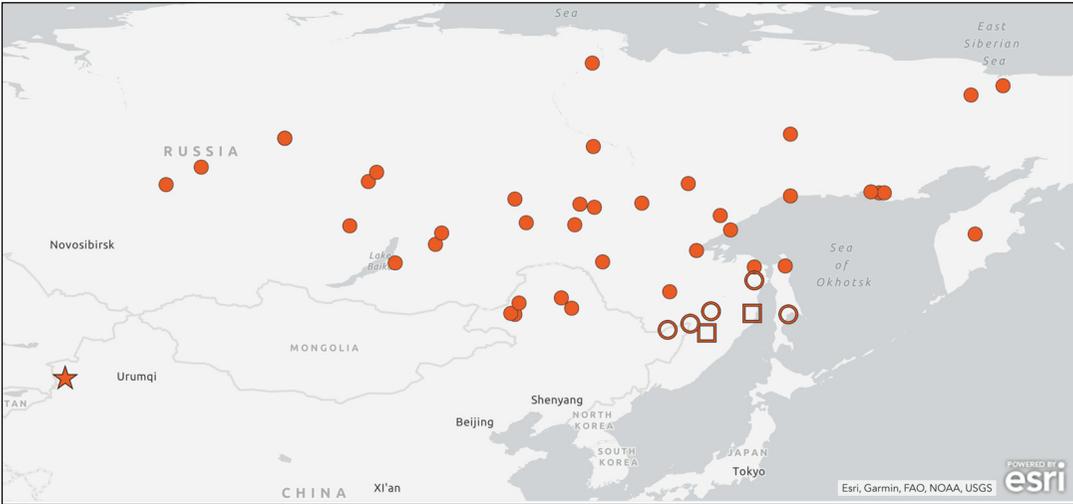
Subgrouping of Uralic

Here I show the subgroups of the Uralic language family in the traditional way. The Uralic language family is divided largely into two branches, Samoyedic and Finno-Ugric, and then Finno-Ugric into two sub-branches, Ugric and Finno-Permic. Finno-Permic includes most languages of the Uralic family and has more subdivisions, but here:

🔵 Finno-Permic

Komi, Udmurt, Mari (Hill Mari, Meadow Mari), Mordvinic (Erzya, Moksha), Finnish, Estonian, Livonian, Votic, Karelian, Veps, Ingrian, Sami

⭐ Ugric languages

Hungarian, Khanty, Mansi

🟦 Samoyedic languages

Nenets, Enets, Selkup, Nganasan

(MATSUMOTO Ryo)
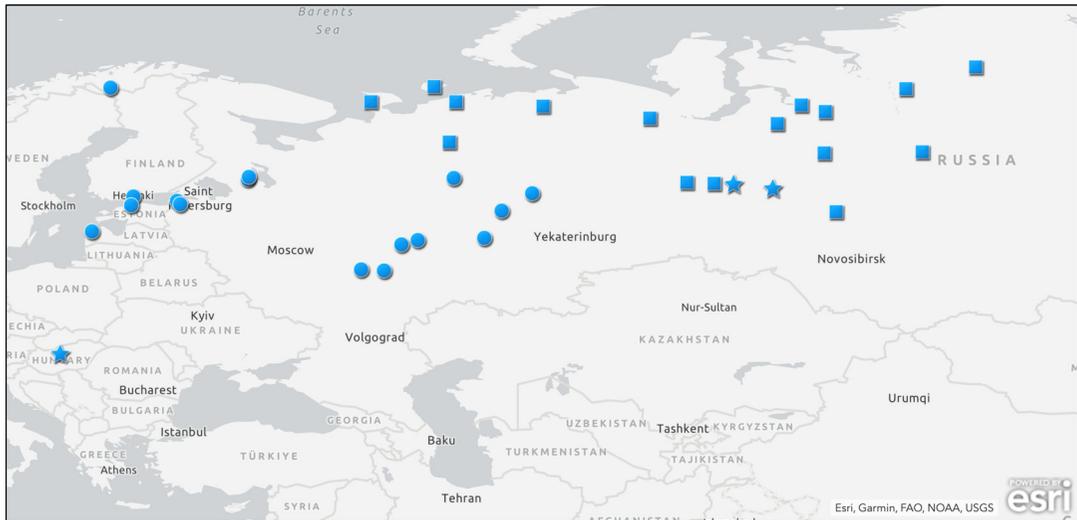
Subgrouping of Mongolic and Turkic

Mongolic and Turkic groups of languages are now considered to be separate language families by the majority of researchers. A classification of languages may differ depending on the features chosen for criteria. The classifications shown below are mainly based on V. Rybatzki (2003) and L. Johanson (1998).

1. Mongolic languages

Northeastern: Dagur
Northern: Khamnigan, Buryad
Central: Mongol, Ordos, Oirad
South Central: Shira Yughur
Southeastern: Monguor, Baoan, Dongxiang, Kangjia
Southwestern: Moghol

2. Turkic languages

Oghuz (Southwestern): Turkish, Azeri, Gagauz, Turkmen, Khorasan Turkic, Kashkay, Afshar

Kipchak (Northwestern):
[Volga-Ural (Northern)] Tatar, Bashkir
[Ponto-Caspian (Western)] Kumyk, Karachay, Balkar, Crimean Tatar, Karaim
[Aralo-Caspian (Eastern)] Kyrgyz, Kazakh, Karakalpak, Nogay

Uighur (Southeastern): Uzbek, Uighur, Sarïg Yughur, Salar

Siberian (Northeastern):
[North Siberian] Sakha, Dolgan
[South Siberian] Tuva, Tofa, Khakas, Shor, Chulym, Altay

Oghur/Bulgar: Chuvash

Arghu: Khalaj

(SAITÔ Yoshio)
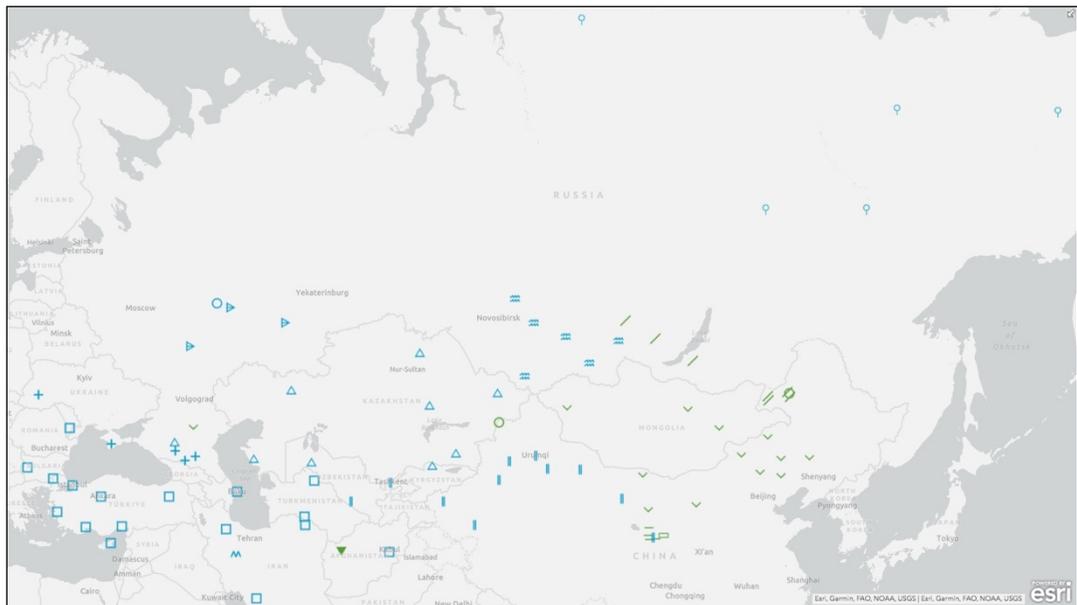
Mongolic ○ Northeastern ╱ Northern ∨ Central ▭ South Central ─ Southeastern
▼ Southwestern
Turkic ▢ Oghuz ▷ Volga-Ural ✛ Ponto-Caspian △ Aralo-Caspian ‖ Uighur
⚲ North Siberian ≋ South Siberian ○ Oghur ∿ Arghu

Figure 1: Subgroups of Mongolic and Turkic.
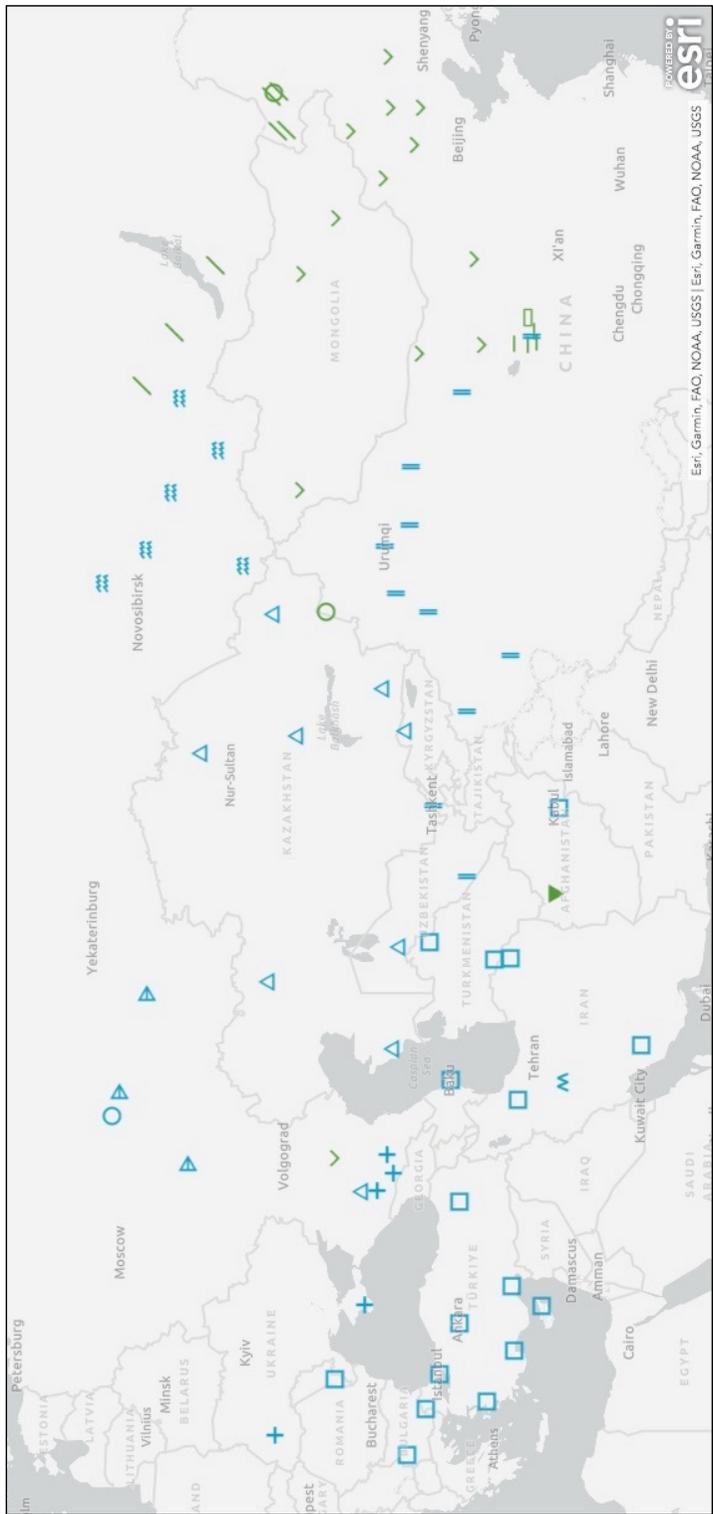
Figure 2:   Subgroups of Mongolic and Turkic except North Siberian Turkic.

Subgroupings of Indo-Aryan, Nuristani, Andamanese, and language isolates in South Asia

I show the subgroupings of the Indo-Iranian branch, with the exception of Iranian, of the Indo-European family, and of the Andamanese family and some language isolates, in the map.

The subgrouping in Indo-Aryan remains controversial. Here, I have simply classified the Indo-Aryan and Nuristani languages as per the following cladogram, with reference to Masica (1991), Eberhard, Simons, and Fennig (2021), and Hammarström, Forkel, Haspelmath, and Bank (2020).
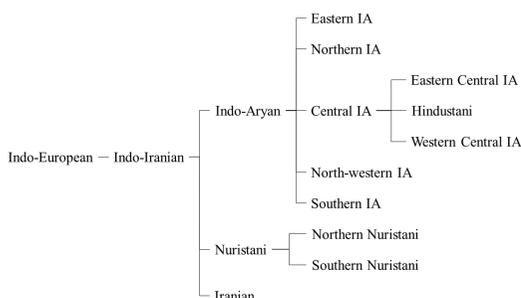
Nuristani is a subbranch of the Indo-Iranian branch and so, of course, parallels the Indo-Aryan and Iranian subbranches. This branch can be subdivided into two groups, northern and southern.

The Andamanese family has two branches, Great Andamanese and Jarawa-Ongan. The former can be further subgrouped into two or three areal groups. The latter branch has two living languages, Jarawa and Öñge. Furthermore, the Sentinelese language is found on the Sentinel island south-west of the Great Andaman. That language, however, remains undescribed as its speakers absolutely refuse to make contact with outsiders, so it cannot be classified anywhere phylogenetically.

(YOSHIOKA Noboru)
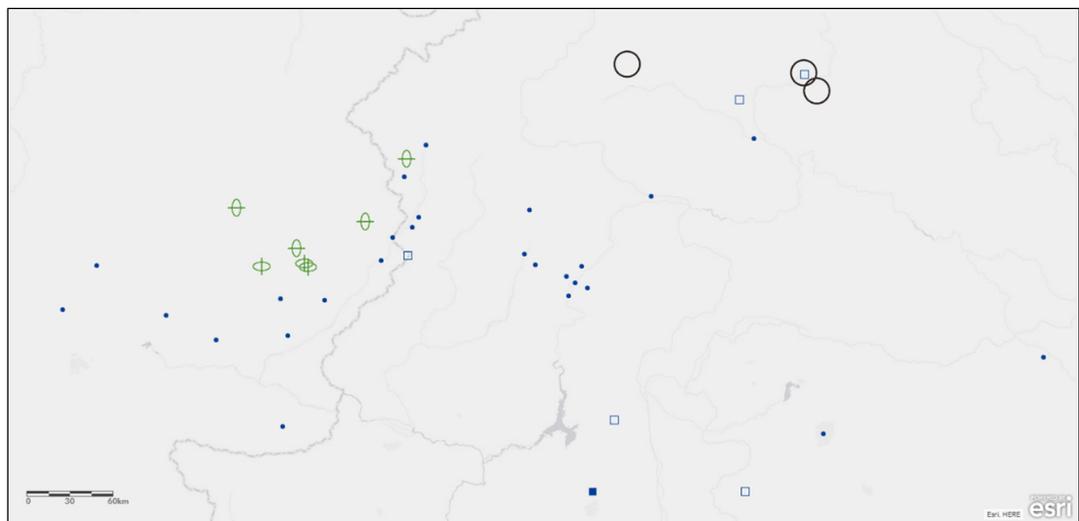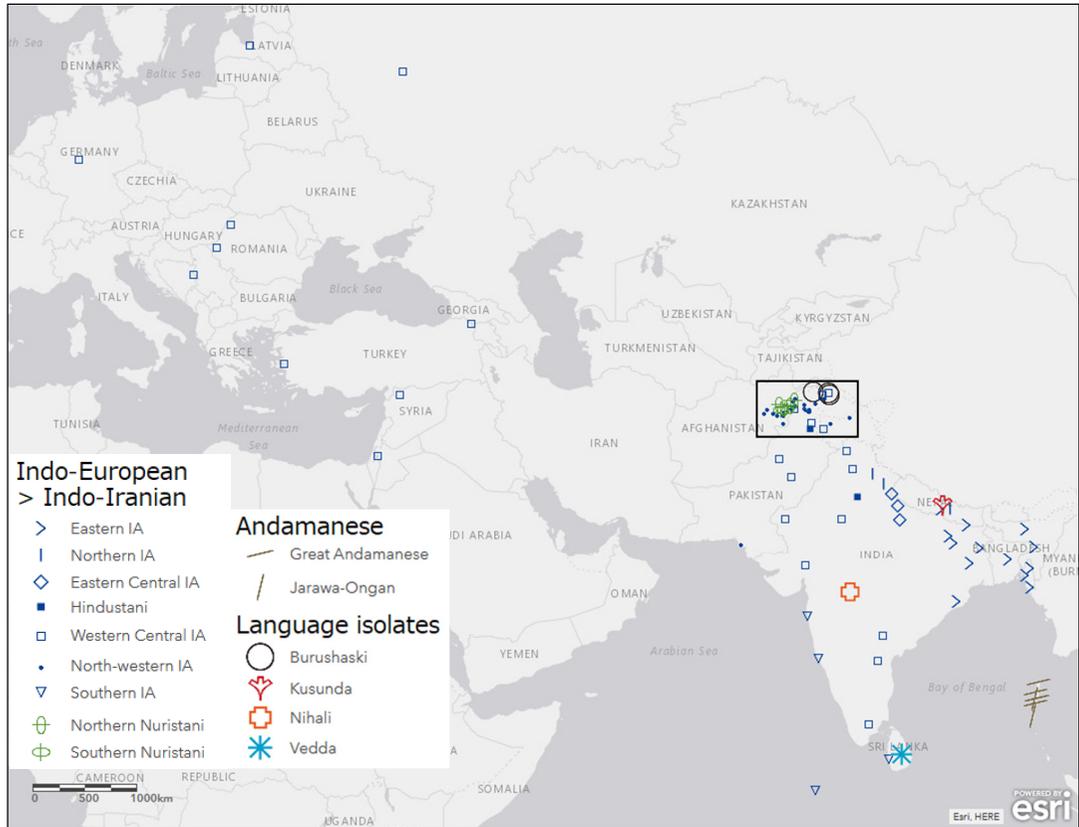
Subgrouping of Dravidian

The Dravidian languages were recognized as a language family as early as 1816 by Francis Whyte Ellis, who was in the civilian service at Madras. Krishnamurti (2003) replaced the earlier tripartite classification of Dravidian languages with the following four subgroups by splitting the erstwhile Central Dravidian based on his genealogical assumptions.

1. South Dravidian (SD I)

Tamil, Malayalam, Irula, Kodagu, Toda, Kota, Kannada-Badaga, Tulu-Koraga*

2. South Central Dravidian (SD II)

Telugu*, Gondi, Konda, Kui, Kuvi, Pengo, Manda

3. Central Dravidian (CD)

Kolami, Naiki, Parji, Gadaba

4. North Dravidian (ND)

Kurukh, Malto, Brahui*

The four-way classification is accepted by most researchers, although inclusion of Tulu-Koraga, Telugu and Brahui in their respective subgroups may be viewed by some as more tentative than conclusive.

The phylogenetic relationship between the four subgroups, which would have a direct implication on the issue of the geographical diffusion of the language family, remains unsettled. Kurukh-Malto and Brahui are isolated from each other as well as from other subgroups. If they comprise a single phylogenetic branch i.e. North Dravidian, their spatial distribution could be attributed to highly migratory nature of their speakers at some point in the past, entailing that south-to-north diffusion of the language family cannot be ruled out.

(KODAMA Nozomi)

╱	SD I
∨	Tulu-Koraga
◇	SD II
▱	Telugu
⊕	CD
+	ND
✳	Brahui
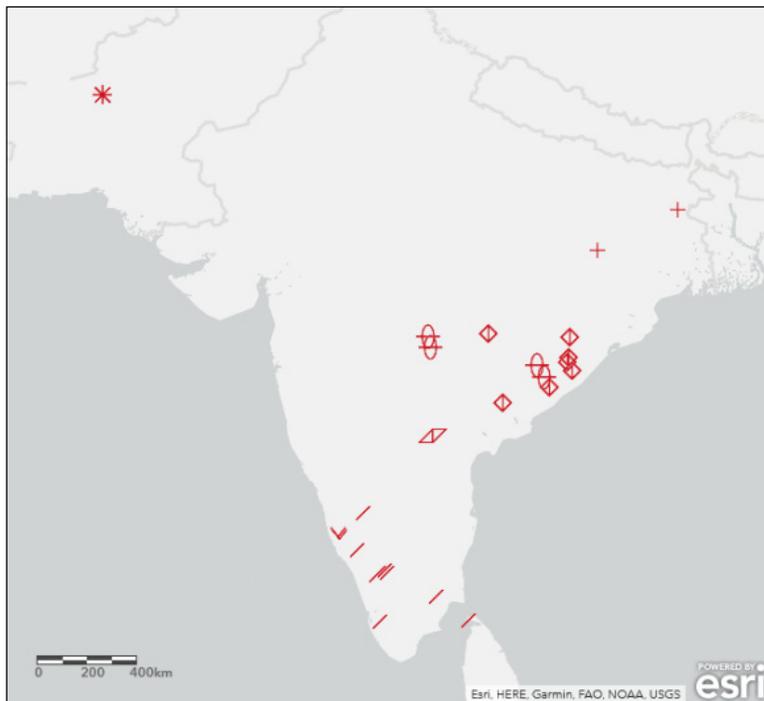
Subgrouping of Armenian and Iranian

Armenian is an independent branch of the Indo-European languages. It is divided further into two major subgroups, namely East and West. The Iranian languages are a subgroup of Indo-Iranian in Indo-European language family. They spread a vast area from Western China (Xinjiang) in the east, to Central Turkey in the west, and from North Caucasus (Russia and Georgia) in the north, to the southern Pakistan and the northern Oman in the south.

In terms of historical and typological linguistics, this branch is generally classified into Eastern and Western Iranian. These are divided further into four subgroups, namely North-Western, North-Eastern, South-Western and South-Eastern Iranian. Each of them has its archaism and innovation, therefore we cannot surmise which language best preserves archaism on the whole.

It is arguable whether Ormuri and Parachi are classified into Western or Eastern Iranian. Efimov (1986: 8) includes them into North-eastern Iranian, while Morgenstierne (1929: 12) classifies them into central position among the Iranian languages.

Note that the subgroup names do not always correspond with the geographical distribution of the modern Iranian languages. For example, Ossetic, although it belongs to North-Eastern Iranian, is spoken in the western region. Also, Balochi spreads rather to the southeastern area while it is classified into North-Eastern Iranian. Map1 shows the distribution and subgrouping of the modern Iranian languages.

(IWASAKI Takamasa)
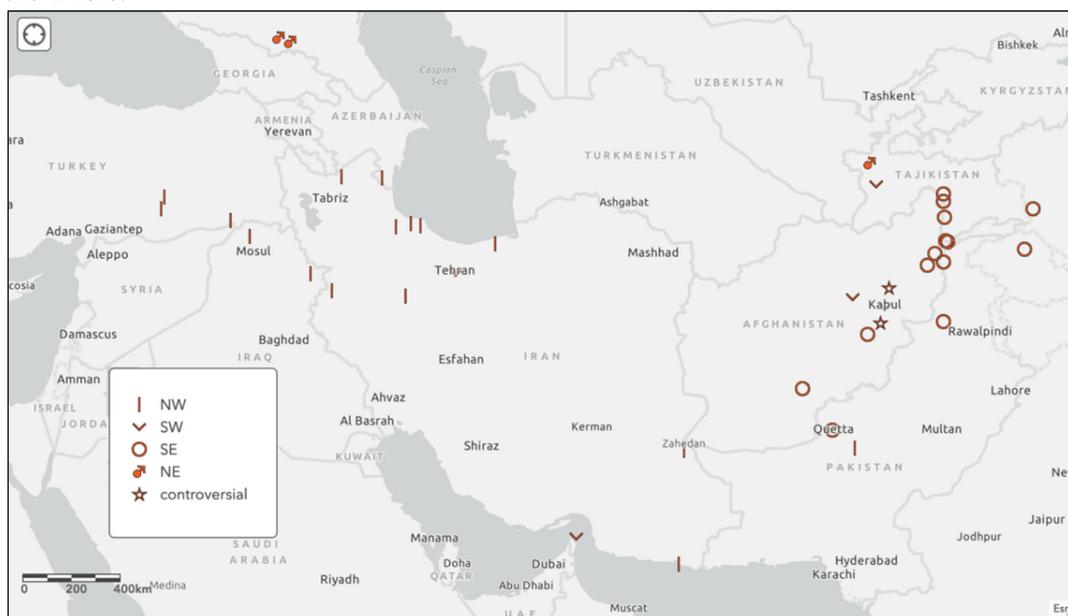
Esri, HERE, Garmin, FAO, NOAA, USGS

Subgrouping in Caucasian languages

Caucasian languages are classified in three language groups: Kartvelian (South-western Caucasus), Abkhazo-Adyghean (North-western Caucasus), and Nakho-Daghestanian (Eastern Caucasus).

Kartvelian includes Kartuli (Georgian), Mingrelian, Laz, and Svan.

Abkhazo-Adyghean includes Adyghe, Kabardian (East Circassian), Abzhywa (Abkhaz), T'ap'anta (Abaza), and Ubykh.

Nakho-Daghestanian are further divided into Nakh languages and Daghestanian languages. Nakh contains two languages Chechen and Ingush, whilst Daghestanian includes Bats, Avar, Andi, Botlikh, Godoberi, Akhvakh, Karata, Bagvalal, Tindi, Chamalal, Bezhta, Hunzib, Tsez, Hinukh, Khvarshi, Lak, Dargwa, Lezgi, Tabasaran, Agul, Rutul, Ts'akhur, Archi, Kryz, Budukh, Udi, and Khinalug.

(SUZUKI Hiroyuki)

● Kartvelian
▬ Abkhazo-Adyghean
0 Nakho-Daghestanian
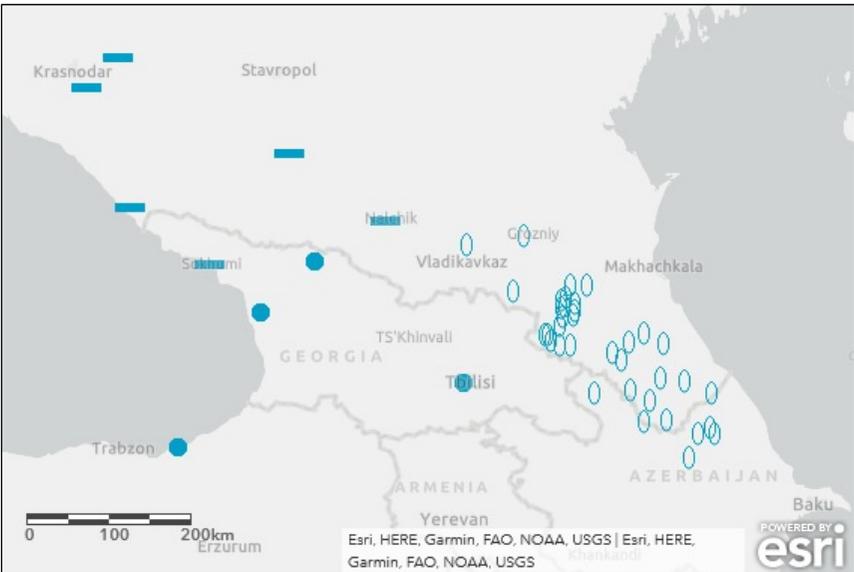
Subgrouping of Semitic

The Semitic is a branch of the Afroasiatic phylum. The earliest attested Semitic is Akkadian in Mesopotamia, which belongs to East Semitic.

In the Syro-Palestinian area there were several Semitic languages such as Eblaite and Ugaritic. Then during the second millennium BCE, Canaanite (Hebrew, Phoenician) and Aramaic emerged. Hezron (1974, 1976) proposed subgrouping of this group as Central Semitic, in which Hetzron grouped Arabic, insted of South Semitic. Aramaic was used as a lingua franca in Babylonian and Persian empires between the seventh and the forth centuries BCE. It remained in use as a literary language until the fifth century CE. Modern varieties of Aramaic survive in a number of linguistic enclaves such as Ma'lūla in Syria (Currently, most of the village residents have fled the country), Tūr 'Abdīn in Western Kurdistan.

Canaanite is a collective term for Hebrew, Phoenician and a few other languages. Hebrew is the language of the Jewish Bible (1200-200 BCE.) and one of the two national languages of Israel now.

Arabic is the most widely distributed Semitic language in the Middle East after the Islamic conquest. Arabic dialect regions are broadly classified into North Africa, Egypt/Sudan, Arabian Peninsula, Syria, and Iraq. And there are peripheral dialects in Malta, Uzbekistan, Chad, Nigeria, Juba in South Sudan and a creole in Kenya.

South Semitic is divided into three groups, Epigraphic South Arabian, Modern South Arabian and Ethiopian. Epigraphic South Arabian is languages of probably between the eighth century BCE and the sixth century CE. Modern South Arabian languages, such as Mehri, Jibbālī, Soqotrī and Hobyōt in Yemen and Oman, probably go back to spoken varieties of Epigraphic South Arabian. To Ethiopian, belong a large number of languages such as Tigre, Tigriña and Amharic, the official language of Ethiopia. Ge'ez is the Classical Ethiopic, the language of the empire of Aksum in first centuries CE.

(NAGATO Youichi)
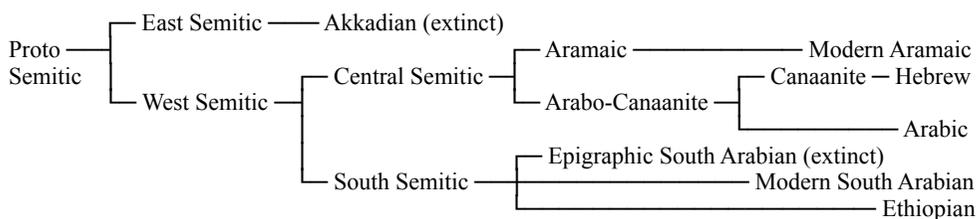
Figure 1: Subgrouping of Semitic (after Hetzron 1974, 1976).
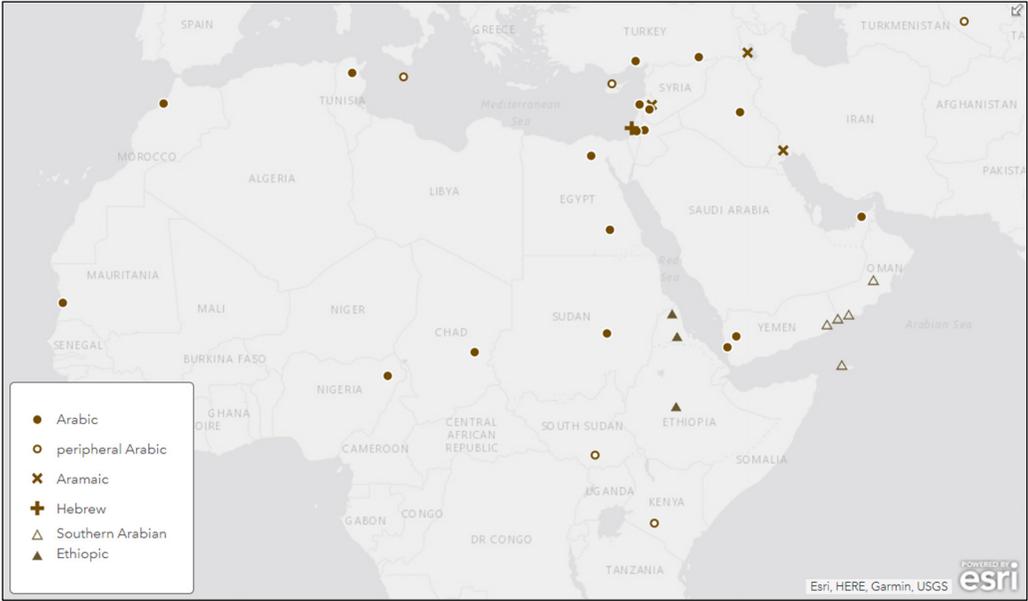
Subgrouping of Nilo-Saharan

For the time being, there is no full consensus about the membership or the subgrouping of Nilo-Saharan. For convenience, we adopt Dimmendaal, Ahland, Jakobi & Kutsch Lojenga's (2019) proposals.

Nilo-Saharan consists of two major branches, Central Sudanic and Northeastern Nilo-Saharan, to these one may add Songhay, Koman and Gumuz (the latter two seem related). Shabo and Kadu languages are sometimes argued within the Nilo-Saharan framework, but they are separated.

The Northeastern branch consists of Eastern Sudanic and the other small branches, Saharan, Mabang, Fur-Amdang, Kunama and Kuliak. Fur-Amdang and Kunama, Eastern Sudanic (ES) and Saharan may constitute a single branch.

Eastern Sudanic consists of northern (n) and southern (s) sub-branches. The northern branch consists of Taman, Nubian, Nara and Nyimang (including Afitti), while the southern branch consists of Berta, Jebel (or 'Eastern Jebel'), Daju, Temein, Surmic and Nilotic branches.
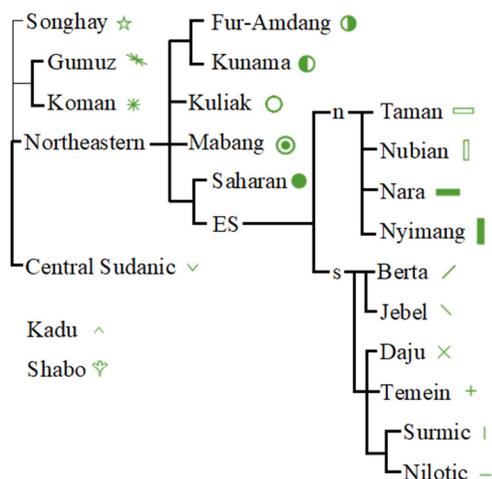
(NAKAO Shuichiro)
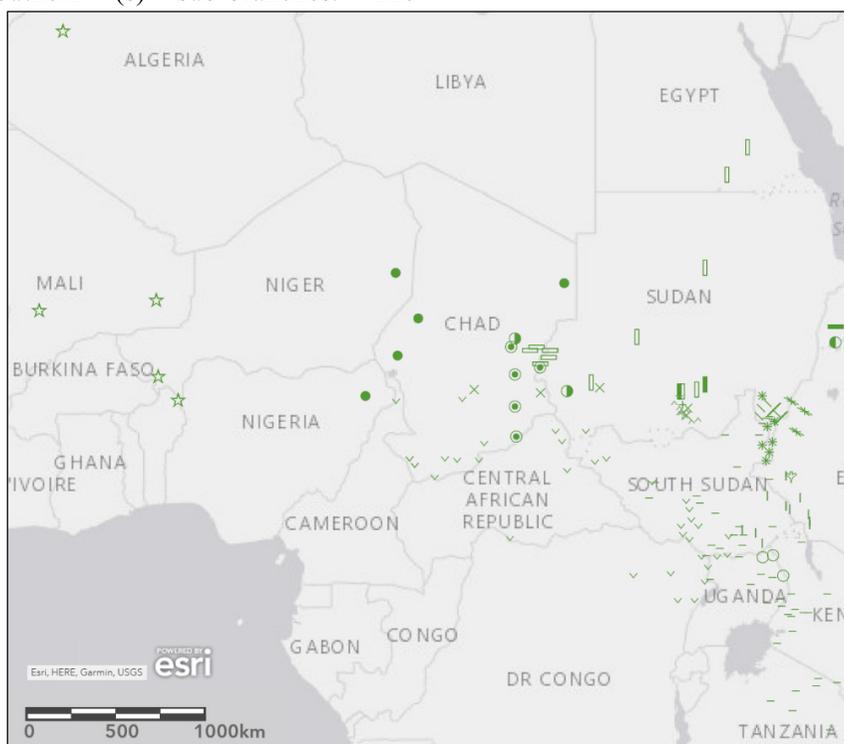
Subgrouping of Niger-Congo

The current understanding of genetic classification of the Niger-Congo languages is established on the basis of Greenberg's (1963) well-known classification of African languages, which classifies NC into six subgroups, namely Mande, West Atlantic (renamed as Atlantic), Adamawa-Eastern (renamed as Adamawa-Ubangian), Gur, Kwa, and Benue-Congo that include Bantu which was previously regarded as an independent genetic unit. Together with Kordofanian, it forms the macro-phylum originally called Congo-Kordofanian, which is equivalent to today's understanding of NC. The classification adopted in this volume follows the simplified model proposed by Dimmendaal and Storch (2016), which is based on Williamson (1989), reflecting major revisions on Greenberg (1963), including reclassification of Eastern Kwa into West BC by Bennette and Sterk (1977). Readers may refer to Williamson and Blench (2000) for a general overview of the genetic classification of NC, and to Watters (2018) for external and internal classification of East BC.

I. Kordofanian
II. Mande
III. Atlantic-Congo
  III-1. Atlantic
  III-2. Ijoid
  III-3. Volta-Congo
    III-3-i. North Volta-Congo including Gur and Adamawa-Ubangian
    III-3-ii. Dogon
    III-3-iii. Kru
    III-3-iv. Kwa
    III-3-v. Benue-Congo including Bantu

(SHINAGAWA Daisuke)

Legend:
- ⊕ Kordofanian
- 0 Mande
- ⌄ Atlantic
- ⌂ Ijoid
- < Gur
- > Adamawa-Ubangian
- ^ Other North Volta-Congo
- ◇ Dogon
- ▯ Kru
- ▭ Kwa
- ■ Bantu
- ▣ Non-Bantu Benue-Congo
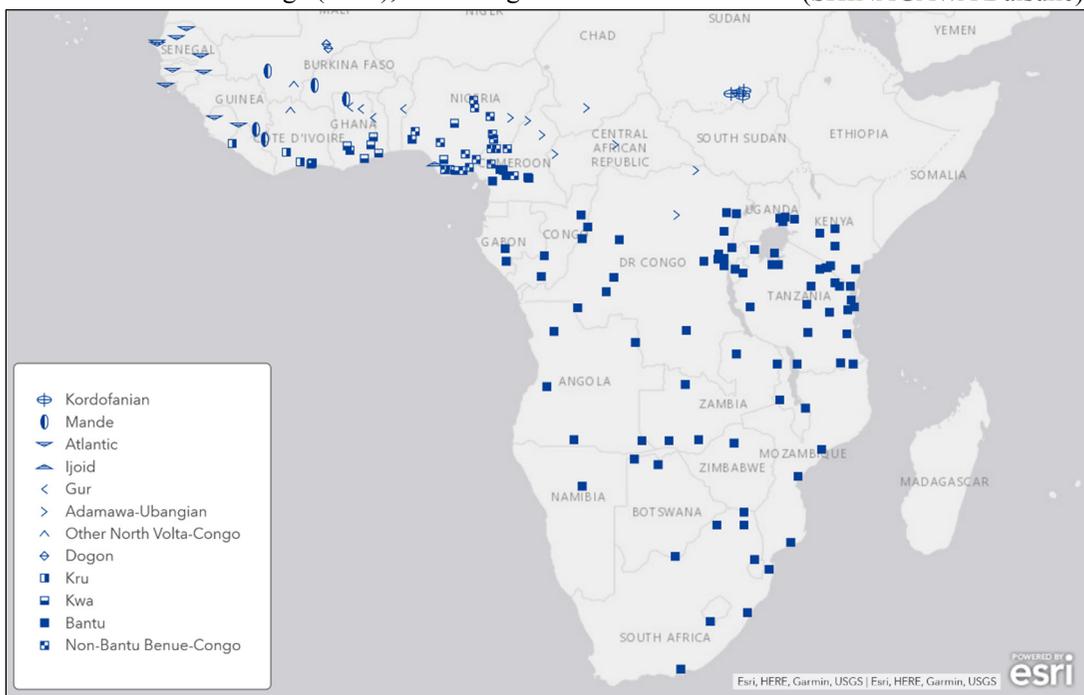
Subgrouping of languages in the Kalahari Basin area

Under the currently accepted genealogical classification presented in Güldemann (2014), the languages spoken in the Kalahari Basin area (hereafter KBA), aka Southern African Khoisan languages, are classified into three language families, namely, Tuu, Kx'a and Khoe-Kwadi. Each family consists of individual language varieties or continua of varieties called language complexes.

Table 1 summarizes the language families in KBA and their constituent subdivisions that are sampled in the present volume. Language varieties are plotted on the below map, where Khoe-Kwadi languages are marked with filled circles, Kx'a with trident marks and Tuu with downward pentagon marks.

(KIMURA Kimihiko, NAKAGAWA Hirosi)

Table 1: Subgrouping of the KBA language families.

Language family	Language (complex)	Variety
Tuu	Nǁng	Nǀuu
	Taa	West !Xoon
		East !Xoon
Kx'a	ǂ'Amkoe	N!aqriaxe
	Ju	Tsumkwe Juǀ'hoan
		Heikkinen !Xuun W
		Heikkinen !Xuun E
Khoe-Kwadi	Namibian Khoekhoe	Windhoek Khoekhoe
	Ghanzi-Hanahai	Naro
		ǂHaba
	Eastern Okwa	Xade Gǀui
		Khute Gǀui
		Gǁana
		Tshila
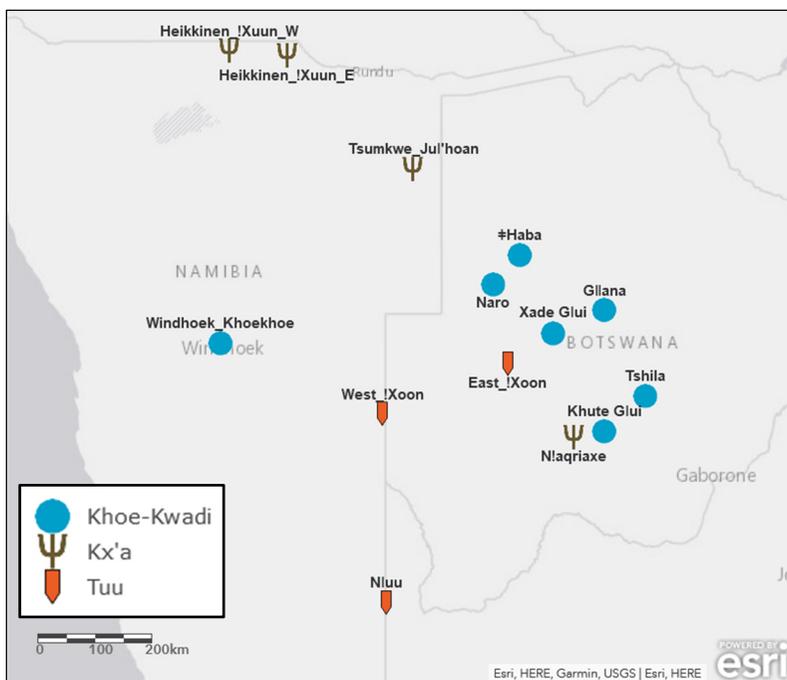
Chapter I

Rat/Mouse

'Mouse/Rat' in Asian and African languages

This chapter includes word forms for mouse or rat in various Asian and African languages. Some languages distinguish rat/mouse (e.g., Altay, Kalderaš Romani) but others (e.g., Korean, Mongolic) do not.

The words used for mouse/rat are characteristic of its diversity. For example, Utsumi (Austronesian) reported that more than half of the lexical items for mouse cannot be classified into any group because of their diversity.

Endo et al. (Kra-Dai) suggests that this is possibly due to its familiarity among people, with it being encountered frequently in day-to-day life. Another possibility, suggested by geneticist Dr. Hitoshi Suzuki, is that because of the abundant varieties of the mouse species, terms used to describe them also differ according to the variations. Fukazawa reported that Ainu distinguishes many word forms according to rat species.

Some terms for mouse/rat came from family names or taboo words. Nakazawa and Yokoyama (Japonic) introduce *oyabito "parents" or *yome "bride" forms for mouse. Iwasaki (Iranian) explains that purg- type in Pamir-Hindukush area came from *paurka- 'grey one' probably owing to avoidance of a taboo word. Names that directly describe rats were avoided in some languages, probably because rats were considered troublesome for agriculture.

(YOKOYAMA Akiko)

Table 1: Main word forms for rat/mouse.

Languages	Word forms
Ainu	*érum ~ erum*
Japonic	*nezumi* type *oyabito type
Korean	*tʃwi*
Sinitic	*ʂu* 鼠 *xau tsɿ* 耗子
Hmong-Mien	*nau*
Tibeto-Burman	*bwəy type *rwak type *b-yəw-n type
Kra-Dai	*hnu type *vau³ / fau³* type *tiu¹* type
Austroasiatic	*kni[i] type *gaŋ type
Austronesian	*buxtsi* type *koḷabaw* or *bolabow* type *tikus* type
Tunguisic	SINGE- type ČAMAKC- type
Uralic	hiir, iir, šyr type pisja type djaŋkal type
Mongolic	*xulgana* type
Turkic	sïčgan/sïčan/šăší type
Indo-Aryan	*mūṣa* type *undura* type
Burushaski	*gírkis* type
Dravidian	*eli* type
Iranian	*mūš* type *purg* type
Caucasian (rat)	v-type (e.g., *virtxa*) q'-type (e.g., *čačan-āq'o*)
Semitic	*faːr* type *antʃˈiwa* type
Nilo-Saharan	*kilt type *s'egbe type *oyio type
Bantu	*-bèbà type *-pʊ type

Abbreviations are as follows:
* : reconstructed form
italicized word : the actual word form in a certain region

'Mouse' in Chukotko-Kamchatkan

Mouse is *pipiqəɬɣən* in Chukchi and *pipiqəlʲŋən* in Alutor and Koryak (Kurebito et al. 2001), showing correspondences between laterals /ɬ/-/lʲ/ and velars /ɣ/-/ŋ/.

In Itelmen, the mouse is named *lʲelʲkʼutʃ* in the northern dialect and *lʲeɬkʼotʃ* in the southern dialect (Kurebito et al. 2001). The differences between the dialects are the lateral /lʲ/-/ɬ/ and the vowel /u/-/o/.

It is obvious that the Chukotko-Koryak group and Itelmen have different words for 'mouse', *pipiqəɬɣən~pipiqəlʲŋən* and *lʲelʲkʼutʃ-lʲeɬkʼotʃ*, respectively.

(ONO Chikako)

A. *pipiqəɬɣən~pipiqəlʲŋən* type
- ■ A-1 *pipiqəɬɣən*
- □ A-2 *pipiqəlʲŋən*

B. *lʲelʲkʼutʃ~lʲeɬkʼotʃ* type
- ○ B-1 *lʲelʲkʼutʃ*
- ● B-2 *lʲeɬkʼotʃ*
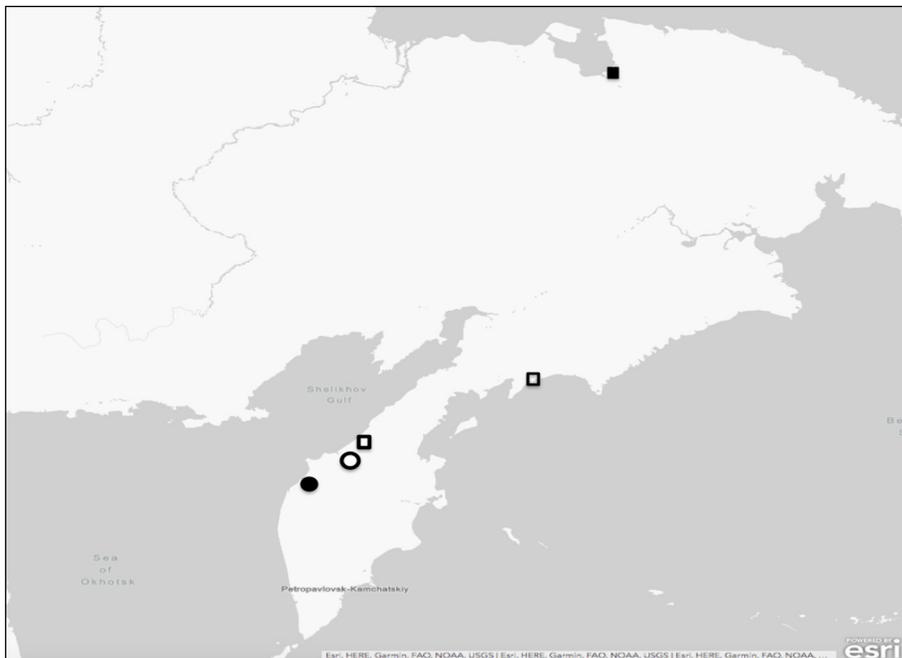
Figure 1.1.1: 'Mouse' in Chukotko-Kamchatkan.

'Rat' in Ainu

The term for 'rat' is classified into a monotonous type, that includes several sub-types. A-1 shows its wide distribution, which indicates that it could be older than the other word forms.

According to Chiri (1976 [1962]), Ainu contains different word forms, depending on the rat species: *toyérum* (lit. ground rat) and *sítoyerum* (lit. large ground rat) for the wild brown rat (*Rattus norvegicus* Erxleben); *húreerum* (lit. red rat), *irúraerum* (lit. rat carrying something), and *harúkarpe* (lit. one making food) for *Ezoakanezumi* エゾアカネズミ (*Apodemus speciosus ainu* Thomas); *niókuy ~ niyókuy* and *oníkuy* (lit. gnawing the root of trees) for *Ezoyachinezumi* エゾヤチネズミ (*Clethrionomys rufocanus bedfordiae* Thomas); and *yukérum* (lit. game animal rat) for the Japanese mouse (*Mus molossinus* Temminck and Schlegel).

(FUKAZAWA Mika)

A. *érum* type
- ○ A-1. *érum ~ erum*
- ◐ A-2. *erumu*
- ⊖ A-3. *érmu ~ erém*
- ◉ A-4. *enum*
- ⊖ A-5. *erúmun ~ erumun*

Figure 1.2.1: 'Rat' in Ainu.

'Mouse/Rat' in Japonic

Regarding the words for mouse/rat, the NEZUMI type (*nezumi*, *nẽdzïmi*, *nezun*, *nïdzïn*, *nïdïn*, *nudzumi*, …) is distributed in mainland Japan, the Amami Oshima, Kikaijima, and Tokunoshima, and the OYABITO types (*ʔwentʃu*, *uyantʃu*, *weedza*, *uyadza*, *oisya*, …) are distributed in the area south of Okinoerabujima. In addition, the YOMONO type (*yomono*, *yo no mono*, *yoru no mono*, *yoru no okata*, *yoru no hito*, *yoosa no mono*, *yumunu*, *yumuru*, *yumudza*, …) is found at several points both in mainland Japan and in the Ryukyus, and the YOME type (*yome*, *yomego*, *yomesama*, *yomezyoo*, *oyome*, *yumizyoo*, …) is also dispersed in mainland Japan and the Ryukyus. There are also other types such as OYAKE (*weeki*, *weheganasi*, *ʔweeganasi*), CYUUCYUU (*tʃiitʃii*, *tʃoitʃoi*, *zizi*, *zyuuzyu*, …), and FUKU(RO) (*ofukuro*, *ofukurosama*, *fukusan*, *ofuku*, *fukuzyoo*, *fukunokami*, …). In Japonic languages, different morphemes are not used for mice and rats, and they are distinguished by compounds such as *hatsuka-nezumi* and *dobu-nezumi*.

The NEZUMI type is distributed near mainland Japan, but OYABITO means "parent person" and is an expression that compares a mouse to a family like YOME 'bride'. YOMONO means "night creature" and originally represented not only mice but also animals such as foxes and raccoon dogs that are generally active at night. CYUUCYUU was derived from the scream of a mouse. The KAAKII type is distributed only on Kikaijima and is thought to be derived from the scream (cf. English *squeak* and Japanese *kiikii*). In this way, forms other than NEZUMI are considered to be innovative because their sources and meanings are clear, and it is presumed that *nezumi* is the oldest form representing mice in Japonic languages. The Japonic form of mouse/rat in Chinese zodiacs used in East Asia is *ne*, although it is not clear which is older, *nezumi*, or *ne*. If we follow the etymology of *nezumi* as "ground dwelling", then *ne* would have no meaning for "a mouse", as *ne* is an abbreviation for *nezumi*. However, *ne* forms compounds such as *norane* 'shrew', *yamane* 'dormouse', and perhaps *kitsune* 'fox', suggesting that *ne* is a morpheme representing "small animals". Therefore, we presume that *ne* is older than *nezumi*. It is understood that the use of relative names such as OYABITO and YOME for mice were taboo words after rice cultivation, when the damage caused by mice became severe.

Comparing *ne* with other language forms, there is *nŭu* (*h-nū*) 'rat/mouse' in Thai, and *hn-* is reconstructed from the tone and initial consonants such as *hnoc* in Sui. If the Chinese *shŭ* 'rat/mouse' is also reconstructed in a form like *hnaʔ*, it is probable that Japonic languages borrowed 'rat/mouse' from these other languages, rather than other languages borrowed from Japonic, because n- > hn- is unnatural.

(NAKAZAWA Kohei and YOKOYAMA Akiko)

	NEZUMI
	OYABITO
/	YOMONO
\	YOME
⌒	FUKU(RO)
^	CYUUCYUU
⌄	NERA
⊙	OYAKE
⊖	KAAKII
0	MAYAA
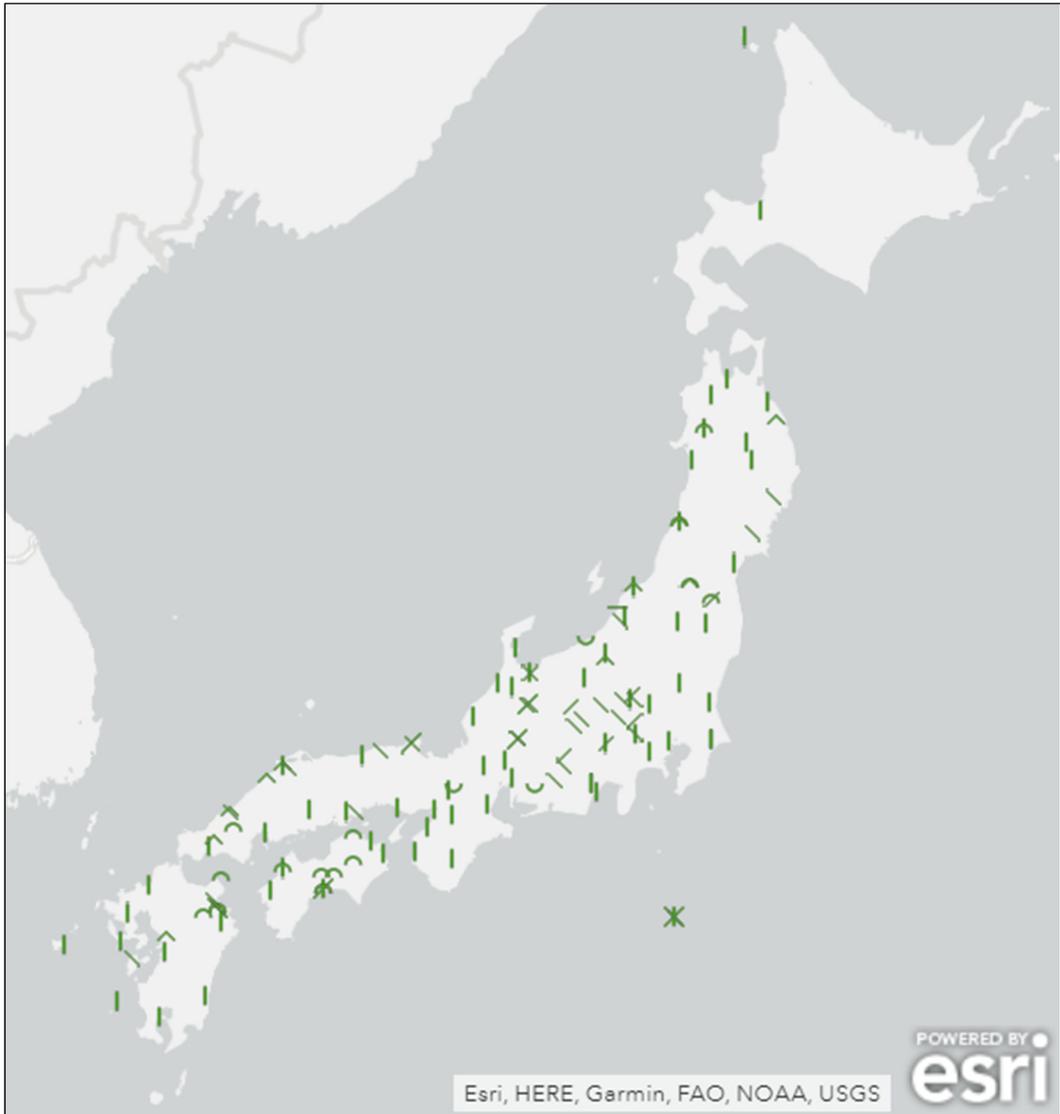
Figure 1.3.1: 'Mouse/Rat' in mainland Japan.
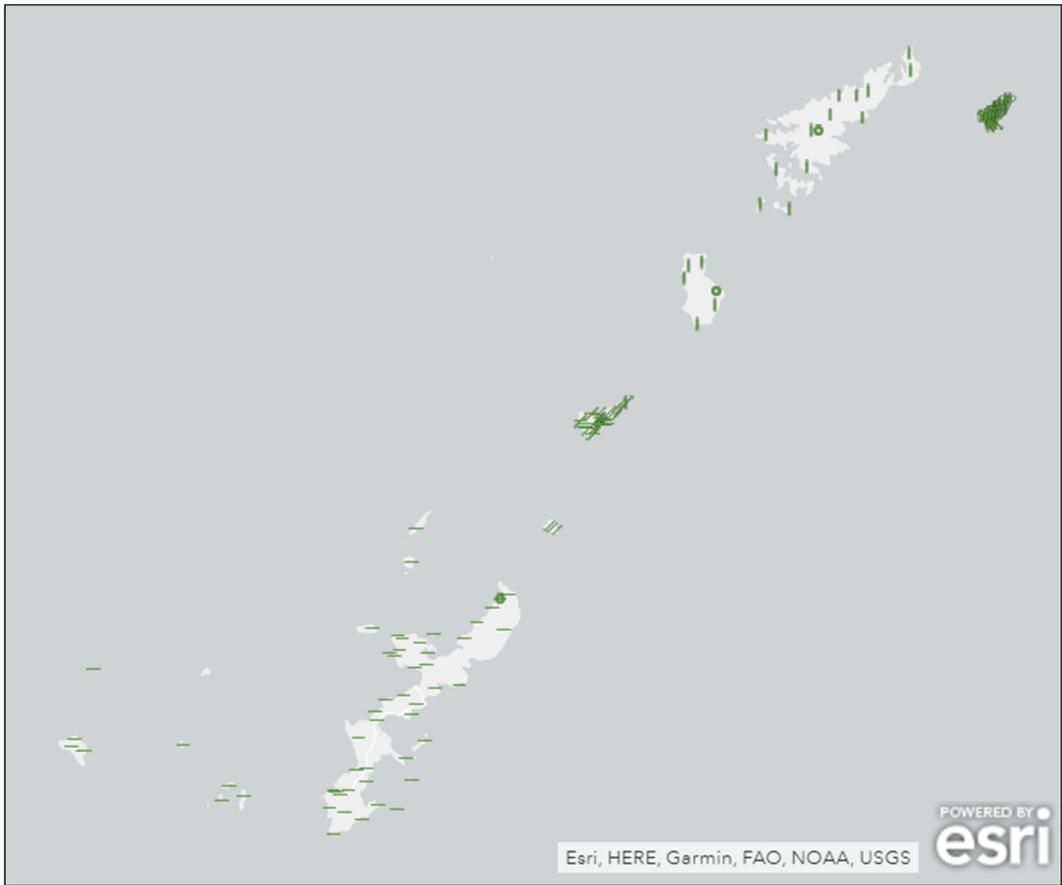
Figure 1.3.2: 'Mouse/Rat' in Northern Ryukyu Islands.
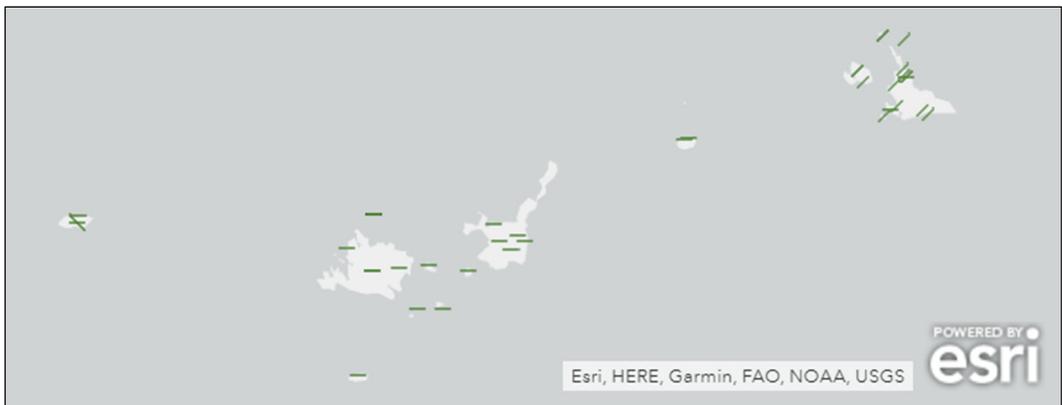
Figure 1.3.3: 'Mouse/Rat' in Southern Ryukyu Islands.

'Mouse' in Korean

Modern standard form for 'mouse' is 'cwi [tʃwi]' and Middle Korean form 'cuj [tsuj]'. Middle Korean is the language spoken from the middle of the 15th century to the end of the 16th century. These two forms are almost the same except for the phonetic realization of the diphthong and the affricate. The mouse and the rat are not distinguished.

A more exact phonetic transcription for the modern form would be [tʃɥi] (with a labial palatal approximant [ɥ], instead of [w]) in many modern dialects including Seoul, but such forms are poorly recorded as such so that we transcribe the modern form simply as [tʃwi] below.

Historically, we can have an older record than the Middle Korean one. In *Jīlín lèishì* (鷄林類事), compiled in the 12th century, the mouse is recorded as '嘴' ("鼠 曰嘴"), so that we can know that this word has not changed essentially from the 12th through the 15th centuries.

Dialect variation is not so great. First, we have a few phonetic varieties derived from the same native etymon shown above, such as the following:

A-1 tʃwi, A-2 tʃi, A-3 ʔtʃi

A-2 is made by dropping the glide [w], and A-3 by adding a glottal closure at the beginning of the word, which is a sign of sporadic reinforcement to express a kind of emphasis.

Secondly, we have a few other forms, having a totally different etymology. For example, a Sino-Korean form [sɔsæŋwɔn] "鼠生員" and its phonetic varieties may be used in many dialects. But such forms can be regarded as a kind of euphemism or a stylistic variant, mainly used for the purpose of avoiding the direct mention of the native word. Therefore these forms are not included in the Map.

(FUKUI Rei)

Figure 1.4.1: 'Mouse' in Korean.

'Mouse' in Sinitic

Based on the stem type, we classify the forms into four types, and subclassify them based on modifier or suffix.

A-1: 鼠ʂu

A-2-1: 老鼠lao ʂu（北京） lɔ tʂʰu（连云港） niãu tshu（厦门）

A-2-2: 老鼠子 nau suei tsๅ（汉源）老水子（成都）老鼠哩（浏阳）

A-2-3: 老鼠儿 lau ʂuər

A-3: X鼠

A-4: 老虫lɔ tʂʰuŋ（银川）

A-5: 老子 lɔ tsๅ（宁波）

A-6: 老伯lɤ pa (休宁)

B-1：耗子 xau tsๅ (承德)

B-2：老耗子 lɔ xɔ tɤʔ（长治） lɔ xɔ lɤʔ (平顺)

B-3: 土耗子 thu xɔ tθๅ（诸城）

C-1: K-L-: ku zʌʔ(平遥) kur(孝义) kəʔ lʌr tɕiʌ(离石) kəʔ lʌ tɕiəʔ(临县) kəur / mau kəur(汾阳) kə lɑ ll tə (汾西) kaŋ kɤʔ lʌʔ(陵川) kə lau ia (吉县) kur（离石）ku zๅ tɯ(万荣)

C-2: 高客kau khæ （红安） kau kʰɤ (tsๅ)(长沙) kɤ khe (双峰)，客人kʰɤ nən（武汉）， 高大爹kau ta tie（武汉）

C-3: 老鼠窟窿（太原）

C-4: 老鼠固 lɔ ɕy ku(新化)

D-1: others: 财神tshai sən（南昌）尖尖嘴 tsieŋ nzieŋ nzuei（福州）梁上君子（成都） 雕花老师（庆元）

A types have or had 鼠 in their stems. The onset of 鼠 ordinally possess alveolar, postalveolar, or retroflex fricative, however, in many dialects, it becomes affricative like [ts-], [tsʰ-], [tʂʰ-] etc. Distribution of affricative onset spread into large area, in the north west, south west, central and south east area (曹 2008:音:106). Most of A types have modifier (or prefix) 老 in the first syllable. Monosyllable type A1 is sporadically distributed in southern area. 老鼠 denoted mouse in Tang period but in Han period, some dialects used 老鼠 for denoting bat.

B types have 耗 in their stems. 耗 ordinally possess velar fricative [x-]. B types are widely distributed in the northern area. 耗 means 'exhaust' and people may think because mouse eat food so named 耗子. However, true etymology is not clear.

C types have k- onset syllable in the words. These types are mainly distributed in Shanxi, Hubei, Hunan and Sichuan province. These types are found around B 耗 type therefor these two types may be related, however further research is needed.

D types include other types.

Reconstructed forms of 鼠 of Middle Chinese and Old Chinese are shown below.

鼠	1	2	3	4
MC	ɕĭo	ɕi̯wo:	-	-
OC	ɕĭa	ɕi̯o	-	nhaʔʔ

1:郭錫良(2010), 2: Karlgren (1957[1997]), 3: Baxter & Sagart (2014), 4: Schuessler (2007)

(YAGI Kenji)

A. *Shu* type

⭕ A-1 鼠 ʂu

◦ A-2-1 老鼠 lao ʂu, lɔ tʂʰu, niãu tshu

⬒ A-2-2 老鼠子, 老水子 nau suei tsɿ

◐ A-2-3 老鼠儿 lau ʂuər

⊕ A-4 老虫 lɔ tʂʰuŋ

⬡ A-5 老子 lɔ tsɿ

B. *Hao zi* type

△ B-1 耗子 xau tsɿ

▲ B-2 老耗子 lɔ xɔ təʔ, lɔ xɔ ləʔ

◤ B-3 土耗子 thu xɔ tθɿ

C. K- type

∠ C-1 K-L- type: ku zʌʔ, kur, kəʔ lAr tɕiA, kəʔ lA tɕiəʔ, kəur, mau kəur, kə lɑ lI tə, kaŋ kɤʔ lʌʔ, kə lau ia, kur, ku zɿ tɯ

╱ C-2 高客 kau khæ, kau kʰɤ (tsɿ), kɤ khe, 客人 kʰɤ nən, 高大爹 kau ta tie

⌣ C-3 老鼠窟窿

⌣ C-4 老鼠固 lɔ ɕy ku

D. other type

✳ D-1 财神 tshai sən, 尖尖嘴 tsieŋ nzieŋ nzuei, 梁上君子, 雕花老师
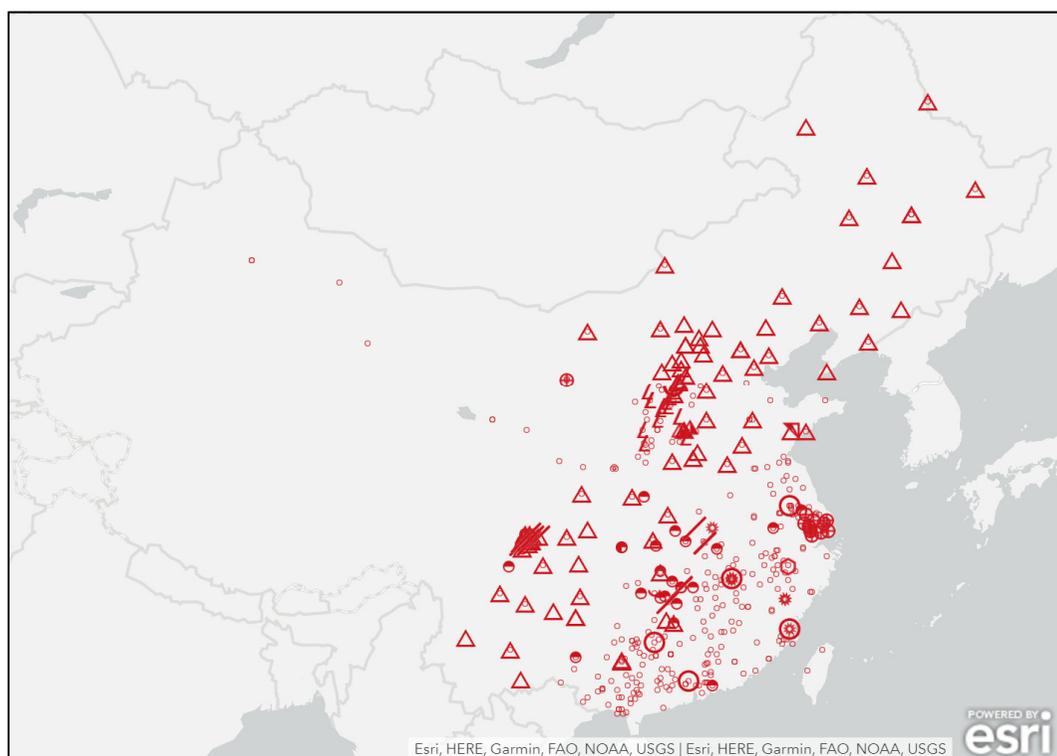
Figure 1.5.1: 'Mouse' in Sinitic.

'Mouse' in Hmong-Mien

There are six types in MOUSE: A: *nau*; B: *bra*; C: *du*; D: *klai*; E: *kjaŋ*; F: *tɕuŋ*.

Since Type A has the widest distribution, and is spread across both Hmongic and Mienic, it must represent the most archaic state for the entry. Type B is concentrated in the western part of Guizhou, suggesting that this form is an innovation that occurred in that area. Note that one lect, Dananshan shows both Type A and Type B forms.

Type C displays a rather vast geographical distribution in Guangxi and Hainan Island. However, this type has only been observed in the Kim-Mun group belonging to Mienic. Therefore, the extension of the distribution can be attributed to the recent emigration of the group, not the retention of the archaic form.

(TAGUCHI Yoshihisa)

Figure 1.6.1: 'Mouse' in Hmong-Mien.

'Mouse' in Kra-Dai

Type A is widespread among Tai and Dong-Shui branches. Li (1977) reconstructs the proto-Tai form as *hnu A1.

Noteworthily, there is a peripheral distribution among the Buyi language, located in the northernmost part of the Kra–Dai area. Type X (vau^3 and fau^3) is in the central area. Type D (la^{31}, etc.) surrounds type X, and type C (nai^5, etc.) is distributed along the outer sides of type D. Thus, according to the principle of peripheral distribution, it is possible to infer the formation process of word forms as type X > type D > type C.

Type T (son^3, etc.) surrounds type S (ron^3) (another peripheral distribution) though its area is smaller. It is possible to reconstruct the changing process, type S > type T. It is also probable that a sound change s- > r- occurred.

In Hainan Island, type R (tiu^1, etc.) is widespread in the Li branch, while type Q (niu^1) is found in a location. According to the phonological change tendency in Li, type Q is an older form, and denasalization occurred in type T.

The other types, B, E, F, G, H, I, J, K, L, M, N, O, P, U, V, and W, are scattered in isolated locations; hence, it is difficult to infer their formation process.

Mouse shows many varieties in comparison with the other animals. It is possibly due to its familiarity among people, being a frequently seen small animal in daily life. Furthermore, Dr. Hitoshi Suzuki pointed out another possibility that there are abundant sorts of mouse, so their terms differ according to the difference in sort.

(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)

A: lu^1, ðɯ^6, mei^6 nu^1, naɯ^1, nau^1, naw^1, nɔu^1, nəu^1, no^3, nɔ^3, n̥ɔ^3, tə no^3, tə^{0}n̥o^3, no$^{3'}$,n̥o^3, nou^1, now^1, nu, nu:^{A1}, nu^{1}la:n^2, nū^6, nuu^1, nuu^2, nuu^5, lak^{8}nu^4, nu^{1}ða:n^2, ni$^{23\text{-}11}$n̥o^{31}

B: na^4

C: nai^5, nɑi^5, qa^{33}nai^{33}

D: la^{31}, lɒ^{55}, lɔ^{55}

E: lja^{44}, ʔa^{3}lja^{45}

F: a^{0}θi^{24}

G: boi^{11}

H: ɗɔ^{33}ɕi^{33}, qa^{0}ɕi^{54}

I: kɯ^{2}keu^4

J: kui^3

K: ma^{34}sa^{44}kuɤi^{44}

L: ma^{4}ha^2

M: mai^{4}zi^2

N: mau^{53}, mə^{53}, mu^{53}

O: mo^{42}lo^{42}

P: ne:ŋ^1

Q: niu^1

R: teu^1, thiu^4, ti:u^4, tiu^1, tiu^4

S: ron^3

T: sɑn^3, son^3

U: tiau^{55}su^{55}

V: tsɔi^4

W: tu^{2,4}kji:u^3

X: vau^3, fau^3

Figure 1.7.1: 'Mouse' in Kra-Dai (enlarged).
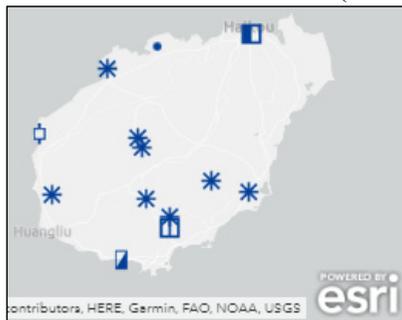
Figure 1.7.2: 'Mouse' in Kra-Dai (Hainan).
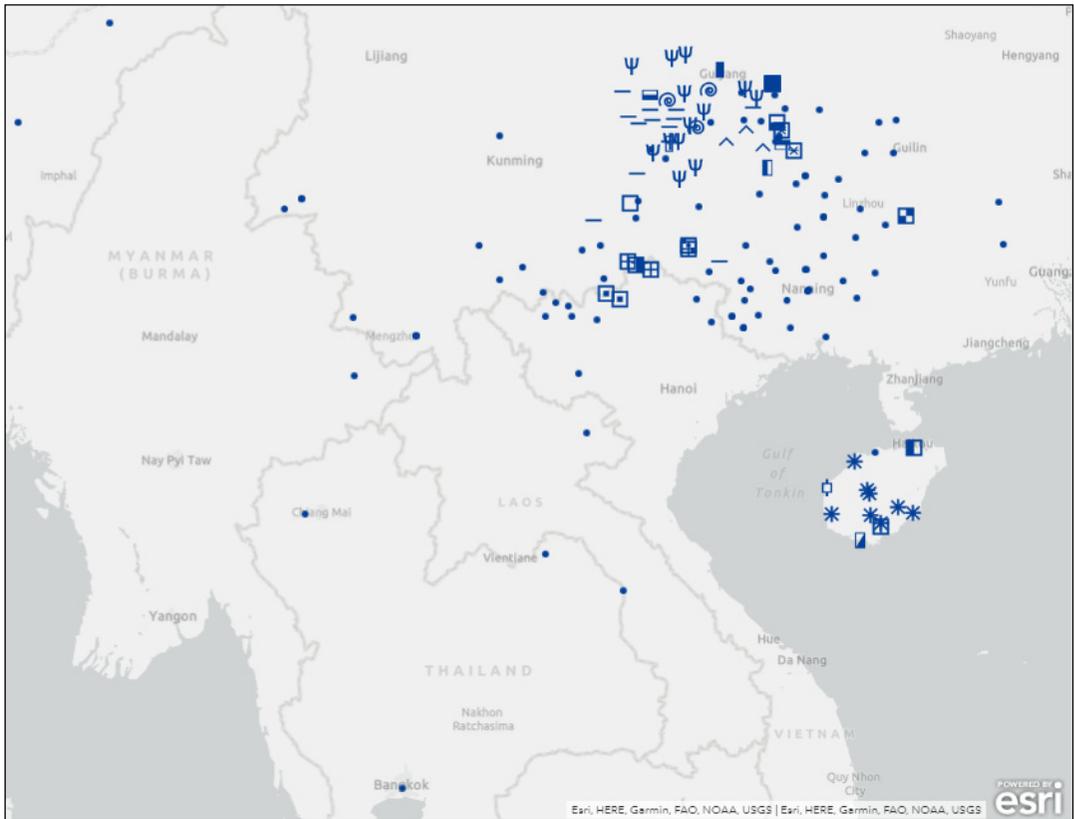
Figure 1.7.3: 'Mouse' in Kra-Dai.

'Mouse/Rat' in Tibeto-Burman

There are four major stems (word roots) for 'mouse/rat' in Tibeto-Burman (TB). Three of these stems trace back to etyma at the Proto-Tibeto-Burman (PTB) level (see STEDT). The word forms for 'mouse/rat' contain word formations, which consist of a single stem, a stem plus an affix, reduplication of a stem, and a compound of two stems. We classify the TB word forms for 'mouse/rat' first into stem types and then into compound types.

Type A is *bwəy (BAMBOO RAT) in the PTB etymon, and Type B is *rwak (RAT) in the PTB etymon.

Type C is derived from *b-yəw-n (RAT/RABBIT/HARE) in the PTB etymon. A similar root, Type A *bwəy, was reconstructed with the meaning 'bamboo rat', so it is often quite difficult to assign a particular reflex to one or the other (STEDT). There are two Chinese comparanda, one with an open syllable 鼠 shu (OC *śįo) and one with a final *-n ('hare' OC *tsɹwən ⪢ *tsʼɹwən) (STEDT).

The etymology of Type D ts is unknown. This word form is often reduplicated.

There are several marginal roots labelled as Type X. cuhā is an Indo-Aryan etymon. sya-n and the first syllable of ṣa-bi-lig are related to *sya-n (FLESH/MEAT/GAME/ANIMAL) in the PTB etymon. tsak might be related to *tsak (FILTHY/DIRTY/RAT) in the PTB etymon or *tsakᴸ (FILTHY/DIRTY/RAT) in Proto-Loloish. The Bai forms sv and ʂv are potentially related to the Chinese 鼠 shu. The other word forms are etymologically unknown. Most of the forms are monosyllabic (with an affix), but

we also found nine types of compound forms: A-compound type (A+B, C+A, and A+D), B-compound type (B+C, C+B, B+D), and B+X (B+ cuhā or tsak), and C-compound type (sya-n+C).

Type A is found in Tibetic (eastern and western parts), Monpa, Lamo, and two sMar languages, whereas Types B and C are widespread across the branches of TB. Type B is found in Qiangic, rGyalrongic, and Lolo-Burmese. The B-compound type (B+X(tsak)) is found in Hani. Type C is found in the central-eastern part of the Tibetosphere (Qiangic and rGyalrongic), at the eastern edge (Tujia, which has been said to have migrated from the Qiangic area), in southern Burma (Karenic), in the North Eastern Indian Areal Group (northern Naga, central Naga, Kuki-Chin, and Jingpho-Asakian), and in Kinnauri. From the distribution, we can conclude that Type C is chronologically oldest. Type D is found in several dialects of Tibetic. Moreover, A-compound type (A+D) is found in central, western, and southern parts of the Tibetosphere (Dingri, Dzongkha, Chiktan, and Nurla dialects of Tibetan).

Word forms for 'mouse/rat' are notably diverse, including various types of compounds. They are mostly native words—that is, loan words are rarely seen. However, the four major stems cover most of the TB area from a geolinguistic perspective.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

✓ A. *bwəy type

wɛo, ɕu³⁵, byoba, etc.; tɕi³⁵po⁵³, ɕu³⁵po⁵³, tɕəwa, ɕəwa, səwa, ᴾtsə wa, ɕwa, etc. (suffixed); a sɿ, etc. (prefixed).

◪ A-compound type

A+B: ptɕitsi, pitse; C+A: zu-bǔy; A+D: phu⁵⁵se⁵⁵.

○ B. *rwak type

xɑ, xe³³, xɪ³³, he³³, ho¹³, fA⁵⁴, gɤ́, ɕyɛ³⁵, ɣo⁵⁵, wu⁵⁵, ɦo⁵⁵, hɛ³³, cwaʔ, ɣuk³¹, kʐoʔ⁵⁵, krwɔʔ, χwɑᶠ, etc.; hɔ̄-tam, ʃi-ok, hoʔ⁴tɕha², ʃiŋyoʔ, ɣɔʔ³¹naʔ³¹, hu³³jum⁵⁵, rapha, ʐa³³pha⁵⁵, etc. (suffixed); ˈdixy, a⁴⁴he³³, etc. (prefixed).

⊗ B-compound type

B+C: rok-yuʔ, hɛ̃³⁵be³¹; C+B: zexʉ, pi⁵³juk⁵⁵; B+tsak: hu³³tam²¹, xu³tsa³¹; B+ts: fɑ⁵⁴tɕhɑ²¹, B+X(cuhā): ˈyi cu, etc.

❘ C. *b-yəw-n type

pi⁵³, piu, pᵊju, bei, byu, yö, ju¹¹, yú, jỳ, ju¹¹, ju, yú, ju⁵⁵, jwɪ̄, ʔpɣʰu, zùu, šʰu, gyuq, etc.; myu ma, yi buk, zi ro, yu pu, yuu⁴, jû, jūu, jû, jú, zu nam, yu-tśa, pɔ jɔ, etc.; (suffixed); zu che, a yi, pʻa-yü´´, pā-zú, kəyvu, kəyù, kuyuk, thüzo, phidžu, etc. (prefixed).

⊕ C-compound type: X (sya-n)+C: sà-zû.

□ D. ts type

tsəgə, tsəɣə, tsuʔyu, etc. (suffixed); tsi-tsi, tshi:tsì, tsə tsə (reduplicated).

∨ X. others

sv, ʂv, cuhā, mudɯ, tsak, nyimu, sà-bi-lig, sya-n, hahta, ŋinsi, kɑtɕingu, cùʔ, me-se, kalók, kjɪ̃²¹nɔʔ³¹, suba, uchi, atɕha, dɯt, yongmüza, azhi, shiok, zu¹sa¹, ŋĕ²¹noʔ²¹, etc.
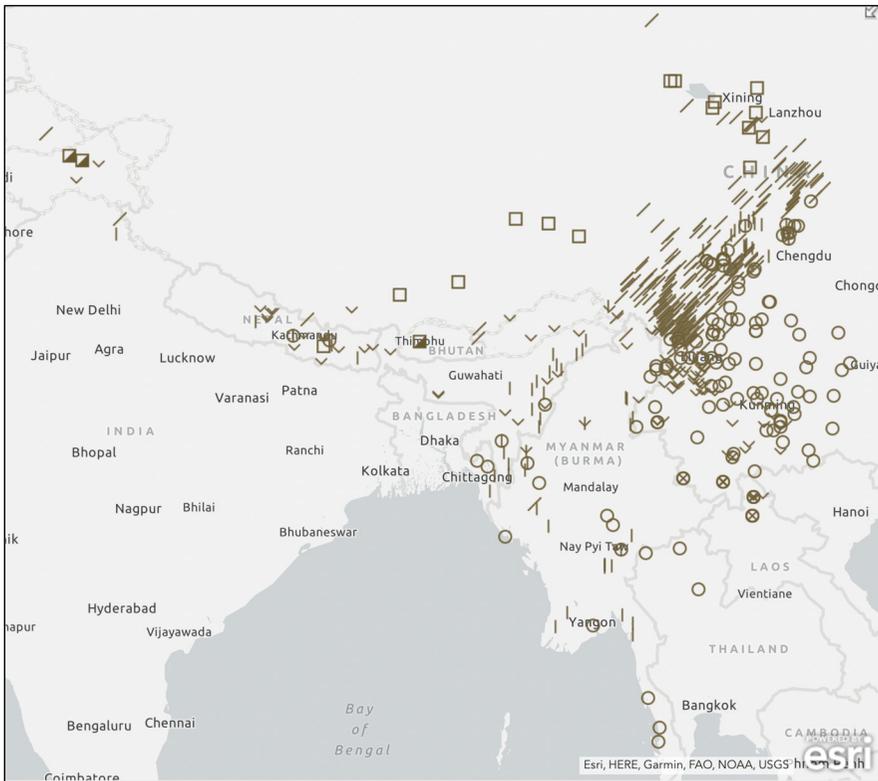
Figure 1.8.1: 'Mouse/rat' in Tibeto-Burman.
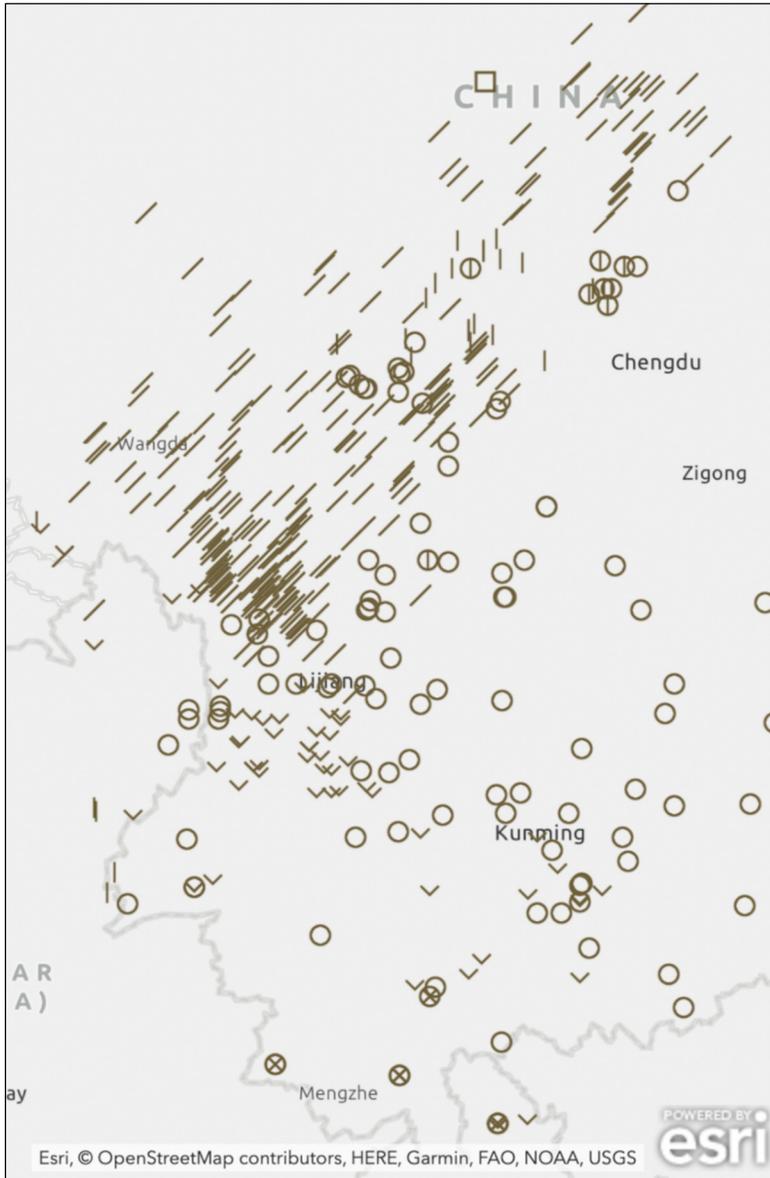
Figure 1.8.2: 'Mouse/rat' in Tibeto-Burman (detailed).

'Rat, Mouse' in Austroasiatic

The word forms meaning "rat" and/or "mouse" in Austroasiatic are classified into five types and others, as follows.

A. *kni[i]ʔ type*
Proto MK: *kni[i]ʔ (Shorto 2006)
Proto Bahnaric: *knɛː (Sidwell 2011)
knɛː (Laven), kənɛɛ (Bahnar), kənɛː (Halang)
Proto Khasic: *kʰnaaj (Sidwell 2012); kʰnaːj (Khasi)
Proto Khmuic: *kneʔ$_1$ (Sidwell 2013); kneʔ (Khmu [Cuang]), knɛːj (Phong), kənéʔ (Khmu [Yuan])
Proto Monic: *kniiʔ (Diffloth 1984); hníiʔ (Nyah Kur [Central]), nɔeʔ (Mon)
Proto Palaungic: *kni(i)ʔ (Sidwell 2010); kᵊnɪ^{1,3} (Danaw)
Proto Vietic: *k-neːʔ (Ferlus 2007); kənaj^3 (Malieng), kəneː^3 (Chứt [Rục])
Aslian: kʌkʌni (Semai) 'mice'
Katuic: (k)naj (Kui), kanaj (Bru), knaj~naj (Kui)
Khmeric: knʌr (Surin Khmer)
Pearic: knaːj (Pear)
Munda: kɔnɛ (Kharia)

B. *gaŋ type*
Proto Palaungic: *gaŋ (Sidwell 2010); kaŋ⁻ (P'uman), kiaŋ² (Wa [Thung Va])

C. *ʔiːk type*
Proto Vietic:*ʔiːk (Ferlus 2007); ʔiːk^1 (Thavung)

D. *ʔoːŋ¹ type*
Mangic: ʔoːŋ^1 (Mang)

E. *kumit type*
Nicobaric: kumit (Car)

F. *Other types* (Munda)
cuʈu (Mundari), gə-rap (Bondo), sarga (Santali), puci (Korku), …

The proto Mon-Khmer form for "rat" and/or "mouse" is reconstructed as *kni[i]ʔ by Shorto (2006). This form is succeeded quite widely among Bahnaric, Khasic, Khmuic, Monic, Palaungic, Vietic, Aslian, Katuic, Khmeric, Pearic and Munda. The initial cluster *kn- is well preserved among them, with some cases in which the shwa vowel is inserted between them, as in Bahnar, Halang, Khmu [Yuan], Vietic and Aslian. Variations are observed in vowels and finals. *i[i] is preserved in Monic and Palaungic, while *i[i]>e(ː) occurred in Khmuic and Vietic and *i[i]>ɛ(ː) occurred in Bahnaric and Munda. It is noteworthy that the final ʔ never coexists with diphthongs a(a)j and ɛːj, which indicates the nature of final ʔ as a final consonant that functions the same as j in the diphthongs.

As for the B type, the proto Palaungic form is reconstructed as *gaŋ by Sidwell (2010), while P'uman and Wa [Thung Va] reflect the initial consonant as a voiceless k.

(SHIMIZU Masaaki,
MINEGISHI Makoto)
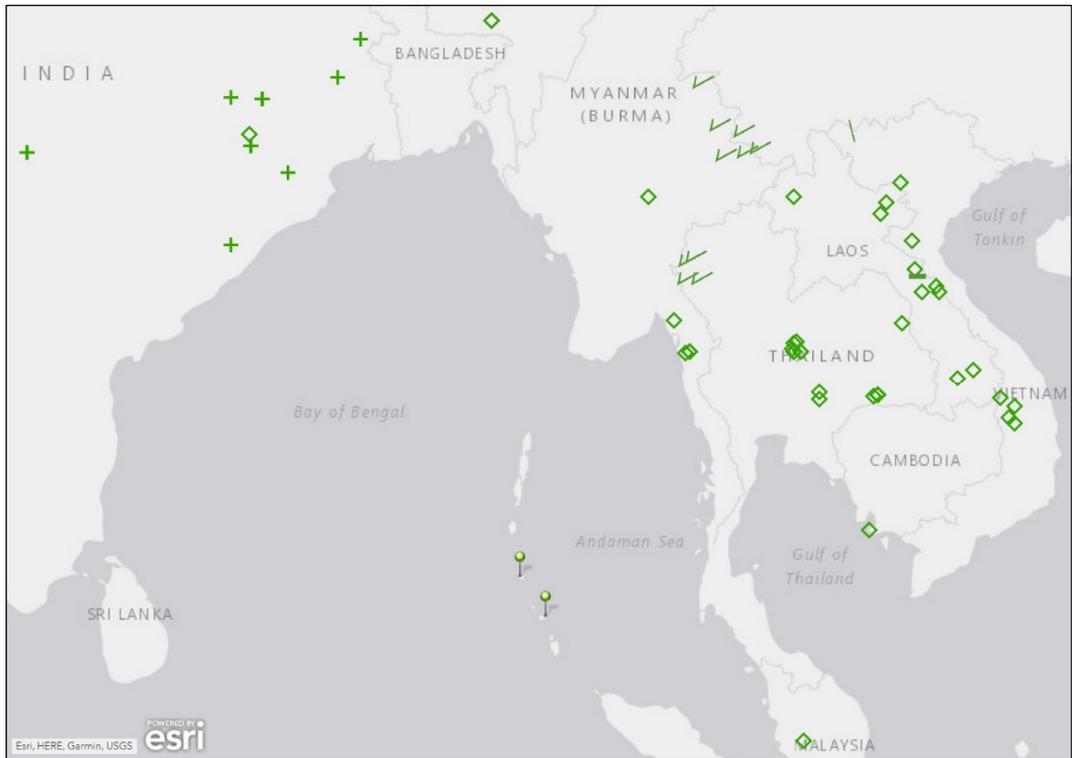
Figure 1.9.1: 'Rat, mouse' in Austroasiatic.

◇	A	*kni[i]ʔ type*
⌐	B	*gaŋ type*
–	C	*ʔiːk type*
\	D	*ʔoːŋ¹ type*
⚲	E	*kumit type*
+	F	(others)

'Mouse' in Austronesian

The form of the lexical item that denotes "mouse" varies widely across Austronesian languages. In the data of this survey, more than half of the lexical items for "mouse" cannot be classified into any group with shared features because they are so diverse in form; these are classified as Type G.

Nevertheless, Types A, B, and C show a certain similalities. Type A forms begin with a bilabial consonant (/b/ or /ɓ/) and /t/, such as *buxʦi* (Tsow), *butit* (Isnag), *birit* (Sundanese), and *ɓokoti* (Wolio). One lexical form, *ūtut* in Kalinga Limos, is classified into Type A because its form resembles *butit* in Isnag and the two languages are close to each other. Type B forms have a word-final syllable that begins with /b/, /ß/, /v/, /w/, or /h/ and are predominantly trisyllabic, Words that have a word-final syllable that begins with /b/, /ß/, /v/, /w/, /h/, or /s/ and are trisyllabic. Typical forms of this type are /koḷabaw/ (Rukai) and /bolabow/ (Molbog): *koḷabaw* (Rukai), *kuḷavaw* (Paiwan), *bɔlabɔw* (Palawan), *bolabow* (Molbog), *ambaw* (Kagayanen and Bangingi Sama), *ßulehuʔ* (Uma), *valesu* (Da'a), *balawo* (Bugis), *balaho* (Konjo), *kalaßo* (Eastern Fijian), *kuðuße* (Western Fijian), and *vualavu* (Malagasy Merina). Types A and B are found in Taiwan, the Philippines, and Indonesia.

Type C forms have a word-initial /t/ and are predominantly disyllabic. A typical form is /tikus/ (Indonesian). Other forms include *tikus* (Murut, Javanese, Madurese, Sasak, and Indonesian), *tikoih* (Acheh), and *titasi-t* (Buru). Type C appears in Sumatra, Java, and the Maluku Islands.

Types D, E, and F consist of forms from two or three languages, and found in the Pacific Islands. Type D forms begin with /g/, such as *goub* (Takia), *gʷoʔua* (Lau), and *gasifou* (Kwaio). Type E forms contain /c/ or /ç/ and a bilabial consonant. Examples are *cībʷi* (Nemi and Cémuhî) and *çībɨ* (Xârâcùù). Type F consists of *kioʔe* (Rapanui) and *ʔiore* (Tahitian).

Other various forms, which are categorized as Type G, appear in every part of the Austronesian region. Examples include *ḵawɬit* (Atayal), *kaẓam* (Atayal), *dagaʔ* (Tagalog), *ɬaŋgam* (Aaklanon), *uŋe* (Sarangani Blaan), *baguduŋ* (Batak Toba), *maɲʧiʔ* (Minangkabau), *bikul* (Balinese), *ʔudu* (Gorontalo), *davo* (Manggarai), *deke* (Ngada), *teʔu* (Sika), *lafo* (Roti), *ŋiro* (Manam), *kari* (Dami), *kuzi* (Mbula), *mwadɛʔ* (Yabem), *dʒu* (Kaulong), *gələŋ* (Tolai), *mun* (Buang), *iraʔpurup* (Adzera), *kikoni* (Kilivia), *itala* (Tawala), *bita* (Motu), *inema* (Mekeo), *kurezu* (Roviana), *nakʰude* (Maringe), *ɣarivi* (Raga), *asu* (Paamese), *kawe* (Lewo), *xasu* (Port Sandwich), *kahap* (North Tanna), *iesukw* (Kwamera), *yīpuu* (A'jië), *waxeli* (Nengone), *te kimoa* (Kiribati), *kicrik* (Marshallese), *kiṯik* (Ponapean), *xeṣi* (Woleasian), *piʧfa* (Rotuman), *kumā* (Tongan), *isumu* (Samoan), and *kimoa* (Mele-Fila).

(UTSUMI Atsuko)

● A: *buxtsi, butit, bɨrit, ɓokoti, ūtut*

◪ B: *koḷabaw, kuḷavaw, bɔlabɔw, bolabow, ambaw, ßulehuʔ, valesu, balawo, balaho, kalaßo, kuðuße, vualavu*

✳ C: *tikus, tikoih, titasi-t*

◥ E: *cībʷi, çībɨ*

➡ F: *kioʔe* (Rapanui) and *ʔiore*

✚ G Other forms: *ḳawłit, kaẓam, dagaʔ, łaŋgam, uŋe, baguduŋ, maɲʧiʔ, bikul, ʔudu, davo, ðeke, teʔu, lafo, ŋiro, kari, kuzi, mwadɛʔ, dßu, gələŋ, mun, iraʔpurup, kikoni, itala, bita, inema, kurezu, nakʰude, ɣarivi, asu, kawe, xasu, kahap, iesukw, yīpuu, waxeli, te kimoa, kicrik, kiʈik, xeʂi, piʧa, kumā, isumu, kimoa*

↘ D: *goub, gʷoʔua, gasifou*
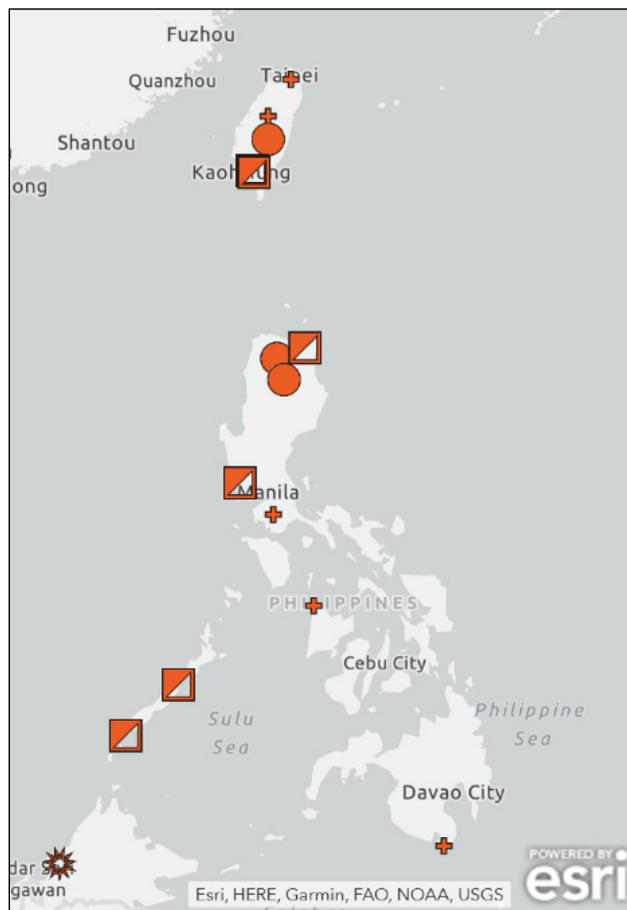
Figure 1.10.1: 'Mouse' in Taiwan and the Phillipines.

Figure 1.10.2: 'Mouse' in Indonesia.

Figure 1.10.3: 'Mouse' in Papua and the Pacific.

'Mouse' in Tungusic

The Tungusic word forms would be classified in some types as below:

   A SINGER-
   B ČAMAK-
   C ənikən (Orochon)
   D xologna (Ewenke dialect)
   E ašiččaŋ (Ewenke dialect)

The type A widely spreads in the western Siberia in Evenki, *siŋereken*, to which form the diminutive suffix *-chan* is added. Most of other Tungusic as Negidal, Udehe, Nanay, Oroch, Ulcha and Sibe are belonging to this type. The type B is secondly spreads in the eastern Siberia in Ewen, *chamakchan*. The other types are only in the North-eastern China as in Orochon and the dialects of Ewenke.

(MATSUMOTO Ryo)

○ A SINGER-
● B ČAMAK-
⊖ C ənikən
▽ D xologna
◿ E ašiččaŋ

Figure 1.11.1: 'Mouse' in Tungusic.

'Mouse' in Uralic

In Uralic there are various forms for 'mouse'. Here I just classified according to the sound forms, it could not be referred to the reason why such many forms they have. The word forms would be classified in some types as below:

A Finnish *hiir*, Livonian *iir*, Mordvin *šyr*
B Sami *sáhpán*
C Mari *kolja*
D Hungarian *egér*

E Mansi *porsuj*
F Khanty *miŋxar woj*
G Selkup *tama*
H Forest Nenets *djaŋkal*
I Tundra Nenets *pisja*
J Enets *tobik*

(MATSUMOTO Ryo)

＼	A	HIIR	☆	F	*miŋxar woj*
●	B	*sáhpán*	⊠	G	*tama*
＋	C	*kolja*	⊕	H	*djaŋkal*
⊖	D	*egér*	□	I	*pisja*
Ψ	E	*porsuj*	■	J	*tobik*
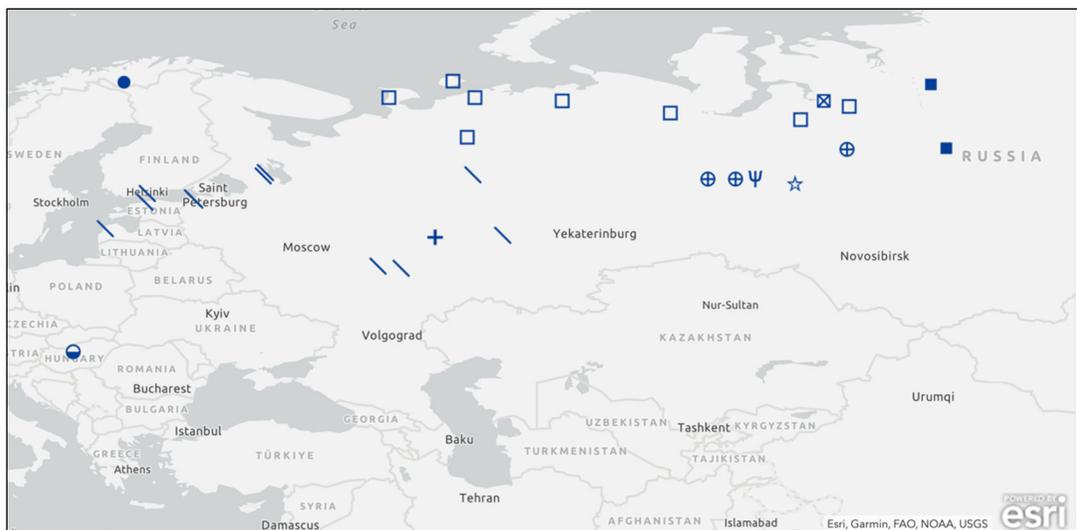
Figure 1.12.1: 'Mouse' in Uralic.

'Mouse/rat' in Mongolic and Turkic

1. Mongolic

The *xulgana* type is dominant and the other types are observed only in the peripheral regions: the *ačikčān* type in Dagur in the northeastern periphery and its modern branch in Xinjiang; *čičixaŋ*-type words and the form *laušə* (< Chinese 老鼠) in some languages in Qinghai and Gansu provinces.

Mongol has the form *xarx* for 'rat,' but *xulgan* is a generic term and covers both 'mouse' and 'rat.'

2. Turkic

Words of *sïčgan* and *sïčan* types and the form *šăší* in Chuvash are cognate, and they are spread among most Turkic languages.

Turkish has another form, *fāre*, which is from Arabic *fār* 'mouse; rat.' According to dictionary definitions, the forms *sïčan* and *fāre* do not seem to be clearly distinguished, but as shown in Savaşçı (2008: 290), they are used for 'rat' and 'mouse' respectively in modern Turkish.

Azeri has *sičan* for 'mouse' and *sičovul* for 'rat.'

The other forms are found in the eastern half of the Turkic-language distribution area:

In southern Siberia, Tuvan and Khakas have *küske*-type words. Tofalar has the form *mïrnēšqa*. Altay distinguishes *čïčkan* 'mouse' and *erlen* 'rat.'

In northeastern Siberia, Sakha and Dolgan have *kutujax*-type words.

Salar in Qinghai and Gansu provinces has *geme*. (Turkish also has *keme* in its provincial dialects. Cf. Turkish *kemir-,* Mongolic *kemele-*, 'to gnaw.')

(SAITÔ Yoshio)

A. *xulgana* type
- ⌢ xulgana, kulgana, xulganǣ, xuluganā, xulgan, falgan, xolganan, xunaglag, xunagla

B. *ačikčān* type
- ⬚ ačikčān, ačikčāŋ, ačigčā

C. *čičixaŋ* type
- | čičixaŋ, sυıdẓaγan

D. loanword
- ╱ laušə (< Chinese)

E-1. *sïčgan* type
- — sïčgan, tïčkan, sïskan, čïčxan, čïčqan, tïšqan, šïškan, sičqon, čašqan, sustan

E-2. *sïčan* type
- ‖ sïčan, sičan, sïččan, sëčan, šïčan, şiγan

E-3. šăší
- | šăší

F. *küske* type
- ∠ küske, güske

G. *kutujax* type
- ✛ kutujax, kutujak

H. mïrnēšqa
- ⋀ mïrnēšqa

I. geme
- ▽ geme

J. loanword
- ○ fāre (< Arabic)
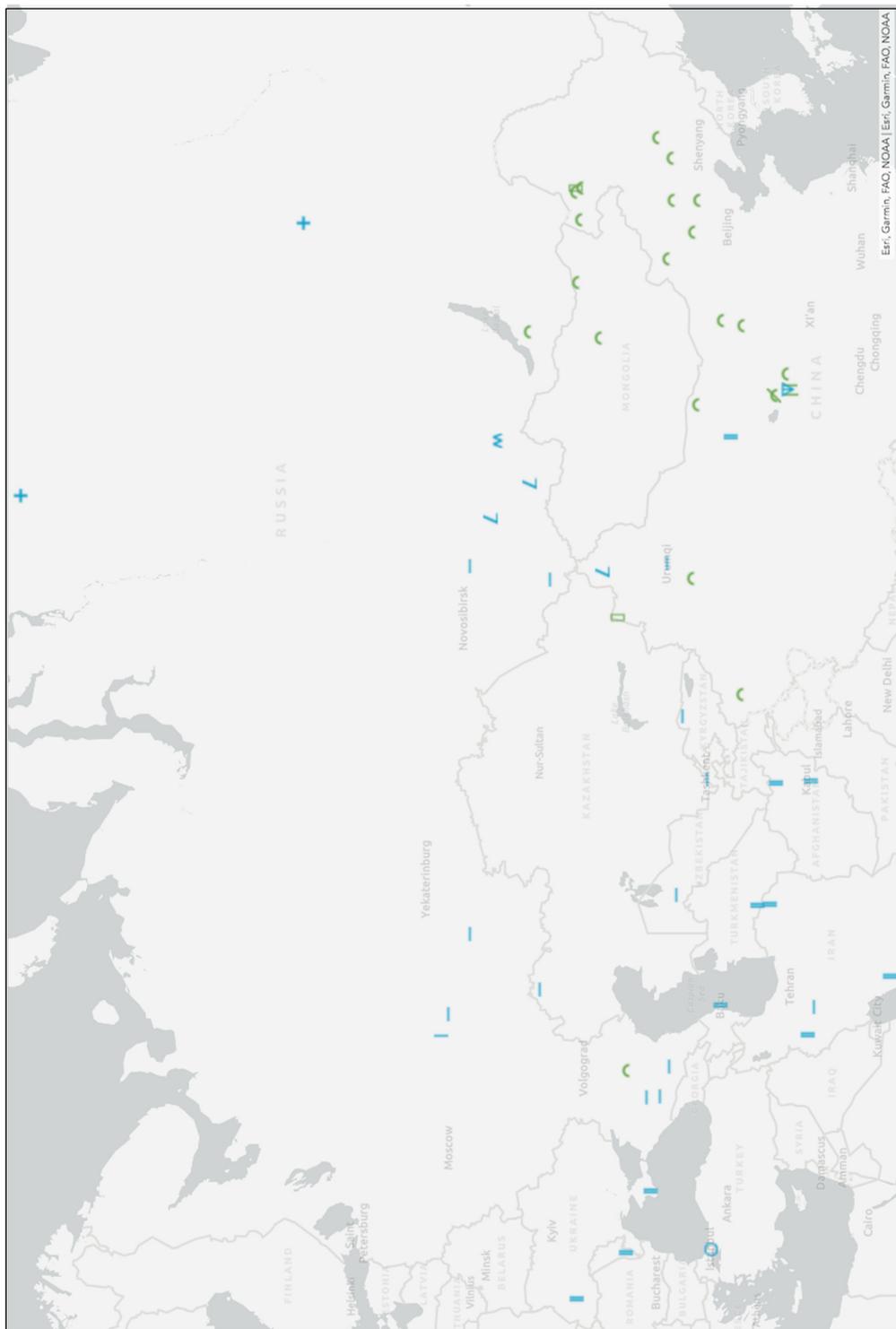
Figure 1.13.1: 'Mouse/rat' in Mongolic and Turkic.
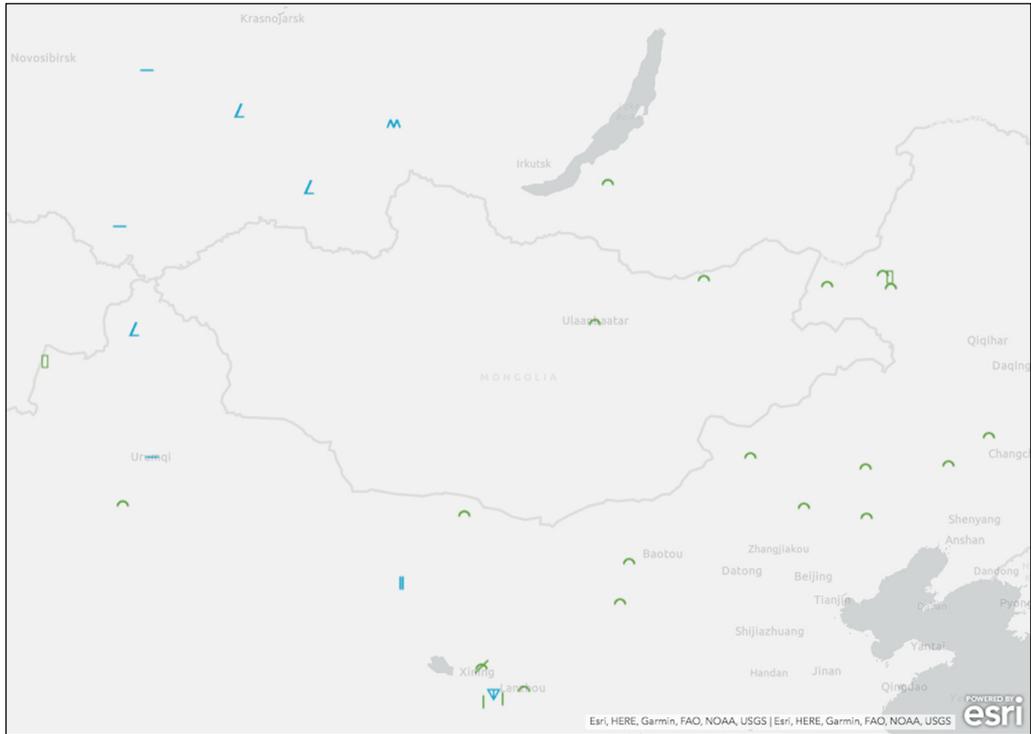
Figure 1.13.2: 'Mouse/rat' in Mongolic and Turkic (magnified).

'Mouse' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. When a language has several words for mise and rats, I targeted 'male adult mouse'.

The distribution of 'mouse' words is mainly the type B in Central India and the type A around it (from the islands to Europe). The type C is distributed mainly in the plains of Pakistan, D is found in the Andamanese languages, and E is for Burushaski.

The most major type is *mūṣa*. This type is derived from Sanskrit *mūṣa* मूष or *mūṣika* मूषिक 'mouse, rat' (-ka is used for diminutive or adjectiviser), derived from Proto-Indo-European *múh₂s* 'mouse'. So this type is cognate with Latin *mūs*, Ancient Greek *mûs* μῦς, and even English mouse. Forms of this type are used by IA and Vedda languages. Vedda has borrowed it from Sinhala. Historically, the /ṣ/ sound has been lost in many of the modern languages in South Asia, so it is often changed to /š/, /s/, /č/, or /z/ sounds. Kalderaš Romani says *kermúso* 'mouse' (while *soboláko* 'rat') and Zargari Romani says *görmiso*, they are composed by *ker*/*gör* and *múso*/*miso*. The latter morphemes *múso*/*miso* is derived from Sanskrit *mūṣa* whereas the former *ker*/*gör* is from Sanskrit *ghara* घर 'house' (In both lects, there is the independent word *kher* 'house').

The *undura* type appear in IA languages and Nihari. They are found in India, except in the north, and in Bangladesh. The origin of Sanskrit *undura* उंदर is etymologically unclear and may be borrowed from any lost substrate language in the Vedic era.

The *ćūha* type appears only in IA languages, which are biased towords larger languages, namely Hindi, Urdu, Panjabi, Saraiki, Sindhi, Gujarati, Nepali, Bengali, and Oriya. The word *ćūha* is not detected in Sanskrit but found in Ashokan Prakrit *ćūha* *𑀙𑀼 'mouse, rat'. So the distribution in the west is reasonable, and surely it has lately spread to Nepali, Bengali, and Oriya. It seems to be an onomatopoeia.

The *tode* type is used in Great Andamanese languages of the Andamanese family. Meanwhile in the Bea language, as of the same branch, in the South Island of Great Andaman uses a quite different form *rogɔ tɛtma* for 'mouse' and the Öñge language, of the Jarawa-Ongan branch, uses *ɛlɛŋe*.

The form *girkis* is used in all the dialects of Burushaki in common. The word can also refer to 'muscles (especially the biceps)' as well as 'mouse, rat', but it probably originally had the former meaning in its etymology.

Kashmiri and Kishtwari employ forms like *gaguṛ* (reconstruction mine) in common, but the etymon is unclear. In Assamese they employ the word *nigɔni* নিগনি 'mouse' which origin is unclear and I treated it so in the Map 1, while they has the word *endur* এন্দুৰ of the type B only for 'rat'. In the Khowar language, they call a mouse *xaláu*. This can be understand as derived from the Sanskrit word *khalapū* खलपू 'one who cleans the threshing floor' (< *khála* खल 'threshing floor' + *pu* पु 'cleaning'). Oriya also has the word *kundu* କୁନ୍ଦୁ 'mouse', which may be borrowed from the Dravidian family.

(YOSHIOKA Noboru)

A. *mūṣa* type (46) △

    mūṣ, musa, muṣo, mŭ̃ṣo, mŭš, mŭ̃šằ, mušä, mušo, mišo, mữ̃ẓu, mữ̃ẓo, mữ̃ẓi, mŭs, mŭ̃šằ, mūsu, musŏ̃, musə, musi, mữ̃sə, müsog, mūsrā, mizok, mūč, močaṭā, pusa

 [*ghara+*]

    kermuso, görmiso

 [*mūṣika*]

    mīdā, misijako, muski, mūẓi, mīyā, miiyaa

B. *undura* type (12) ○

undur, undar, undir, undru, ūndrā, ūndhar, ũdar, induro, indur, ĩ̆dur, honḍar

C. *ćūha* type (10) ▬

    čūhā, čūhī, čūho, čūō, čūā, čuvɔ, čuyā

D. *tode* type (3) ⌣

    to de, ṭoḍe, de

E. *gírkis* type (3) ▱

    gírkis

F. others

    [*gagur* type (2)] gagur, gəgiṛ; nigɔni, kɔṛeŋ, xaláu, kundu, kūo, rogɔ tɛtma, ɛlɛŋe, yaŋuat
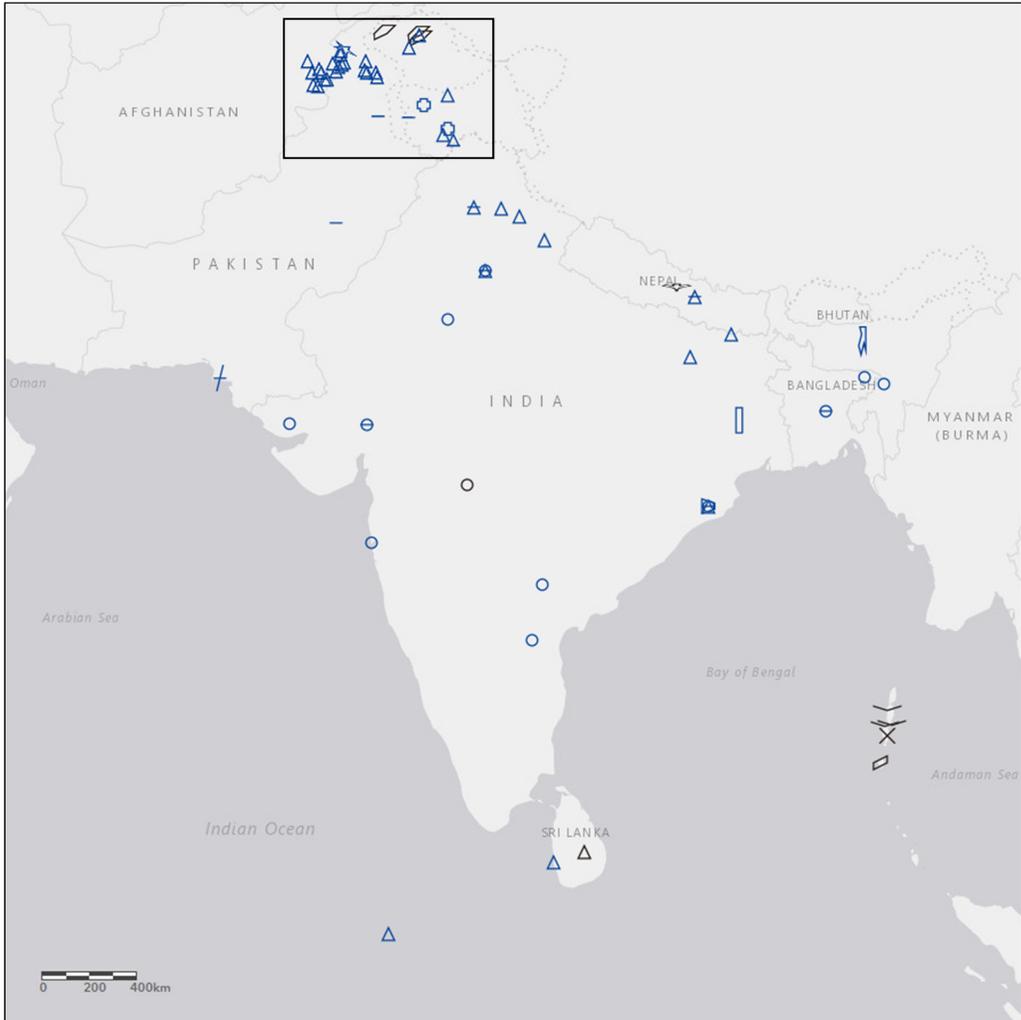
Figure 1.14.1: 'Mouse' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).
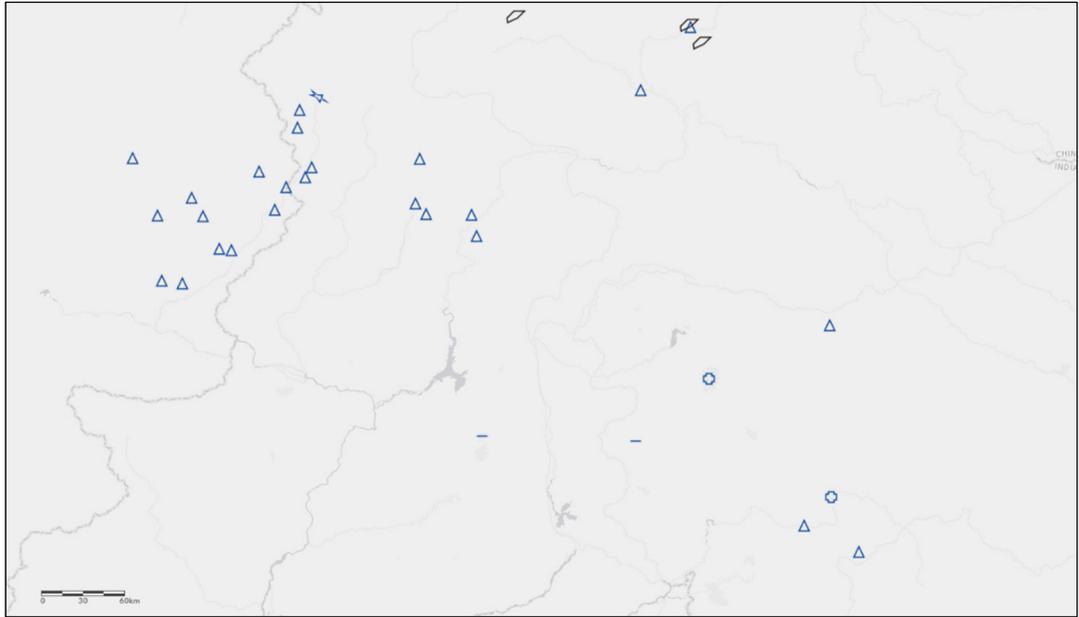
Figure 1.14.2: 'Mouse' in northern Pakistan (the area encloed by the rectangle in Figure 1.14.1).
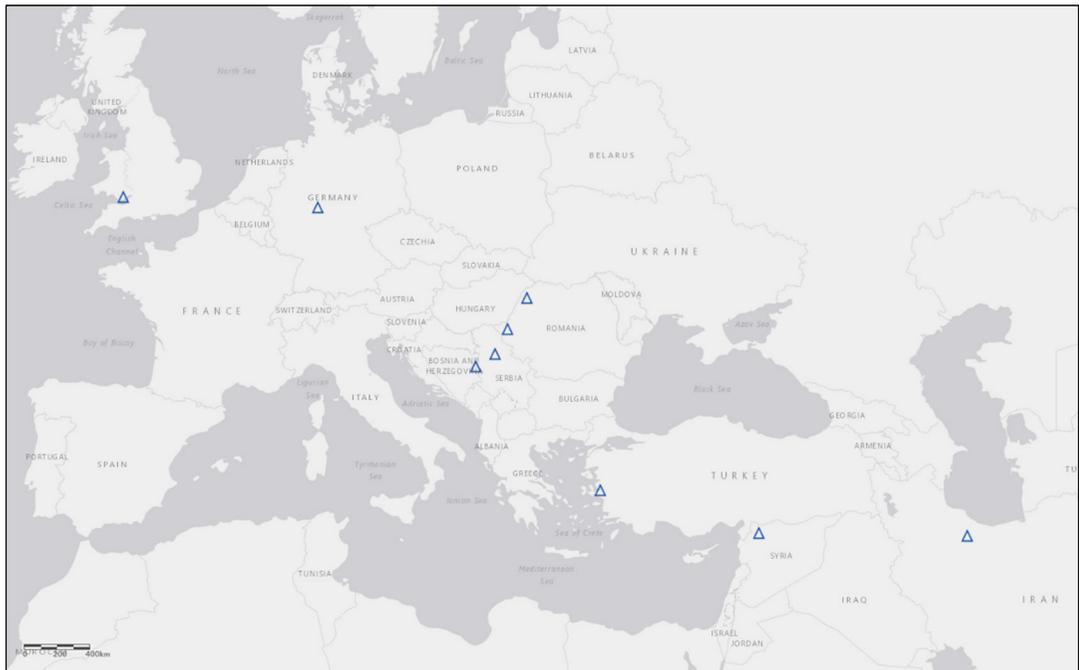
Figure 1.14.3. Types for 'Mouse' in Indo-Aryan languages outside South Asia.

'Rat' in Dravidian

The Dravidian etyma for 'a rat' are a common term for genera *Rattus* (rat), *Mus* (mouse), *Bandicota* (bandicoot rat), *Suncus* (musk shrew). *A Dravidian etymological dictionary* (Burrow & Emeneau 1984; DEDR) identifies four etyma, which may be ultimately cognate although evidence for this possibility is inconclusive.

The most widespread etymon is ELI (DEDR #833). Its reflexes in South Dravidian are *eli*, *ili*, *isy*, *eyj*. Suffixed forms such as *elka* and *eluka* as well as *el*, *elli*, *alli* etc. are found in South Central and Central Dravidian. Brahui *hal* is possibly one of the few retained Dravidian animal terms in this geographically isolated language.

The other three etyma are innovative forms exclusive to a genealogical clade. ORLI (DEDR #994) are found only in the Kui-Kuvi-Pengo-Manda clade of South Central Dravidian. Kurukh & Malto clade shares reflexes of OSGA (DEDR #941) and ERGO (DEDR #673) for 'field mouse'.

(KODAMA Nozomi)

∕ eli ＼ ili ━ isy, eyj

∨ eluka, elka ‹ elli. allī › el

▮ hal

△ orli,urli,oḍri,ori'i

◇ osge/eɽge, osgā/eɽgo:
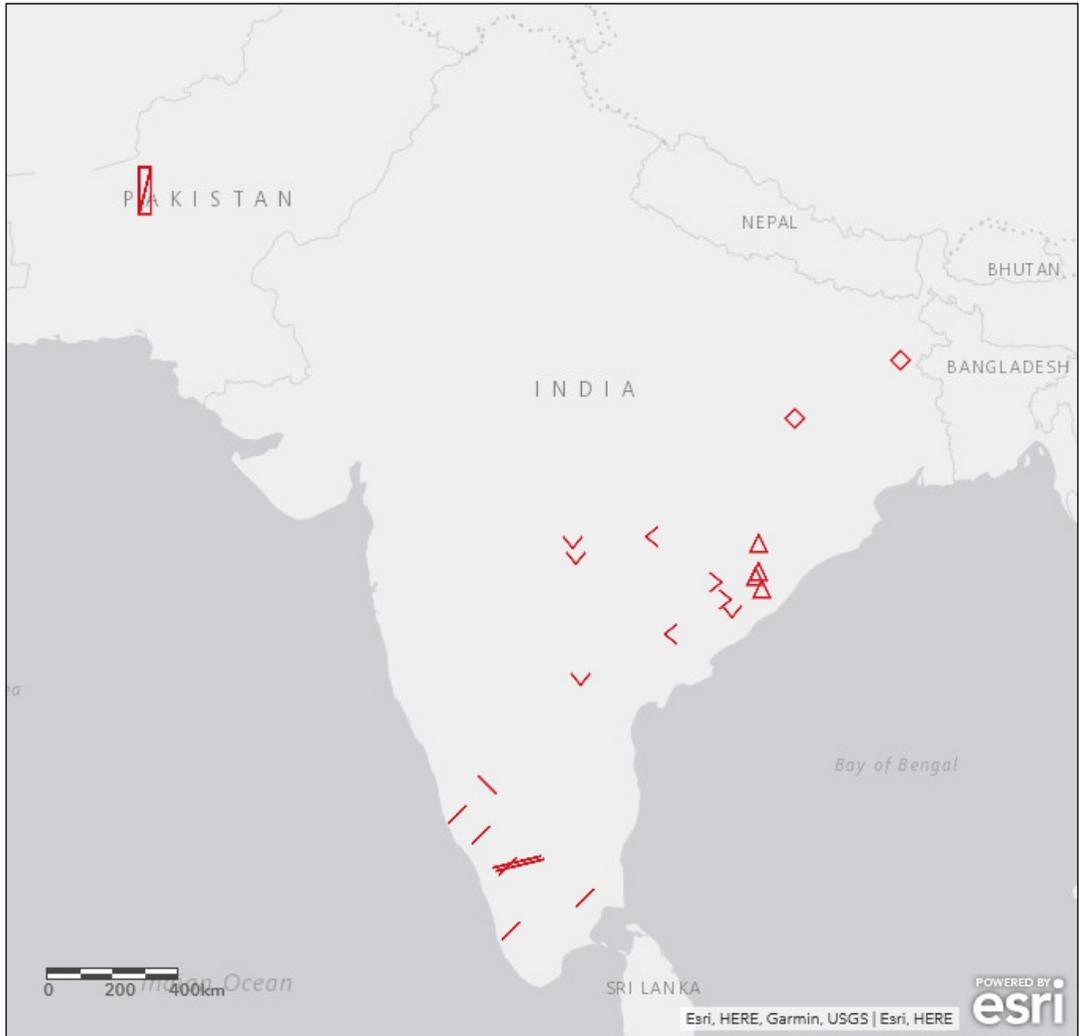
Figure 1.15.1: ‘Rat’ in Dravidian.

'Mouse' in Iranian

As seen in Figure 1, Type A mūš, which is the most common type in the Iranian languages, is distributed in the vast area except for languages in the Pamir-Hindukush region, Ormuri in Afghanistan and Pakistan, Parachi in Afghanistan, and Mazandarani in Iran. It ultimately traces back to the Proto-Indo-European (PIE) word *mūs- 'mouse'. Some Northwestern Iranian words of this type show rhotacism (ex. Zazaki *merre* 'mouse' and maybe further r/l alternation, ex. Gorani *mıla*).

The distribution of Type B is limited to the Pamir-Hindukush area. All Pamir languages have this type. It derived from the Proto-Iranian (PIr.) *paurka- 'grey one

(diminutive)', which replaced the original Iranian word for 'mouse' probably owing to avoidance of a taboo word.

Languages with Type C word are scattered in some peripheral areas, namely the Caspian coast and the eastern Afghanistan and western Pakistan.

Forms extended with suffix (-ak- < PIr. *-aka-, mainly adds diminutive) are frequently observed in many languages in Type A through C.

Type D *baluṛ* is unique to Parachi, whose origin remains unknown. It might be a borrowing from a non-Iranian language.

(IWASAKI Takamasa)
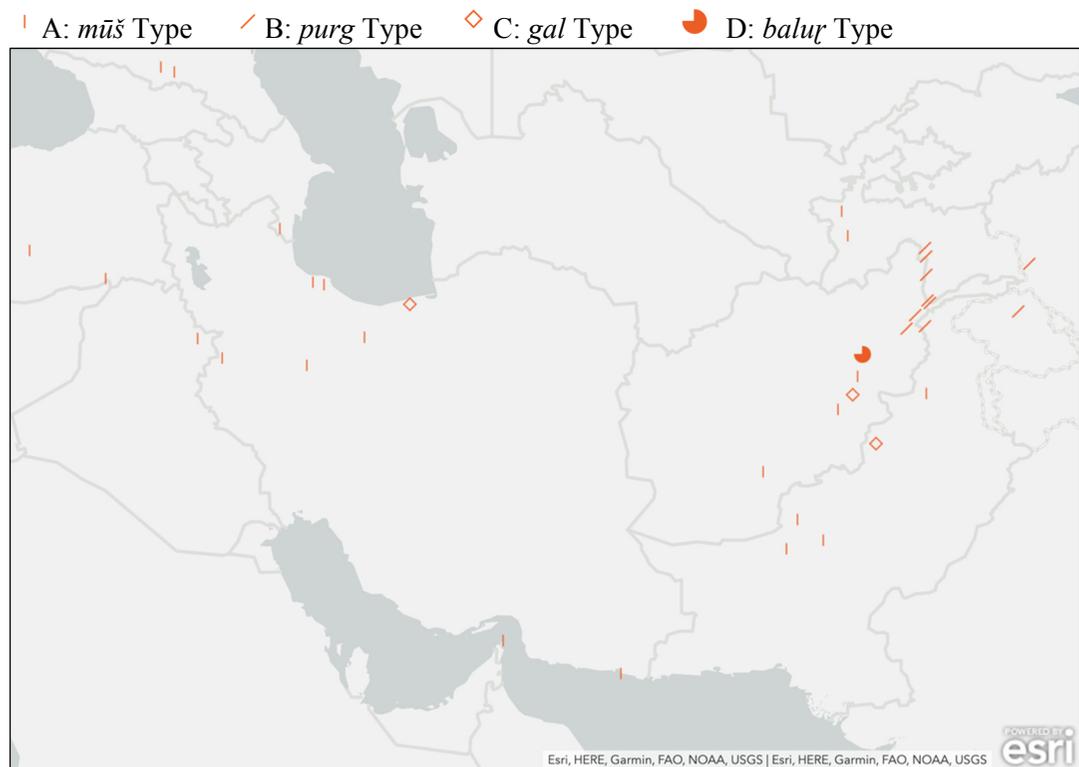
Figure 1.16.1: 'Mouse' in Iranian

'Rat' in Caucasian languages

The data provided by Klimov & Khalilov (2003:232) show that some target languages differentiate 'rat' from 'mouse', while some do not. Fewer word forms are recorded for 'rat' than for 'mouse'. In addition, we find a lexically large variation among the word forms for 'rat'.

The distribution and morphology of each type are as follows:

Type A: in Kartvelian languages, as a compound of 'donkey' and 'mouse'; Type B: in Abkhaz, as a compound of 'mouse' and 'big'; Type C: in Adyghe; Type D: in Kabardian, as a compound of 'mouse' and 'big'; Type E: in Chechen and Khvarshi; Type F: in Ingush; Type G: in Andi; Type H: in northern Daghestanian languages; Type I: in southern Daghestanian languages, and Type J, miscellaneous: in Dargwa and Udi.

The forms in Andi, Karata, Bagvalal, Dargwa, Archi and Rutul are common to those for 'mouse'.

(SUZUKI Hiroyuki)

- A: v-type; *virtxa*, *vir-i*, *wiršduɡw*.
- B: *a-hʷɨnapdɨw*.
- C: *šʷaje*.
- D: *dzɪʁʷəšxʷə*.
- E: *mukadaχka*.
- F: *cicxolɡ*.
- G: *hink'k'u*.
- H: χ-type; *χorχodobo*, *ãχʷa-hek'oča*, etc.
- I: q'-type; *čačan-ãq'o*, *noq'q'on*, etc.
- J: others; *waca*, *daptapaj*, *k'eč'namel*.

Figure 1.17.1: 'Rat' in Caucasian languages.

'Mouse' in Semitic

A. *fa:r* type (●) is Arabic form. *faʔr* (Yemen) is the same as the Classical Arabic. In Egypt /fɑ:r/ with an emphatic /ɑ/. In Morocco *faṛ* with an emphatic *ṛ*. *lufár* (Nubi in Kenya) is a form combinated with the definite article **l-*.

B. *umm-si:si* type (○ *umm* 'mother' of *si:si:* 'young rat') is found in Arabic in the Sudan belt: *umm-si:si* (Sudanese Arabic), *amsi:si:* (Chadian Arabic).

C. *ʕ-k-b-r* type is distributed in the Syria region. *axbar* (▲ Hebrew עַכְבָּר) is taken over from Biblical Hebrew *ʕakbar* and is the cognate of Pheonician *ʕ-k-b-r* ⵔⴻ𐤓𐤁𐤊𐤏 (cf. Akkadian *akbaru(m)* 'jerboa'.) *ʕaqubra* (▼ Aramaic of Koy Sanjaq) is related to Classical Syriac *ʕuqbro:* (ܥܘܩܒܪܐ) and probably related to *ʕ-k-b-r*. Behnsted & Woidich (2011: 385) reports that *ʕakba:ri:* is found in Arabic in San'a of Yemen. The Syriac form *ʕuqbro* may be related to *ʕ-q-r-b* 'scorpion' (Syriac *ʕeqarbo:* ܥܩܪܒܐ, Ugaritic *ʕ-q-r-b* ◀⊣⊳ 𐎁).

D. *ʔ-r-q-j-b* type is South Arabian form: *ʔa:rqajb* (▽ Mehri), *ʔark'eéb* (△ Hobyot). This type shares the *r, q, b* consonants with *ʕuqbro:* (Syriac).

E. *antʃ'iwa* (☆ Tigrinya ኣንጭዋ) and *ʕans'a:j* (★ Tigre) of North Ethiopic are probably related. *a:jit'i* (✳ Amharic አይጥ) may be related to these.

F. *mu:ʃ* (✚ Bukhari Arabic in Uzbekistan) is borrowed from Iranian (cf. Tajik *muʃ* муш, Persian *mu:ʃ* موش).

None of these modern forms seem to be related to Akkadian forms *pe:ru:ru:tu(m)* 'mouse', *χumṣi:ru* '(large) mouse' *puʃχu:* 'rat'.

(NAGATO Youichi)

A. *fa:r* type
- ● *faʔr, fa:r, faṛ, lufár*

B. *umm-si:si* type
- ○ *umm-si:si, amsi:si:*

C. *ʕ-k-b-r* type
- ▲ *axbar*
- ▼ *ʕaqubra*

D. *ʔ-r-q-j-b* type
- ▽ *ʔa:rqajb*
- △ *ʔark'eéb*

E. *antʃ'iwa* type
- ☆ *antʃ'iwa*
- ★ *ʕans'a:j*
- ✳ *a:jit'i*

F. Other
- ✚ *mu:ʃ*

Figure 1.18.1: 'Mouse' in Semitic.

'Mouse' in Nilo-Saharan

What follow are heuristically reconstructed (marked with a hash #) Nilo-Saharan roots for 'horse' (93 languages surveyed). We simplify diacritics and notations for the [±ATR] feature in the original data. For languages that do not (seem to) distinguish 'mouse' vs. 'rat', we use the term for 'rat'.

No Nilo-Saharan roots for 'mouse/rat' seems safely reconstructable, but there are three possible cross-branch roots.

Type A #*kilt*, which is attested in Taman (Assangori *kinit*, Misiirii *kurut*), Surmic (Majang *kilt*), Shabo (*kilta*), Kadu (Krongo *kili*) and less possibly Nubian (Kenuzi *skitte*, Midob *ekkendi*) and Saharan (Kanuri *jilwa*, Tudaga *kuur*) branches. The fact that Shabo and Majang share the root may be due to later contact.

Type B #*s(')eg(b)e* has three variants, B1 #*segbe* attested in Central Sudanic Sara-Bongo-Bagirmi ('Beli *hegbe*, Baka *sige*, Kara *se'b*; cf. **S-Rgb-* by Boyeldieu 2000a), B2 #*yegə* attested in the Sara sub-branch of the same branch, and B3 #*s'igi* attested in Koman (Komo *s'ik*, Opo *cigi*).

Type C #*der* (cf. **dér/dēr* 'rat' by Ehret 2001) is found in Eastern Nilotic (Maa *en-deroni*), Ik (*der*) and less possibly Central Sudanic Moru-Ma'di (Ma'di *idre*, Lugbara *edroo*). The inclusion of Central Sudanic in this type, however, is dubious because /dr/ in Moru-Ma'di is a phoneme rather than a consonant cluster. Ehret (2001) considers the Eastern Nilotic reflexes as borrowings from Kuliak.

There are a few branch-unique roots, such as Type D #*ŋiŋse* for Daju (Dar Daju *ñiŋse*, Darfur Daju *ŋəŋse*, Shatt *nyingas*) and Type E #*oyio* for Western Nilotic Lwo (Luwo *aywiio*, Shilluk *yiejo*, Päri *yio*, Acoli *oyoo*). Other roots categorized as Type F are attested by only one or two languages in the list.

(NAKAO Shuichiro)
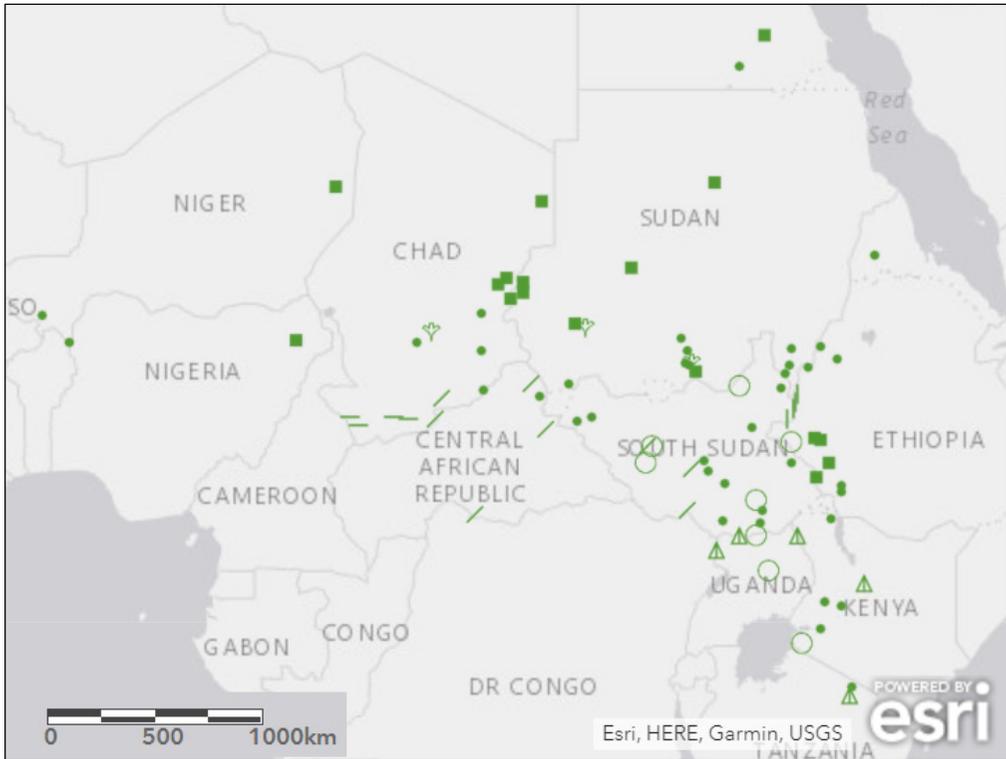
Figure 1.19.1: 'Mouse' in Nilo-Saharan.

- ■ A. *#kilt*
  B. *#s'egbe*
- ╱ B1. *#segbe*
- ─ B2. *#yegə*
- | B3. *#s'igi*
- △ C. *#der*
- ☥ D. *#ŋiŋse* (Daju)
- ○ E. *#oyio* (Lwo)
- ● F. The other types
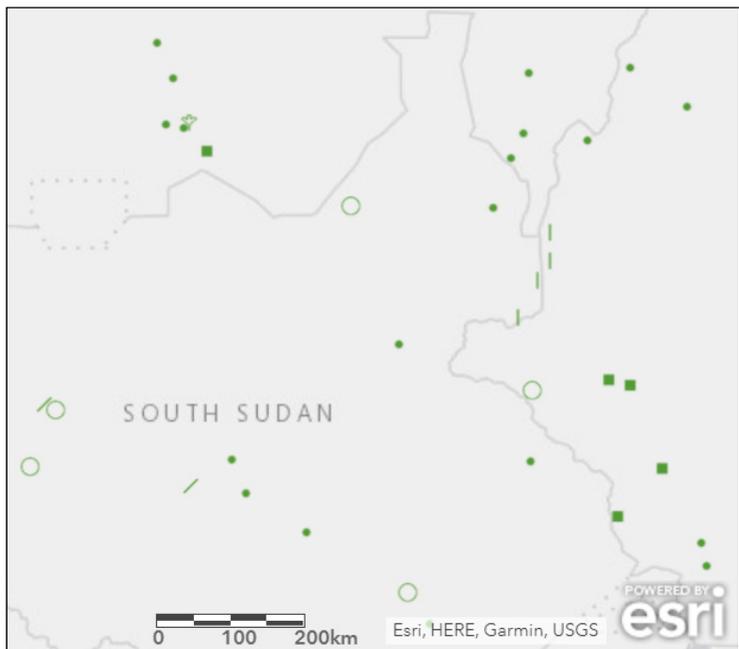
Figure 1.19.2: 'Mouse' in Nilo-Saharan around South Sudan.

'Rat' in Bantu

In the Proto-Bantu lexicon (Meeussen 1969, Bastin *et al.* 2002), four distinct forms are reconstructed as a nominal stem denoting 'rat' or 'mouse', namely i) *-*bèbà*, which has a variant form *-*bìbà*, ii) *-*béndé*, iii) *-*kòcòè*, and iv) *-*pókò*. While they show different geographical distribution patterns, with more or less substantial overlaps with each other, at least some of them can be regarded as denoting different species.

*-*bèbà* [BLR-MAIN-117] (as registered in Bastin *et al.* (2002)), along with its variant *-*bìbà* [BLR-VAR-168], is broadly distributed in the Eastern and Southern zones including D, E, F, G, J, K, L, M, N, and S (for the Bantu zones, see Guthrie (1967–71) and Hammarström (2019) for the latest version). Its salient feature in terms of geographical distribution is that this term highly tends to overlap with other forms that are also traced back to Proto-Bantu, e.g., the combination of *-*bèbà* +*-*béndé* is attested in Manyika [S13a] (*mbewa*; *mbende*); the combination of *-*bèbà* + *-*kòcòè* is attested in Kami [G36] (*mbewa*; *ngoso*); the combination of *-*bèbà* + *-*pókò* is attested in Kimbu [F24] (*imbeva*; *mpuku*), etc. Such distribution patterns might suggest that this term tends to be used as a generic term.

*-*béndé* [BLR-MAIN-149] shows almost complementary distribution with *-*bèbà,* i.e., its descendant forms are mainly distributed in the North-Western zones including zones A, B, C, and H. While this term is also observed in combination with other forms, especially with *-*pókò* as in Bulu-Bene [A74] (*mbiene*; *mpuku*) and Bobangi [C32] (*mende*; *kapuku*), Guthrie

(1970: 37) estimates that this term may have denoted 'a particular striped or spotted species' at the PB stage.

*-*kòcòè* [BLR-MAIN-1873] is broadly distributed in the Central Savannah area spreading over zones B, F, G, L, M, P and S, e.g., in Bondei [G24] *ngoswe*; Luba-Katanga [L33] *ŋkoswe*, and Namwanga [M22] *ekuza.* As for its meaning, Guthrie (1970: 289) suggests that the term may have referred to a 'house-rat'.

*-*pókò* [BLR-MAIN-2642] shows the widest distribution among the four etymons in PB, spreading across all zones except for interlacustrine Zone J. Based on the distribution, Guthrie (1970: 45) estimates that this term may have denoted either 'rat in general' or its 'commonest species.'

Besides the numerous forms that can be relatable to these reconstructed etymons, considerable varieties of forms for (various species of) 'rat' are also observed. For example, <*kuli*> and its relatable forms are distributed across South-Eastern zones e.g., in Pogolo [G51] *likweri*; Matengo [N13] *likuli*; and Makonde [P23] *nkule*. There are also various forms that are shared only within a group of immediate genetic branching, e.g., <*kwende*> in languages spoken in the corridor between Lake Malawi and Lake Rukwa, <*ndugi*> in Logoori-Kuria group [JE40] spoken in the eastern shore of the Lake Victoria, <*kindu*> in the Central Kenyan languages [E50], and <*koikoi*>, which is exclusively shared within the Chaga languages spoken in the slope area of Mt. Kilimanjaro [E60].
(SHINAGAWA Daisuke and KOMORI Junko)

Forms traced back to reconstructed PB forms

- ⬡ *-bèbà   [BLR-MAIN-117]
- ▽ *-pókò   [BLR-MAIN-2642]
- ⌃ *-kòcòè   [BLR-MAIN-1873]
- ═ *-béndé   [BLR-MAIN-149]

Other common forms

- ‖ <kenge>
- ⌄ <kindu>
- ╱ <koikoi>
- ⌒ <kuli>
- ╲ <kwende>
- ⌣ <ndugi>
- Ψ <mbiling
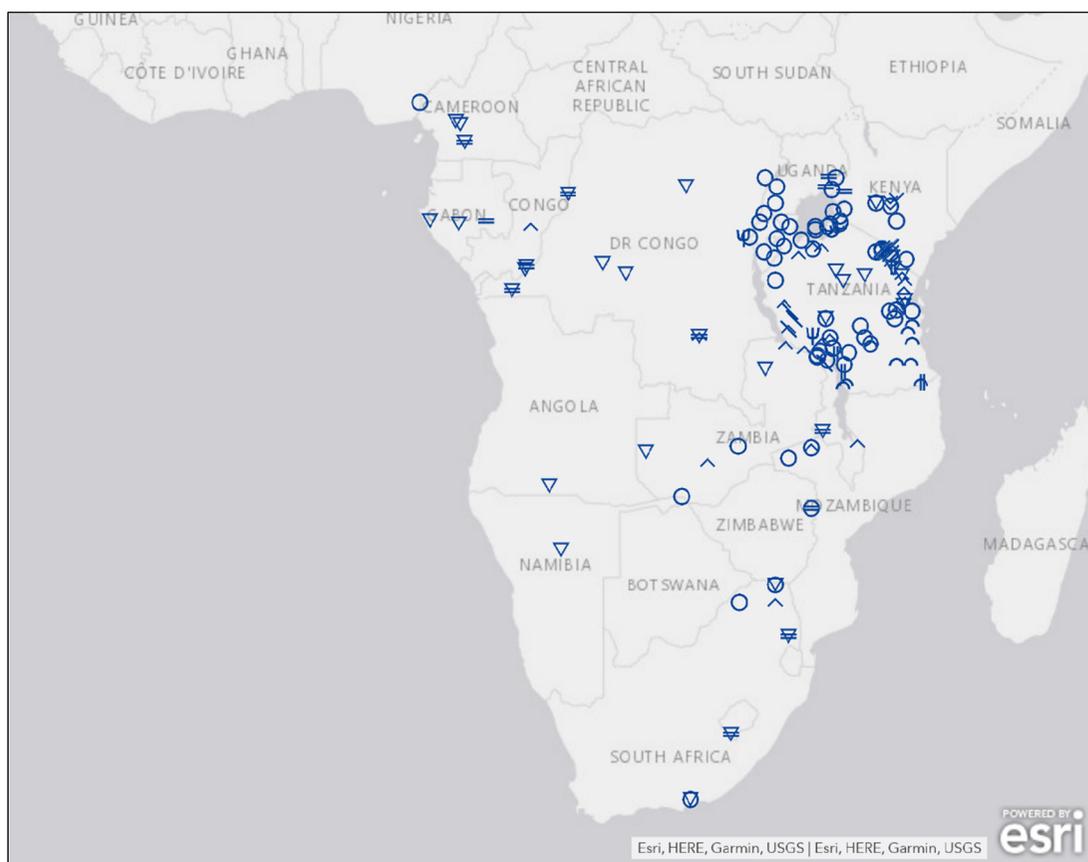
Figure 1.20.1: 'Rat' in Bantu.
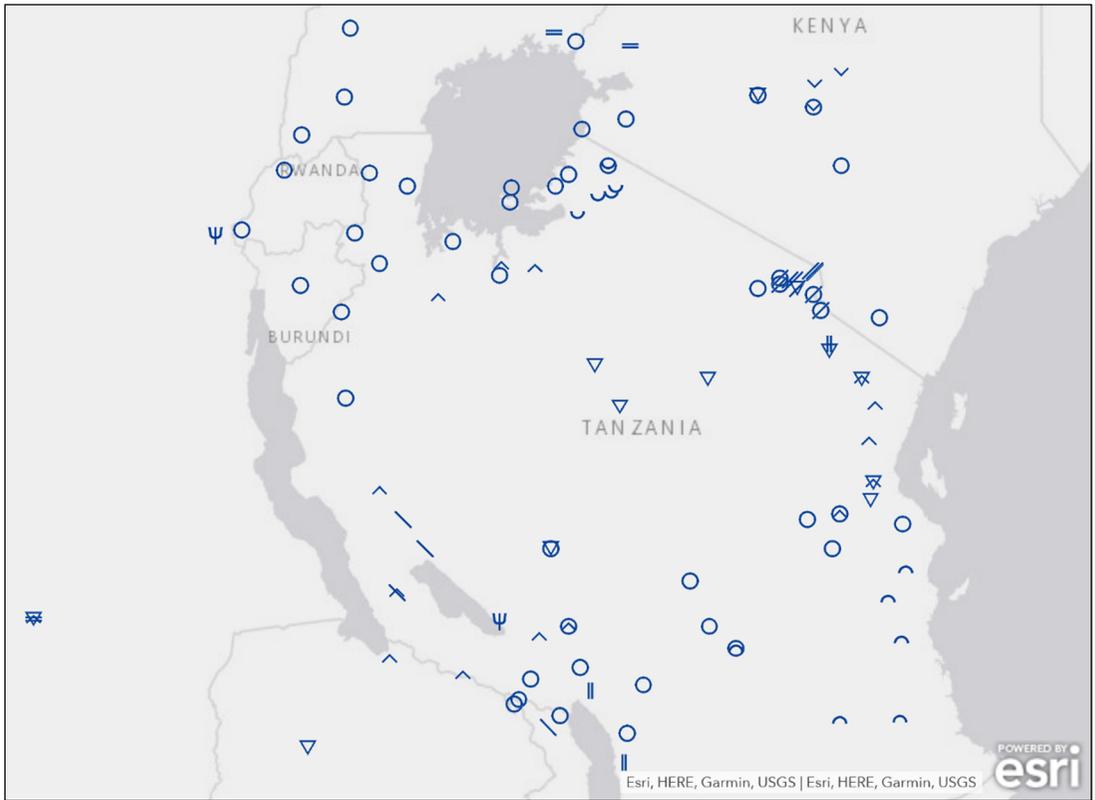
Figure 1.20.2: 'Rat' in northeastern Bantu zones.

'Mouse' in the Kalahari Basin area

As summarized in Table 1, seven types of the word "mouse" appear in 15 sample languages in the Kalahari Basin area (KBA). Word forms included in each word type share the same etymological origin, thus, *si-ŋǂuˤe* (B1) and *ŋǂuˤje* (B2) are cognates integrated into Type B.

All seven types are distributed within one of the three language families, Tuu, Kx'a, and Khoe-Kwadi. Each etymological word type is not distributed across different language families; that is, word forms in Types A and B are observed only within the Tuu family, Type C and D in the Kx'a family, and Types E, F, and G in the Khoe-Kwadi family.

Table 1: Geographical variation of "mouse".

	Tuu	Kx'a	Khoe-Kwadi
A	A: *tiɾi*		
B	B1: *si-ŋǂuˤe* B2: *ŋǂuˤje*		
C		C: *ŋǁuũ*	
D		D1: *gǀhui* D2: *ǀhṹì*	
E			E1: *ŋǂuni* E2: *ɲúní* E3: *ŋǂuˤni*
F			F: *gǁàu*
G			G: *tùrú*

(KIMURA Kimihiko, NAKAGAWA Hirosi)
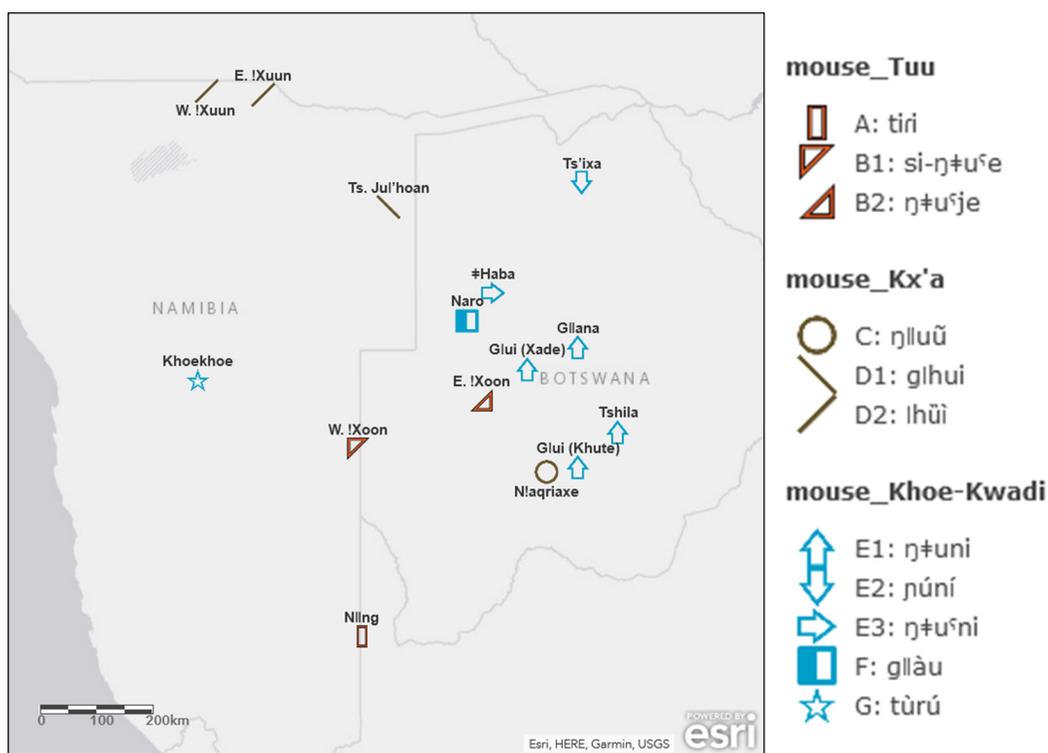
Figure 1.21.1: Geographical variations of "mouse" in KBA.

Chapter II

Chicken

'Chicken' in Asian and African languages

The terms for 'chicken' tend to begin with the sound *k-*, *t-*, *d-*, *tʃ-*, *dʒ-*, *m-*, *s-*, or *h-*. Words beginning with *k-* are widespread in Asia and Africa. However, since the *k-* words often originate from the onomatopoeia of chicken clucking, it could have occurred separately in different regions, not only due to language contact or borrowing. For example, Hmong-Mien *kəi* and Kra-Dai *kai* are (or are thought to be) borrowed from Sinitic, whereas Old Japanese *kake* is not assumed to have the same origin as Sinitic.

The words beginning with the *t-*, *d-*, *tʃ-*, and *dʒ-* consonants appear prominently in Mongolian, Turkish, Korean, and Semitic languages, and a band-like spread is visible on the Asian and African maps. Nevertheless, the etymological relationship between the words in each language is unclear.

Chicken domestication occurred about 1,000 years ago in India and Southeast Asia, respectively (Yonezawa and Sasaki 2016). Notably, the words for 'chicken' beginning with *m-* are common in Austronesian, Austroasiatic, and Proto-Indo-Iranian languages in the regions where domestication occurred early.

The words for 'fowl' (cf. Turkic *takā*, Semitic *firaːx*) and 'bird' (cf. Ainu *cikáp*, Proto-Tibeto-Burman **daw*) are often found in the original sense of 'chicken.'

In some cases, a word for 'chicken' may not be distinguished from that for 'hen' (cf. Nilo-Saharan), and also mean 'hen' in a narrow sense (cf. Mongolic). When distinguishing between 'roosters' and 'hen,'

a gender-indicating word (cf. Turkic) or the sound markers for gender may be used (cf. Sanskrit).

Table 1: Main word forms for 'chiken'.

Languages	Word forms
Ainu	*niwátori* type (< Japonic) *cikáp* < 'bird'
Japonic (J)	*tori* type < 'bird' *niwatori* type < 'garden bird' *kokeko* type: OJ *kake* [*op] *hoohoo* type [*op]
Korean (K)	*tʌk, tak* (< MK tʌrk)
Sinitic (Sn); Chinese (C)	鸡 type: *tɕi, ki, kue* (< MC *kiei* < OC *kie*) 鸡子 type: *tɕi tsɿ, kɛi tɕi* 鸟儿 type: *tɕir, tɕi ər*
Hmong-Mien	*kəi* type (< probably Sn)
Kra-Dai (KD)	*kai* type (< PKD < Sn)
Tibeto-Burman (TB)	PTB **b(y/r)a* type [*op] 'bird or bee' PTB **k-rak* type 'fowl or chicken' [*op] PTB **haːr* type 'bird, fowl, or chicken' PTB *w(a/u)* type 'bird, egg, wing, or fowl' WrT *de* type Proto-Karenic: **chjaNᴬ* type PTB **daw* type 'bird'
Austroasiatic	Proto-Khmuic etc.: **(s)ʔiar* type Proto-Katuic: **ʔndruuj* type Proto-Pearic: **hlɛːk* type Proto-Vietic: **r-kaː* type Munda (Proto-Kherwarian): **sim* type *chaaŋ* type *manuk* type *moan* type *ciəj⁶* type
Austronesian	*sīyop* type *(m)anuk* type *ayam* type *moa*

	Onomatopoeia type: *koka, kuʈuʔ, tsiotsio, kirek* [*op]		*h-* type *d-* type *kp-* type
Tungusic	*kakara*, *xaxara* [*op] *čoko* type *nai*	Semitic	*dadʒa:dʒ* type *dʒida:d* type *derho* type *farru:ʒ* < 'chick' *fira:x* < 'fowl' *ʈe:r* < 'bird'
Uralic	*kana* *saraz* *čywe* type *kureg* type	Nilo-Saharan (NS)	#*kokor*/#*koko* type [*op] #*kanda* type (< PNS **k-nd-*) #*kunza* type (< PNS **k/ng-Rnj-*) #*dirbad* type #*(n)gweno* type (< PNS **ng-n-*)
Mongolic (Mg)	*takā* type (< Turkic 'fowl')	Bantu (B)	PB **kókó* type PB **kókó* type (variant: **kókó*) PB **kúba* type
Turkic (Tk)	'Hen' *tavuk* type (< OTk *takïgu*) *taqqïnjaq* *takā* type (< Mg *takā* < Tk 'fowl') *mekijan* type *kuš* (< 'bird') *anəš* 'Rooster' *xorus* type (< Persian) *ätäč* type *tanaq* *avtăn* (< *avt-an* 'singing') *guŋgu* (cf. C *gong* 公 'male')	Kalahari Basin area	*hunder* type *kooko* type (< B *koko*) [*op] *kʰúú-kʰúú* [*op] *hūkú* *goro* *ʔánɨ́*

Abbreviations are as follows:

*op: onomatopoeia,

#: (heuristically/author's) reconstruction.

(FUKAZAWA Mika)

South Asia	Sanskrit **kukkuʈá/í* type 'rooster/hen' [*op] *mr̥gás* type (< Proto-Iranian < Proto-Indo-Iranian (PIIr)) #*khaini* type [*op?] *qarqámuc* type [*op?] Sanskrit *ćaʈaka* type < 'sparrow' [*op] #*kombḍa* type [*op?]
Dravidian	'Fowl' *kor* type *kozi* type *qéru, xēr* (cf. Tamil *kēru* 'to crackle (as a hen)')
Iranian (Ir)	*kerk* type (< PIr **kr̥ka-*) [*op] *murɣ* type (< PIr **mr̥gá-* 'bird' < PIIr '(wild) game animals') *tuxi* type *čūrī* type
Caucasian	'Rooster' *m-* type < 'male' *r-* type *t-* type

'Chicken' in Ainu

The term for 'chicken' (*Gallus gallus domesticus* Brisson) in Ainu is divided into three types. Type A is a borrowed type from *niwatori* にわとり or 'chicken' in Japanese, and Type B is an original word in Ainu that means 'bird.' The forms in Type C are classified into a mixed type, which consists of the Japanese *niwatori* and Ainu *cikap.* The term of C-3, *cisecikah*, seems to be a calque for the Japanese 'niwatori': *cise* 'house' and *cikah* 'bird' in Sakhalin dialect. Note that the Sakhalin dialect /-h/ [x] is a special phonemic variant, that is substituted for the coda /-p, -t, -k, -r/ [p', t', k', ɾ] in Hokkaido dialects (cf. Chiri 1973 [1942]: 471–472; Tamura 2000: 20).

(FUKAZAWA Mika)

A. *niwátori* type
- 🔴 A-1. *niwátori ~ niwatori*
- ⭕ A-2. *niyátori*

B. *cikáp*
- ◇ *cikáp*

C. Mixed type
- ◈ B-1. *(niwátori) cikap*
- | B-2. *niwácikap*
- Ψ B-3. *cisecikah*

Figure 2.2.1: 'Chicken' in Ainu.

'Chicken' in Japonic

For chicken, the most common type is TORI (*tori*, *tui*, *tuɭ*, …), followed by NIWATORI (*niwatori*, *niwadori*, *niwattori*, *niyattori*, *newattore*, *niwatoi*, *niyatoi*, *nyattorime*, *myaatui*, *meeduru*, …) and KOKEKO (*kokkako*, *kakero*, *goka*, *gugu*, …) types.

TORI is originally a word that refers to birds in general, but it seems to be an expression based on synecdoche, which has come to be called simply TORI even for chickens that are the most familiar livestock birds. NIWATORI means "garden bird" and refers to the bird you keep. KOKEKO is derived from the scream, and the oldest attested form for chicken in Japonic is also *kake*, which seems to be the proto-Japonic form for chicken. However, the KOKEKO type found in various dialects seems to be derived in parallel from the scream rather than the retention of the proto Japonic, since we cannot find any sound correspondence with each form. The HOOHOO type (*hoohoo, hooho*, *hoho*, *ho*, *pappa*) may also derived from the scream of the chicken or other birds. NIWATORI may also have been made in parallel from *niwa* 'garden' and *tori* 'bird', but as a compound word consisting of *niwa* and *tori*, both forms of *niwatori* and *niwadori* are possible. However, almost all dialectal forms correspond to *niwatori*, not *niwadori* (if the proto form were *niwadori, the modern form in Tohoku dialects would be ×*niwãdori*), suggesting that these forms were inherited from the old noun phrase of *nipa tu tori* 'bird in the garden' (> *niwattori* > *niwatori*). The NIWATORI type in Ryukyuan languages such as Kuroshima *meeduru* do not directly correspond to *nipa tu tori*, suggesting that these forms are calques of Japanese. Although the etymology of Yonaguni *mita* is uncertain, we presume that this word came from *miya-tori-a (> *meetorya > *miituya > *mita*) "garden bird" (-*a* is a diminutive). In Yonaguni, *hatu*, which refers to pigeons/doves, also refers to birds in general. Similarly, there is a possibility that TORI originally referred to chicken and later expanded to birds in general. However, this is precluded by the fact that *kake* for chicken existed in Old Japanese and *tori* made various compounds such as *atori* 'finch' and *tidori* 'plover'.

Since *kake* is thought to be a word derived from the scream of the chicken, comparison with other languages may not make much sense.

(NAKAZAWA Kohei and YOKOYAMA Akiko)

ꞁ TORI
– NIWATORI
⌒ HOOHOO
⌃ KOKEKO
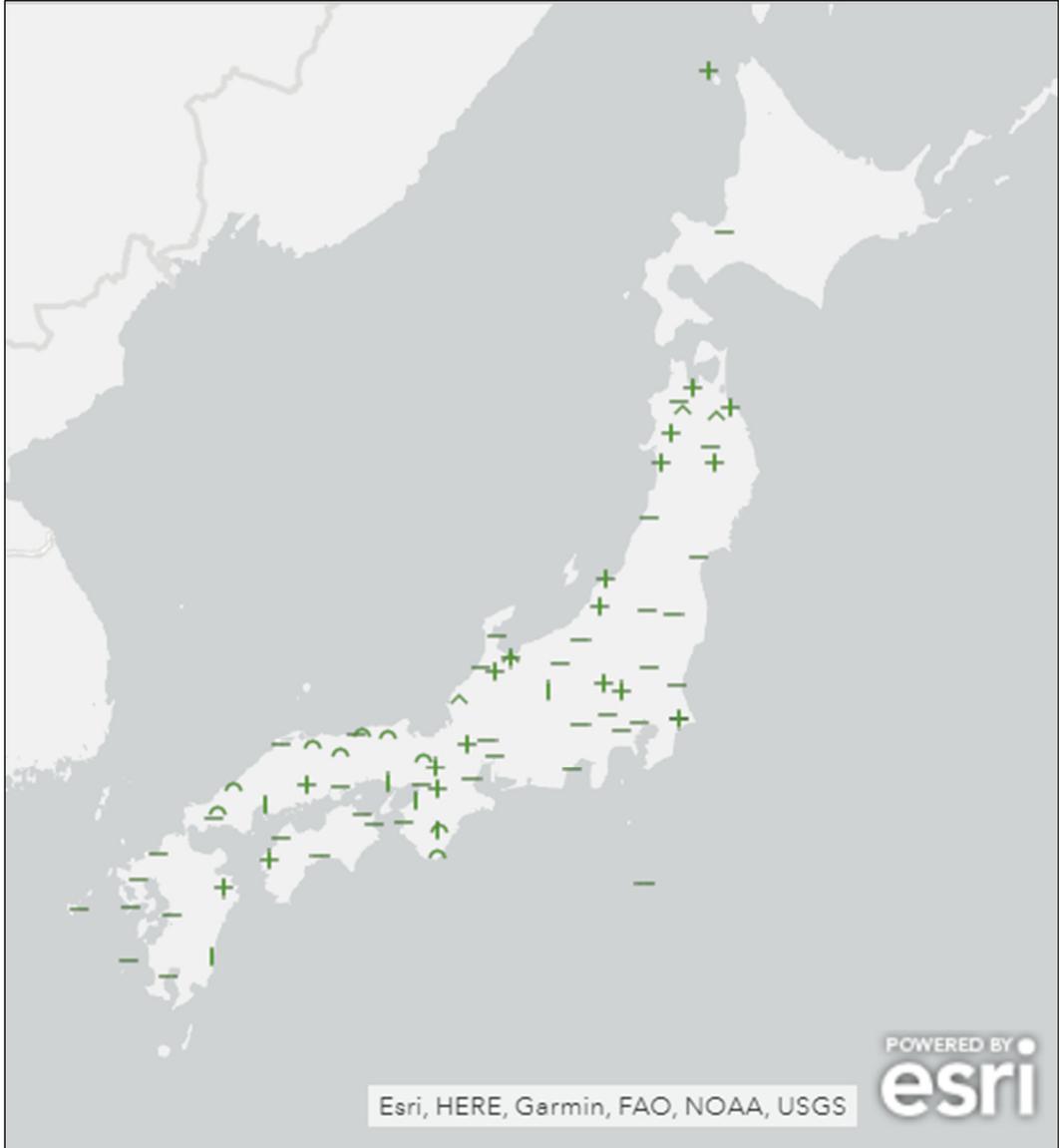
Figure 2.3.1: 'Chicken' in mainland Japan.

Figure 2.3.2: 'Chicken' in Northern Ryukyu Islands.
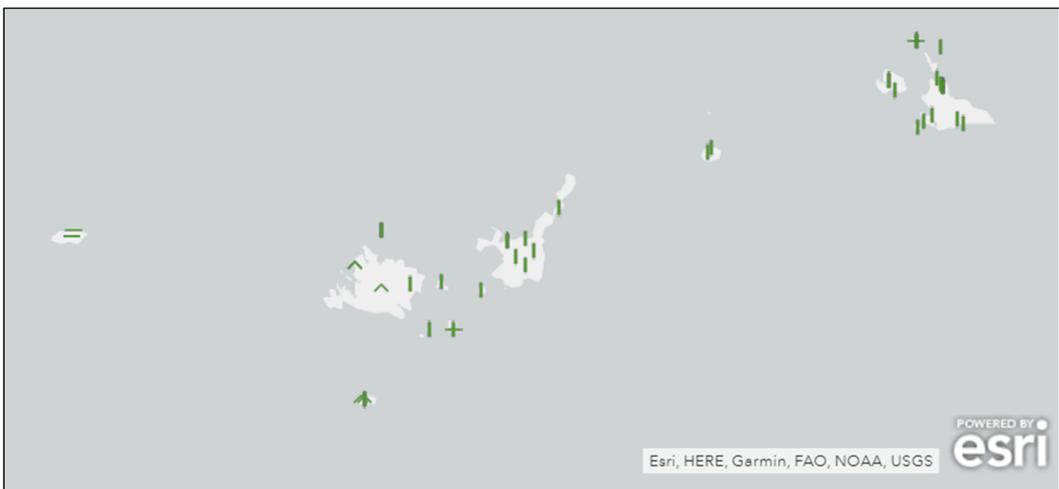
Figure 2.3.3: 'Chicken' in Southern Ryukyu Islands.

'Chicken' in Korean

Modern standard form for 'chicken' is 'tark'. The final consonat cluster '-rk' is a kind of morphophonemic transcription, actually pronounced as '[tak]' when spoken in isolation. The cluster '-rk' appears only when followed by a particle beginnning with a vowel. Middle Korean form was 'tʌrk'. Middle Korean is the language spoken from the middle of the 15th century to the end of the 16th century.

Modern and Middle Korean forms are almost the same except for the vowel. The vowel 'ʌ' has lost its phonemic status and merged with the vowel 'a' in many dialects except for the Cheju dialect. Also, the final consonant cluster '-rk' was pronounced as such in Middle Korean.

Dialect variation is not so great. There are only a few varieties concerning the quality of the vowel and the selection of final consonant in the morphophonemic alternaiton of the final consonant cluster.

A-1 tʌk, A-2 tak

A-1 is found only in Cheju dialects, which preserve the Middle Korean vowel /ʌ/.

(FUKUI Rei)
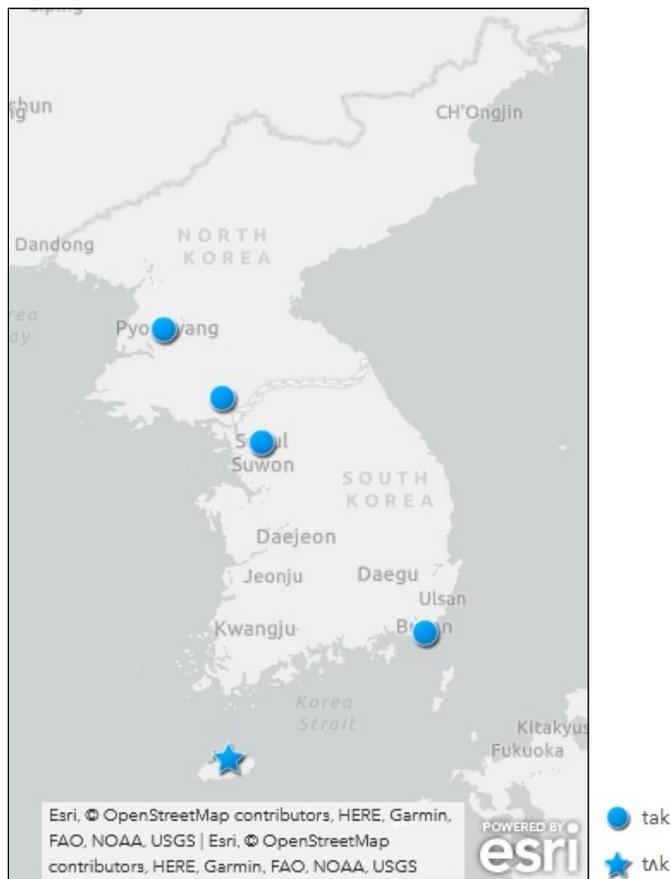
Figure 2.4.1: 'Chicken' in Korean.

'Chicken' in Sinitic

We classify the words based on stem types at first, then subclassify based on types of suffix or modifier.

A1: 鸡 tɕi (北京) ki (柳州) kue (厦门）

A2: 鸡子　tɕi tsɿ (信阳) kɛi tɕi（汝城）

A3: 鸡儿　tɕir (陽原) tɕi ər (太原)

A4: X鸡　小鸡哦siau tɕi ɤ (林县)

B: Others: 头牲　dʰou sen (祁阳)

Almost all of the words denoting Chicken distributed in China have 鸡 in their stems, and most of them are monosyllable type. However, from central to north west part, many dialects have suffix (A2, A3). Other types are very rare.

Reconstructed forms of 鸡 of Middle Chinese and Old Chinese are shown below.

	1	2	3	4
MC	kiei	kiei	kej	-
OC	kie	kieg	*kˤe	*kě

1:郭錫良(2010), 2: Karlgren (1957[1997]), 3: Baxter & Sagart (2014), 4: Schuessler (2007)

Old forms of 鸡 had [k-] in onsets, however many forms of northern dialects have affricates onset because of palatalization.

(YAGI Kenji)

A.

○ A-1 鸡 tɕi, ki, kue

△ A-2 tɕi tsɿ, kɛi tɕi

□ A-3 tɕir, tɕi ər

○ A-4 小鸡哦siau tɕi ɤ

B.

● B-1 Others: 头牲dʰou sen
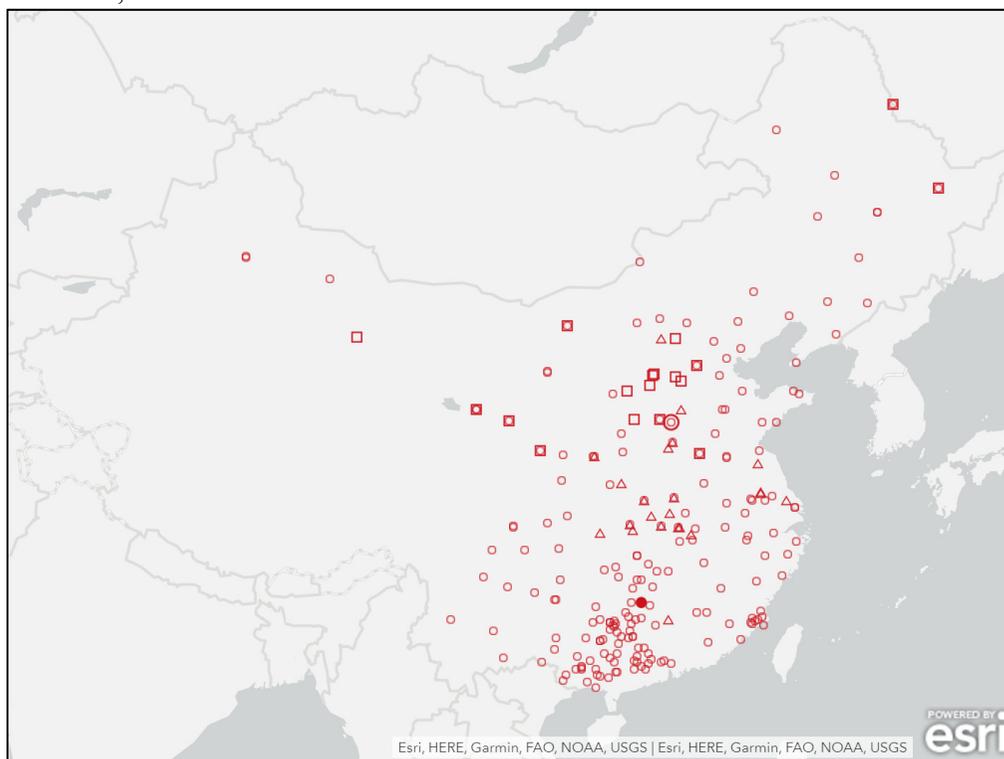
Figure 2.5.1: 'Chicken' in Sinitic.

'Chicken' in Hmong-Mien

There is only one type in CHICKEN: A: *kəi*.

This entry has only one type, thus exhibiting a uniform distribution. The source of this word must also be external, probably Sinitic. However, since similar forms are widely distributed throughout East and Southeast Asian languages, it is difficult to determine the origin of this word.

(TAGUCHI Yoshihisa)

A

Figure 2.6.1: 'Chicken' in Hmong-Mien.

'Chicken' in Kra-Dai

All forms in Kra–Dai belong to type A, of which the proto-form is reconstructed as *kai B1 by Li (1977). It is treated as a loan word from Sinitic. Subtypes are phonological varieties.

Type A1 has an unaspirated velar initial consonant and is distributed among the Tai and Ong-Be branches. Types A4, A5, A6, and A7 can be included in this class with vowel differences.

Types A2, A8, and A9 have uvular initial consonants. They are found among the Kra and Dong-Shui branches, which show archaism in many cases. Liang and Zhang (1996: 922) reconstructed *q- for this word in proto-Kra-Dai, regarding it as retention. In comparison, Ostapirat (2000: 224) reconstructed *ki A in proto-Kra, treating q- as a later innovation.

Type A3 is found in the Li branch on Hainan Island. The form for chicken is not found in Norquest (2016); however, he reconstructs *kh- for such an ordinary sound correspondence.

(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)

A. *kai* type

- A1: kai[1], kɐi[1'], kay[2], cay[2], kăj[3], kay[3], kaj[B1], kai[5], kɑi[5], ka:i[5], kay[5], tə[0]ka:i[5], kay[6], kai[35], kai[45]
- A2: qa:i[5], qa[24], qai[322], qai[33], qai[53], qai[54], qɛ[31], qɛ[55], qɤ[44], ɗɔ[33]qi[24], ʔai[24], a:i[5]
- A3: khai[1]

- A4: kjai[5], ʨai[5]
- A5: kɤɤ[3]
- A6: ki[5], ci[1]
- A7: kua[31](hau[31])
- A8: la[3]qɑ[45], lɔ[33]qe[44]
- A9: li[35]qɛ[44]

Figure 2.7.1: 'Chicken' in Kra-Dai.

'Chicken' in Tibeto-Burman

There are eight major stems (word roots) for 'chicken' in TB. These stems are etyma of the proto-level forms of Proto-Tibeto-Burman (PTB; see STEDT). They contain word formations that consist of a single stem (plus an affix) or compounds. We first classify the forms for 'chicken' into stem types and then into the compound types.

The etymology of Type A is *b(y/r)a (BIRD/BEE) in the PTB etymon. The etymology of Type B is *k-rak (FOWL/CHICKEN) in the PTB etymon, which is hypothesized as reduced from an onomatopoetic form (Alves 2015). Type C is derived from *ha:r (BIRD/FOWL/CHICKEN) in the PTB etymon, which is related to the Proto-Austroasiatic ʔIAR 'chicken' (Alves 2015). Type D is derived from *w(a/u) (BIRD/EGG/WING/FOWL) in the PTB etymon. Type E is from Written Tibetan (WT) de 'chicken'. The etymology of Type F is *chjaN^A (CHICKEN) in Proto-Karenic, and Type G is *daw (BIRD) in the PTB. Type H is *kak (CHICKEN) in the PTB etymon. This root is a possible allofam of *k-rak (the etymon of Type B) (STEDT).

In addition to the major types, there are several marginal roots (Type X). Among them, s-ŋak is 'bird' in the PTB origin, and kukhrī has an Indo-Aryan origin. The other forms are etymologically unknown. Moreover, the PTB etyma contain several meanings in addition to 'chicken'. Most of the above-mentioned forms have a single stem, but we also found several types of compound forms of the A-compound type: A+B, A+X(tɕy, phrug), and B+A; and the D-compound type: D+B and D+X(chi).

Type A is geographically widespread and is found in the northern and central-eastern parts of the TB area (Tibetic, Qiangic, and rGyalrongic) and in northern India (Darma). Type B is widespread across the branches of TB. This type is found in the southern and central-eastern parts (Qiangic and Lolo-Burmese, Bai, Trung, Tani, and Deng). Type C is found in two areas, northern Yunnan (Qiangic, Naxi, and Malimasa) and north-eastern India (central Naga, Chin, Meithei, and Tangkhulic). Type D is found in northern Sichuan (Qiangic) and north-eastern India (Angami-Pochuri, Sal, and Tangkhulic groups). Type E is found in the eastern part of the Tibetosphere. Type F is only found in southern Burma (Karenic), and Type G is found in north-eastern India and Bangladesh (Bodo-Garo). Type H is scattered from northern Burma (Rawang) to southern Tibet (Monpa and Basum) to central Nepal (Newar). Furthermore, Nishi (1990) has suggested that this form is common in the Tamang-Ghale-Kaike group.

From the perspective of the relationship with the domestication of the chicken, in areas where domestication was early (e.g., in the southern part of TB), Types B, H, and C (meaning 'chicken') are dominant. In areas where domestication was delayed (Tibetan Plateau and the Chittagong Hill Tracts), Types A and G, in which the original meaning is 'bird' (not directly 'chicken'), are found.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

— A *b(y/r)a* type

pɕa, ɕa, ɸɕa, pya, bja, wɕawo, pwa, ptsia³³, pkwaʔ, pwaʔ, patʃu, tsjaᴴ, ptsa³, za²¹, ᴾtso, ᴾtɕa, tɕa, sɔ, etc; ɸɕamu, ɕawo, tɕapo, etc. (suffixed).

⊖ A-compound type

○ B. *k-rak* type

ʈá, zo³⁵, ro⁵³, ra³⁵, ɦæˑ⁵³, zo⁵⁵, va³³, cɛʔ, zi³³, ze³³, ɣa²¹, ɣo¹³, ɣo³³, ɣua³⁵, tɕɛʔ, je⁴⁴, kraʔ, krɔ', kzɑ²⁴, kɪɑi⁵⁵, kzuaʔ⁵⁵, ğâʔ, nghoq, ha³³, woʔ, a³³, kjɔʔ³¹pho⁵³, paro, puruk , porok , ja³³tɕhuʔ³³, lakyon, ɣa³¹pha³⁵, ha³³tɕi̠³³, etc.; a⁵⁵zi³³, a⁵⁵zɻ³³, ɑ⁵⁵ɣɑ⁵⁵, a³¹xa³³, a²¹ha³³, etc. (prefixed).

✳ C. *ha:r* type

æ̃ʎ, æ²¹, hən, a:i, ʔǎar, ho¹noʔ³, yen nao, rààw, etc.

△ D. *w(a/u)* type

zy, ẏ, jy, jì, wu, u³¹, etc.; thevü, aunak (prefixed).

△ D-compound type

▷ E. *de* type

te, de, etc.; tewo (suffixed)

▢ F. *chjaN* type

ʃĩ⁵⁵, cha³³, ɕí, ɕja³¹, chânɴ, etc.

□ G. *daw* type

do, tó, tau; dupisa, daosa (suffixed).

◎ H. *kak* type

khā, khɑ³³, khaʔ⁵³, khɛ ma.

⌄ X. others

s-ŋak, tɕy, etio, haku, kukhurā, chaṅts. me⁵⁵tio⁵⁵, tiu⁵³, chi, xo:co, etc.
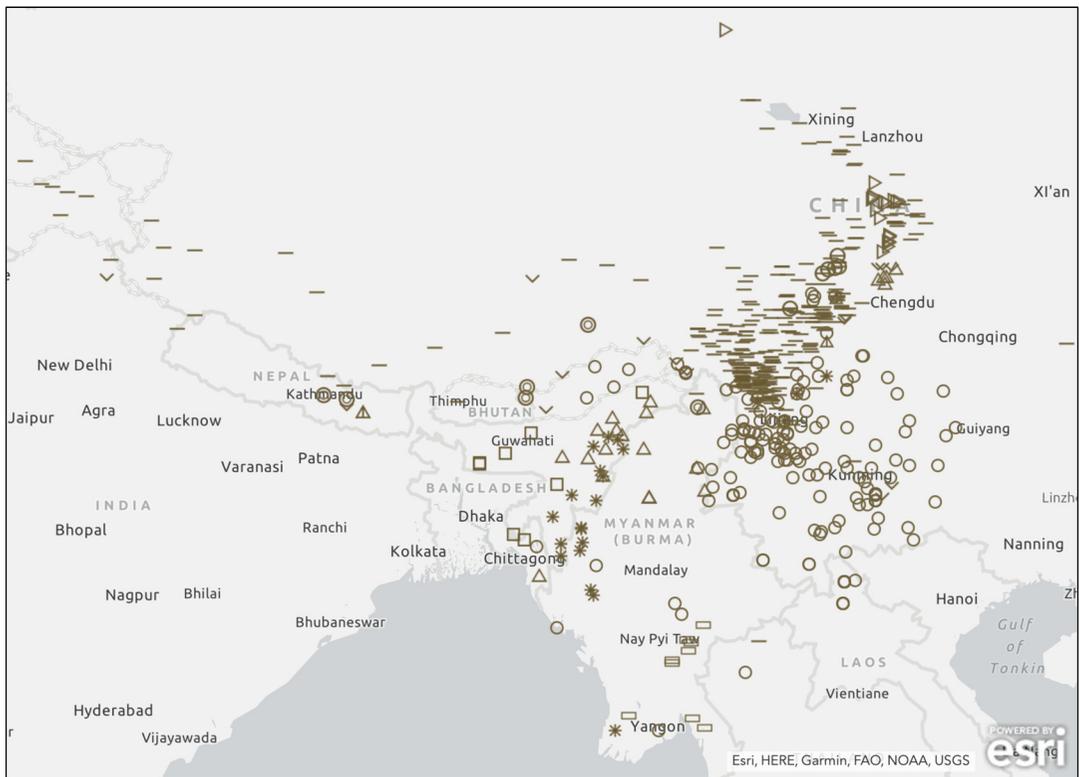
Figure 2.8.1: 'Chicken' in Tibeto-Burman.
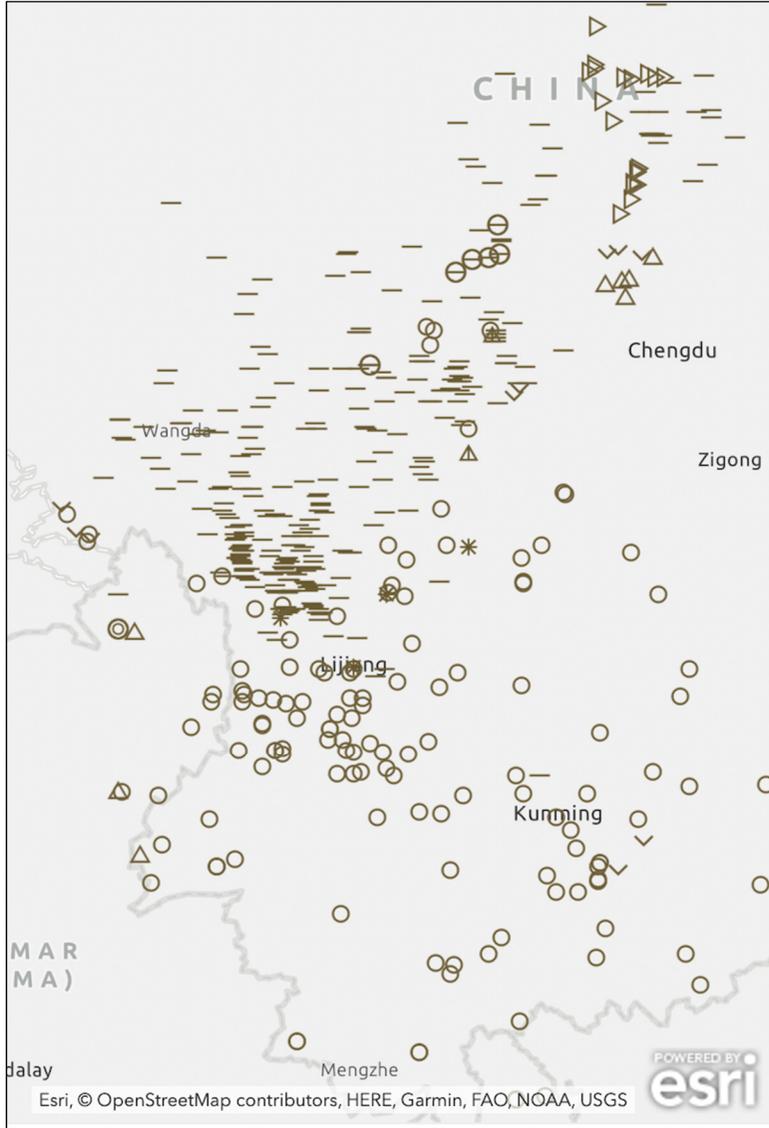
Figure 2.8.2: 'Chicken' in Tibeto-Burman (detailed).

'Chicken' in Austroasiatic

The word forms meaning "chicken" in Austroasiatic are classified into ten types, as follows.

A. *(s)ʔiar type*
Proto Bahnaric: *ʔiar (Sidwell 2011); ʔjɛr (Bahnar [Golar]), ʔjar (Mnong [Central]), ʔi̱ (Sedang)
Proto Khasic: *sʔiar (Sidwell 2012); sʔiar (Pnar [Jowai]), sʔi (War [Amwi])
Proto Khmuic: *(s)ʔiər (Sidwell 2013); hʔiar (Khmu), ʔiːr (Phong), ʔeːl (Khsing-Mul)
Palaungic: ʔiər (Palaung)

B. *ʔndruuj type*
Proto Katuic: *ʔndruuj (Sidwell 2005); ndruuj (Ngeq), nt'ruəj (Souei), ntruaj (Pacoh)

C. *hlɛːk type*
Proto Pearic: *hlɛːk (Headley 1985); liək (Chong [Samre]), lɛːk (Chong [of Chantaburi])

D. *r-kaː type*
Proto Vietic: *r-kaː (Ferlus 2007); kaː¹ (Thavung), rəkaː (Chứt [Rục])

E. *sim type*
Munda: *sim (Proto Kherwarian: Munda 1968); sim (Santali [Bodobelghoria]), siːm (Santali [Heben])
Katuic: siem (Katu [An Diem])

F. *chaaŋ type*
Monic: chaaŋ (Nyah Kur), caɲ (Mon)

G. *manuk type*
Aslian: manuk (Kensiu)

H. *moan type*
Khmeric: moan (Khmer)

I. *ciəj⁶ type*
Mangic: ciəj⁶ (Mang)

J. *ʔajam type*
Loan from Malay: ʔajam (Temiar)

The most widely distributed word form in Austroasiatic is the A type *(s)ʔiər*, which is shared among Khasic, Palaungic, Khmuic and Bahnaric. Proto Khasic and proto Khmuic forms have an initial *sʔ-, while the others have only ʔ-.

The B type *ʔndruuj* is widely seen in Katuic.

The C type *hlɛːk* is common among Pearic languages. Two forms cited here (Chong [Samre] and Chong [of Chantaburi]) no longer preserve the proto initial *h-.

The D type *r-ka* is common in Vietic, in which the monosyllabic languages no longer preserve the presyllable *r- (e.g., *gà* in Vietnamese).

The E type *sim* is widely seen in Munda, while Katu [An Diem] uses the form *siem*, which possibly belongs to the E type. Since it is the only form found in Katuic, where the B type *ʔndruuj* is most common, the form *siem* might be regarded as a case of distant borrowing.

(SHIMIZU Masaaki,
MINEGISHI Makoto)
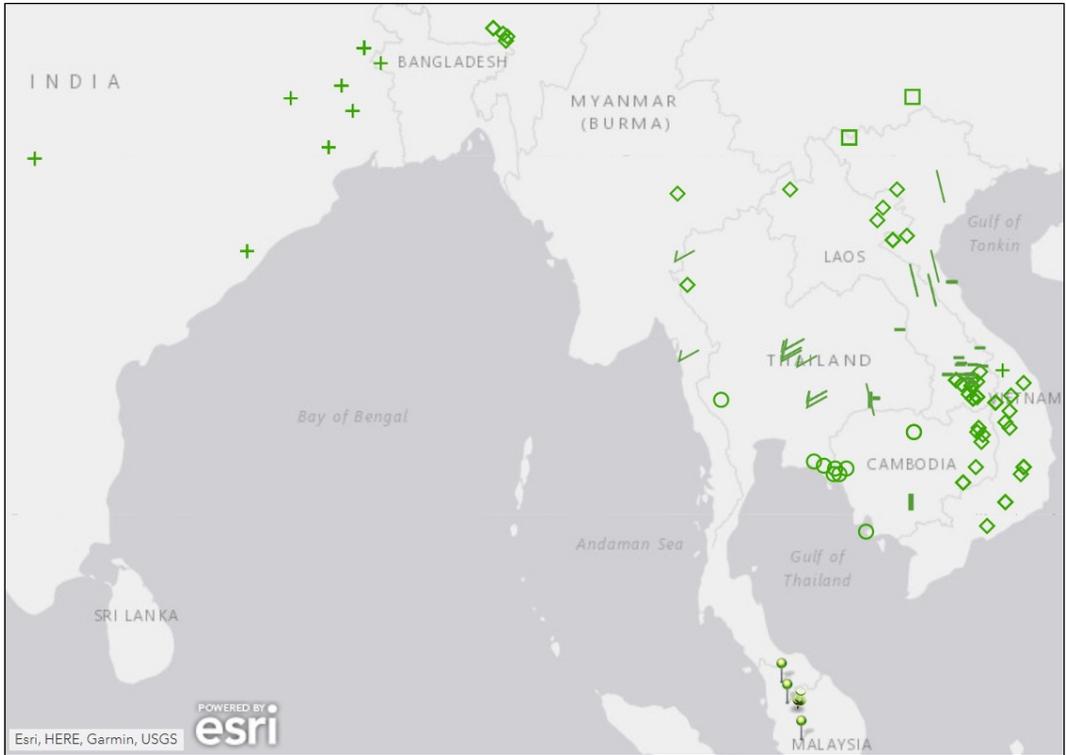
Figure 2.9.1: 'Chicken' in Austroasiatic.

◇	A	*(s)ʔiar type*
–	B	*ʔndruuj type*
○	C	*hlɛːk type*
\	D	*r-kaː type*
+	E	*sim type*
↙	F	*chaaŋ type*
⸙	G	*manuk type*
│	H	*moan type*
□	I	*ciəj⁶ type*
⸙	J	*ʔajam type*

'Chicken' in Austronesian

As in many other parts of the world, chickens are extremely important livestock in the Austronesian region.

There are various word forms that are assumed to be derived from an onomatopoeia that imitates a rooster crow or chicken call, such as *ʦiotsio*, *koka*, *kuka*, *kuʈuʔ*, and *kiokio*. These forms are grouped together as Type E.

Other forms that do not seem to be derived from onomatopoeia are classified into four types: Types A, B, C, and D. Type A forms contain /s/ and a half vowel. *Sīyop* (Kalinga Limos), *sisiw* (Tagalog), and *isiw* (Aklanon) are examples of this type.

Type B forms have /a/ and a nasal or /l/ or /y/, which include /manuk/ and /anuk/, the most frequently found forms. This type consists of two major subtypes. B-1 is MANUK type, which include *manok* (Paiwan), *manuk* (Kagayanen, Batak Toba, and Manggarai), *manu* (Da'a, Wolio, Ngada, Sika, and Roti), *manuʔ* (Uma and Bugis), *manɔʔ* (Acheh), *maŋ* (Manam), *man* (Mbula), *maluʔo* (Gorontalo), *meniŋ* (North Tanna), *meyʌ* (A'jië), *menu* (Kinbati), and *malek* (Ponapean). The other subtype B-2 is ANUK type, such as *anuʔ* (Isnag), *anuk* (Sarangani Blaan), and *anuk* (Murut).

Type C forms begin with /a/ and end with a nasal, the most typical form being /ayam/. The examles of this type are: *ayam* (Minangkabau, Indonesian, and Sundanese), *adʒam* (Madurese), *hayam* (Sundanese), and *dʒaŋaŋ* (Konjo).

Type D comprises just one form, /moa/, which is found in Tongan, Samoan, and Tahitian.

Word forms that seem to be derived from onomatopoeia are categorized as type E. Type E-1 forms have two /k/ sounds or two /ʔ/ sounds, such as *koka* (Rukai), *kuka* (Paiwan), *kiokio* (Eastern Fijian), and *ʔuʔui* (Rotuman). Type E-2 forms have /k/ and /t/ phones, such as *kuʈuʔ* (Javanese) and *pitik* (Balinese). Type E-3 forms have /t/ or /ʦ/ sounds such as *ʦiotsio* (Tsou), *tepu-t* (Buru), *toru* (Dobel), *tataro* (Dami), *tatariʔ* (Adzera), *tītān ato* (Paamese), and *tō* (Western Fijian). Type E-4 consists of word forms /k/ and /r/ sounds. Examples include *kirek* (Takia), *kokaro* (Nyindrou), *kukurə* (Irarutu), *kəkaruk* (Tolai), *qɔqɔrɛɣ* (Buang), *rekorekʷa* (Kilivia), *kokoroku* (Motu), *kokorako* (Roviana), *kʰokorako* (Mannge), *karaikoa* (Lau), and *lakaporo* (Lewo).

There are various other forms which are categorized into Type F. *Wayłuŋ* (Atayal), *ʦiotsio* (Tsou), *ekiŋ nomal* (Kaulong), *ro-mʷelaul* (Port Sandwich), *reia* (Kwamea), *yalek* (Nemi), *ɟá* (Cémuhî), *watitewe* (Nengone), *pɯao* (Marshalese), and *maɾüxa* (Woleaian), are the examples of type F.

Type A is found in the Philippines. Type B has the widest area of distribution with a few languages in Taiwan and the Philippines, many languages in Sulawesi, and some languages in Sulawesi. Moreover, Eastern Nusa Tenggara, Papua, and the Pacific islands have Type B languages. Type C is primarily found on the islands of Sumatra and Java. Type D is found exclusively in the South Pacific islands. Type E, which consists of the various forms presumably derived from onomatopoeia, also has a wide distribution but is predominantly found in Papua, Solomon Islands, and the Pacific islands.

(UTSUMI Atsuko)

⌒ A: word forms containing forms contain /s/ and a half vowel: *sīyop, sisiw, isiw*

Ψ B-1 MANUK type: *manok, manuk, manu, manuʔ, manɔʔ, maŋ, man, maluʔo, meniŋ, meyʌ, menu, malek*

Y B-2 ANUK type: *anuʔ, anuk,anuk*

◯ C: forms begin with /a/ and end with a nasal: *ayam, adʒam, hayam, dʒaŋaŋ*

◎ D: *moa*

— E-1 onomatopoeia with two /k/ or /ʔ/: *koka* (Rukai), *kuka* (Paiwan), *kiokio* (Eastern Fijian), and *ʔuʔui ber, bea, pea, beya*

= E-2 onomatopoeia with /k/ and /t/: *kuṭuʔ* (Javanese) and *pitik* Various forms

‖ E-3 onomatopoeia with /t/ or /ts/: *tsiotsio, tepu-t, toru, tataro, tatariʔ, tītān ato, tō*

▼ E-4 onomatopoeia with /k/ and /r/ sounds: *kirek, kokaro, kukurə, kəkaruk, qɔqɔrɛɣ, rekorekʷa, kokoroku, kokorako, kʰokorako, karaikoa, lakaporo*

✚ F Other forms: *wayłuŋ, tsiotsio, ekiŋ nomal, ro-mʷelaul, reia, yalek, ɟá, watitewe, puao, marɯ̈xa*

Figure 2.10.1: 'Chicken' in Taiwan and the Phillipines.

Figure 2.10.2: 'Chicken' in Indonesia.
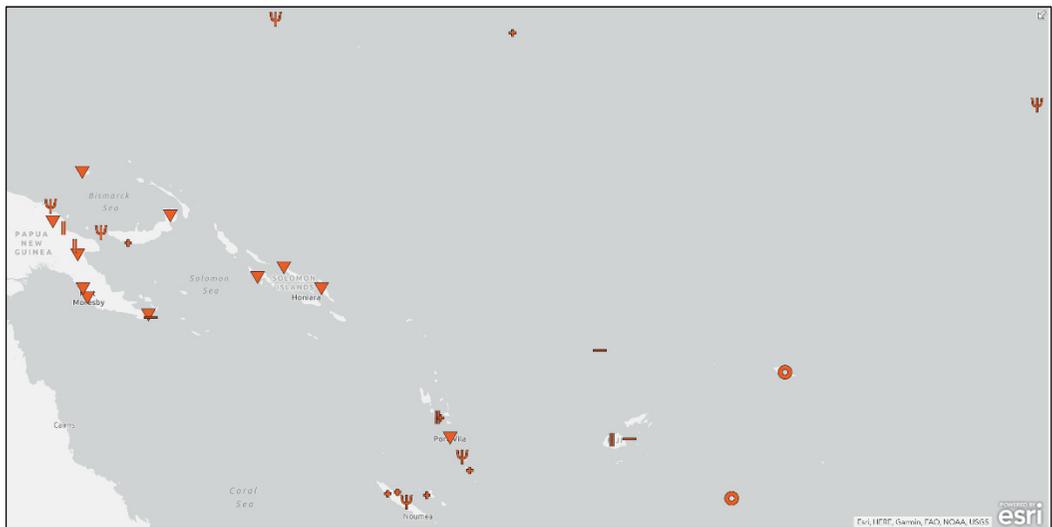
Figure 2.10.3: 'Chicken' in Papua and the Pacific.

'Chicken' in Tungusic

In Tungusic languages word forms for 'chicken' would be classified in some types as below:

A KAKARA:Orochon *kakara,* Ewenke *xaxara*

B ČOKO: Hezhe *toko,* Sibe *tʂɔqɔ,* Nanay *chiko*

C Udehe *nai*

   As you already know, the languages which have words for chicken are situated in China or near China. The other languages, such as Evenki, Ewen, Negidal, Ulich and Uilta etc. which are living in Siberia, have no word (or if they have, it is borrowing from Russian: ex. *kuuritsa*). This is because Tungusic people traditionally has no poultry farming.

(MATSUMOTO Ryo)

A KAKARA

B ČOKO

C *nai*

Figure 2.11.1: 'Chicken' in Tungusic.

'Chicken' in Uralic

In Uralic there are various forms for 'chicken' as below. Here I just classified according to the sound forms, it could not be referred to the reason why such many forms they have.

A KANA: Finnish *kana*, Veps *kana*, Livoninan *kanā*

B SARAZ: Mordvin *saraz*

C ČI-: Mari *čywe*, Komi *čipan*, Hungarian *tyúk*

D KUREK: Udmurt *kureg*, Khanty/Mansi *sis-kurek*

A type includes all the languages of Balto-Finnic branch, their forms have very little difference. It is interesting that C type and D type cover the different branches of languages. For example, D type KUREK are observed in Udmurt of Permic branch and Khanty-Mansi of Ugric branch. On the other hand, the other Permic Komi and the other Ugric Hungarian have similar form C type.

In the northern area, Samoyedic and Sami languages have no form for chicken. It would be because it is too cold to have the poultry culture.

(MATSUMOTO Ryo)

— A KANA
● B SARAZ
□ C ČI-
☆ D KUREK
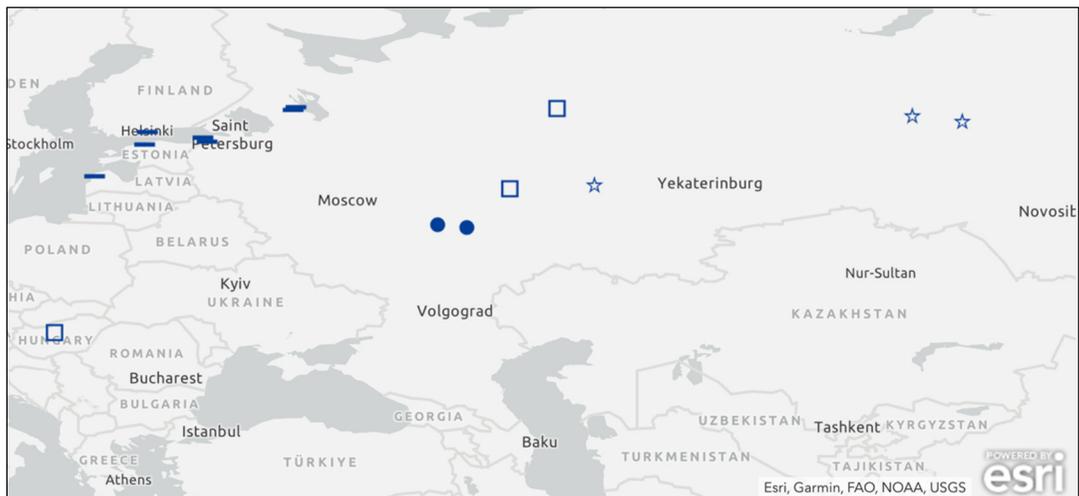
Figure 2.12.1: 'Chicken' in Uralic.

'Chicken' in Mongolic and Turkic

1. Mongolic

Mongolic languages tend to use a generic word for 'chicken,' and the distinction of male and female chickens is made by adding a gender-indicating word. (But, the word for 'chicken' may mean 'hen' in a narrow sense.)

The word for 'chicken' in most languages is a *takā* type. The Mongolic *takā*-type words for 'chicken' and the Turkic *tavuk*-type words for 'hen' are cognates. It is considered that the Turkic word, which originally meant 'a domestic fowl,' entered Mongolic sometime in the past (cf. Clauson 1972: 468).

Dagur uses the form *kakrā*. The word may be related to a Tungusic word for 'chicken' (cf. Orochen *kakara*), which may be of an onomatopoetic origin.

Moghol in Afghanistan has a Persian loanword, *murg* and the form *nultu*, which literally means 'having beaks' (*nul*, a Persian word for 'beak'; *tu*, a Mongolic suffix designating possession).

2. Turkic

Turkic languages tend to distinguish male and female chickens.

Hen

In most modern languages, the cognates of the Old Turkic word *takïgu*, which we call *tavuk*-type words here, are used. The Chuvash form *čăx* also belongs to this type.

Uzbek has the forms *tovuq* and *makijon*. Uighur uses the form *mekijan*.

Salar in Qinghai and Gansu provinces has the form *anəš* (cf. *ana* 'female').

In southern Siberia, Tuvan and Altay use a Turkic-originated Mongolic word (*dagā*, *takā*). The Tofalar form is *taqqïnjaq*. In Altay and Shor, the word *kuš*, which originally meant 'bird,' is used for 'hen.' (Altay also has the form *takā*.)

Sakha and Dolgan in northeastern Siberia use a Russian loanword (*kūrussa*, *kurisa* < Russian *kuritsa*).

Rooster

The *ätäč*-type words are found in Tatar, Bashkir and Kazakh.

A Persian loanword is used in the southern region with modifications (*xorus*, *xoruz*, *xoras*, *koraz*, etc.).

Chuvash uses the form *avtăn* (< *avt-an*, 'singing').

The forms *erkek tavuk* (Karachay), *askïr dagā* (Tuvan), *askïr taqqïnjaq* and *er taqqïnjaq* (Tofalar) are made by attaching a gender-indicating word to the noun for 'chicken.'

Shor uses the form *tanaq*, which is close in form to the Khakas word for 'hen,' *taŋax*.

The Salar form is *guŋgu* (cf. Chinese *gong* 公 'male').

Khakas' *petux*, Altay's *pötük*, and Sakha's *bötūk* in northeastern Siberia are all from the Russian *petux*.

The words for 'rooster' in most Turkic languages are loanwords, which may indicate that the gender distinction manifested in later periods.

(SAITÔ Yoshio)

A. *takā* type

⬚ tæxā, daxjā, tɛxā, dæxʲā, dɪxā, daxā, takā, takǣ, taxʲān, taxjā, daχɢa, tɯɢa, taχa, tɯχa, təχa, tχa, taɢaū

B. modern loanword

⌒ **B-1.** kakrā (< Tungusic)

△ **B-2.** murg (< Persian)

C. *tavuk* type

▭ **C-1.** tavuk, tojuq, towuk, towux, toux, touq, tawux, tavux, tavïk, tavïq, tawïk, tavuq, tawïq, tōk, tovuq, toxu, tahɢaɣə, taŋax

▐ **C-2.** čăx

D. taqqïnjaq

○ taqqïnjaq

E. *takā* type

⬚ takā, dagā

F. *mekijan* type

❘ mekijan, makijon

G. kuš

═ kuš

H. anəš

◎ anəš

I. modern loanword

∠ kūrussa, kurisa (< Russian)
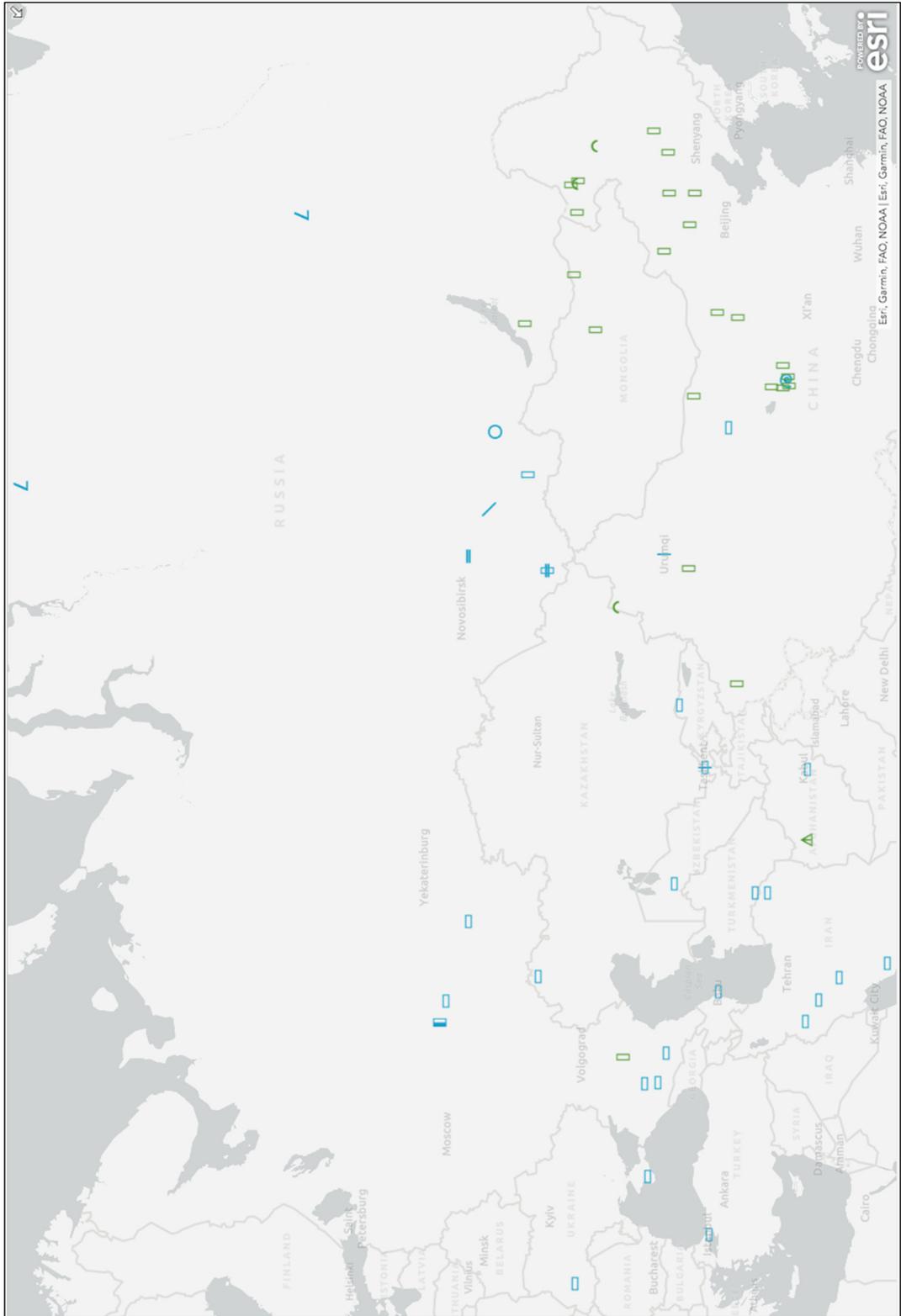
Figure 2.13.1: 'Chicken' in Mongolic and 'hen' in Turkic.

Figure 2.13.2: 'Chicken' in Mongolic and 'hen' in Turkic (The Mongolian Plateau and its vicinity magnified).

A. *xorus* type

▽ xorus, xoros, xurus, xoras, xoruz, xo'roz, horoz, horaz, koroz, koraz, qoraz (< Persian)

B. *ätäč* type

Ⅿ ätäč, ätäs, əteš

C. tanaq

＼ tanaq

D. avtăn

✳ avtăn

E. guŋgu

✚ guŋgu

F. type with a gender-indicating word

▯ **F-1.** erkek tavuk

▯ **F-2.** askïr dagā

◑ **F-3.** askïr taqqïnjaq, er taqqïnjaq

G. modern loanword

∠ petux, petuk, pötük, bötūk (< Russian)
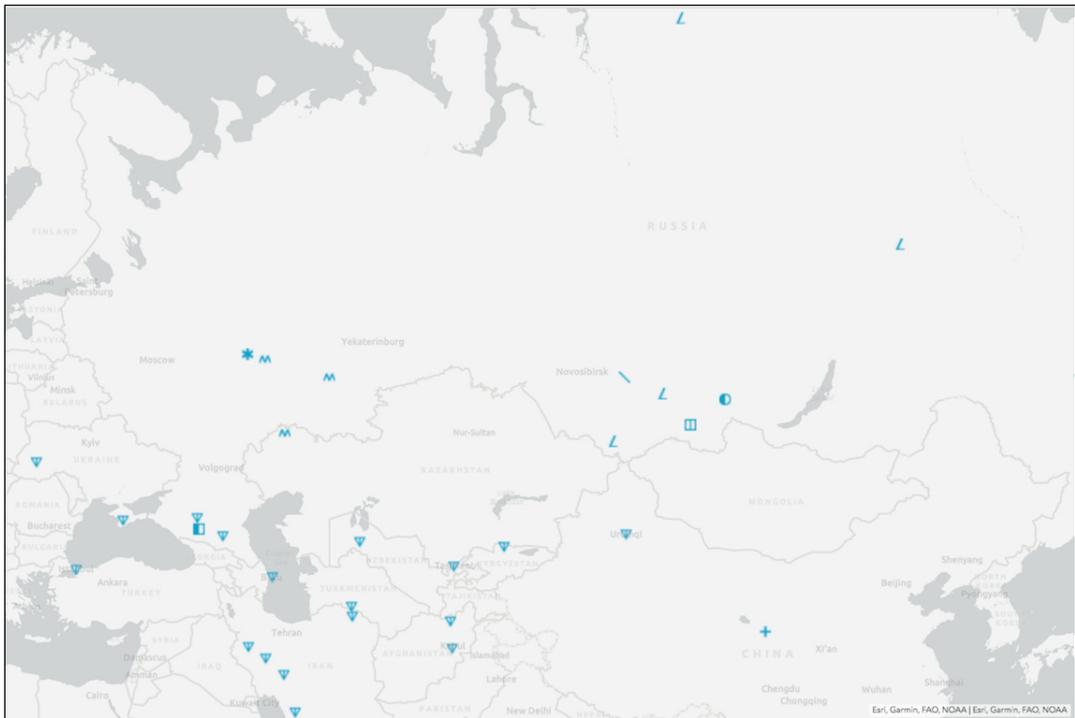
Figure 2.13.2: 'Rooster' in Turkic.

'Chicken (Hen)' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. When a language has several words for chickens, I targeted 'hen (female chicken)'.

As for the distribution of 'chicken' words, Indo-European languages in South Asia widely employ the type A, but in the West Coast of India there is exclusively occupied by the type F. The type C is for European Romanic languages. Burushaski and neighbouring IA languages employ the type D. The distribution of the type B is difficult to explain with any reasonable tendency. And Type E is seen mainly in the northwest part, but also in Nepal.

The most major *kukkuṭá/í* type A is derived from Sanskrit onomatopoeic **kukkuṭá* *कुक्कुट* 'cock' and **kukkuṭí* *कुक्कुटि* 'hen'. Forms of this type are used by most IA and Nuristani languages. If a language distinguishes 'cock' and 'hen', they often use /a, o, u/ sounds for the former and /i/ sound for the latter. In the Chitrali languages, however, employ somewhat different forms for the part of 'chicken' along with the common morpheme *nar* 'male' for 'cock'; Khowar *narqṓqu* 'cock' vs. *kahák* 'hen', and Kalasha *narkū́ku* vs. *kakawáŋk*. It shows a different system from Dameli, which has a simpler formation, that is, *kukuř* 'cock' vs. *kukuř-pai* 'hen'. Note that there may be some non-cognate forms in the type A, because the forms are based on the sounds of the birdcall of chickens.

The *mṛgás* type appears in IA languages and the Mixed Great Andamanese language. The Sanskrit form is *mṛgá* *मृग* and refers to 'wild beast, deer', so the modern IA words are not derived from it. The PIA form is also **mṛgás* with the meaning, while Proto-Iranian is **mṛgáh* 'hen, bird' (< PII **mṛgás* 'forest animal'). Nowadays there are languages in South Asia use the forms derived from the Iranian origin via Classical Iranian *murɣ* مرغ 'hen, bird', and then it has reached up to Mixed Great Andamanese *murgithire* as a loan word compounding with *thire* 'child'; And thus its meaning 'chicken' may only refers to 'chick' actually.

Both the *khaini* type C and the *kombḍa* type F are etymologically unclear (reconstructions mine), and here reconstructed by me. These may also be onomatopoeias. The former, C, is used in European and Iranian Romanis (whereas Jerusalem Domari employs *kukar* of the type A), and the latter, F, is used in Marathi, Konkani, and Nihali as concentrating at the west of central India.

The fourth major type *qarqámuc* is in all Burushaski lects and two IA languages, Shina and Domaaki. The IA languages no longer hold any alternative words for 'chicken'. This type is onomatopoeic, too.

Next, the *ćaṭaka* type is used in the Hindukush range and in Nepali. The original Sanskrit onomatopoeic word *ćaṭaka* चटक means 'sparrow' from its twitter voices.

There are few exceptional forms for 'chicken' in South Asia. Mixed Great Andamanese *(bhuku)mɔčɔ* and Jarawa *mɔiča* seem to have relationship each other. Kharia Thar *saŋkɔe* 'hen' is etymologically mysterious, but the combination with *laŋgɔe* 'cock, rooster' makes for an interesting word formation.

(YOSHIOKA Noboru)

A. *kukkuṭá/í* type (53) ◇

    kukkaṛī, kukkiṛ, kukkhṛī, kukuṛā, kukuṛī̃,
    kukuṛé̃, kukǔ̃ṛ, kukiṛ, kukṛī̃, kukaḍa,
    kukḍi, kukuḷu, kikiḷiya, kukuṛī̃, kukúr,
    kukar, kəkir, kíkir, kukhro, kəkyuř,
    kukuř-pai, kukwī̃̃, kakawáŋk, kakwéki,
    kukǎ̃ī̃, kukǔ̃, kukú, kokō, kok, kākók,
    kahák, kakág, kuguī̃, qakok, koráī, kūri

B. *mṛgás* type (7) ✂

    murg, murgi, murɔgi, murɨgi, murgithire,
    murɣ, margho

C. *khaini* type (7) ✗

    khaini, kaini, khanji, kanni, kakni, qaqini

D. *qarqámuc* type (5) ⭍

    qarqaámuc, qarqámuš, karkaámuš

E. *ćaṭaka* type (4) ▯

    čēriḍ, cǎ̃raṭik, čarǎ̃ṇṭak, čallo

F. *kombḍa* type (3) ◸

    kombḍā, komba, kombo

G. others

    [*moičo* type (2)] mɔiča, (bukhu)mɔčɔ;
    télu, saŋkɔe, taap
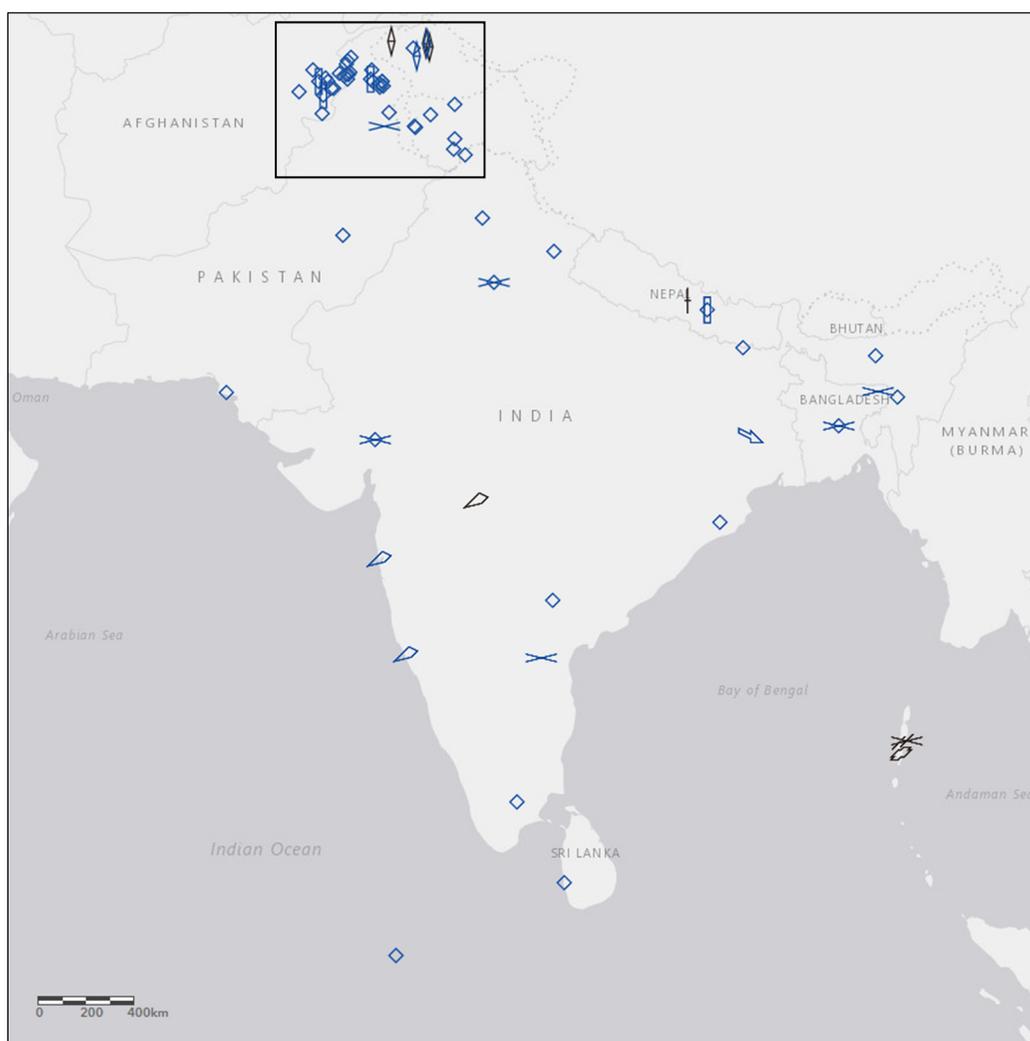
Map 2.14.1: 'Chicken' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).
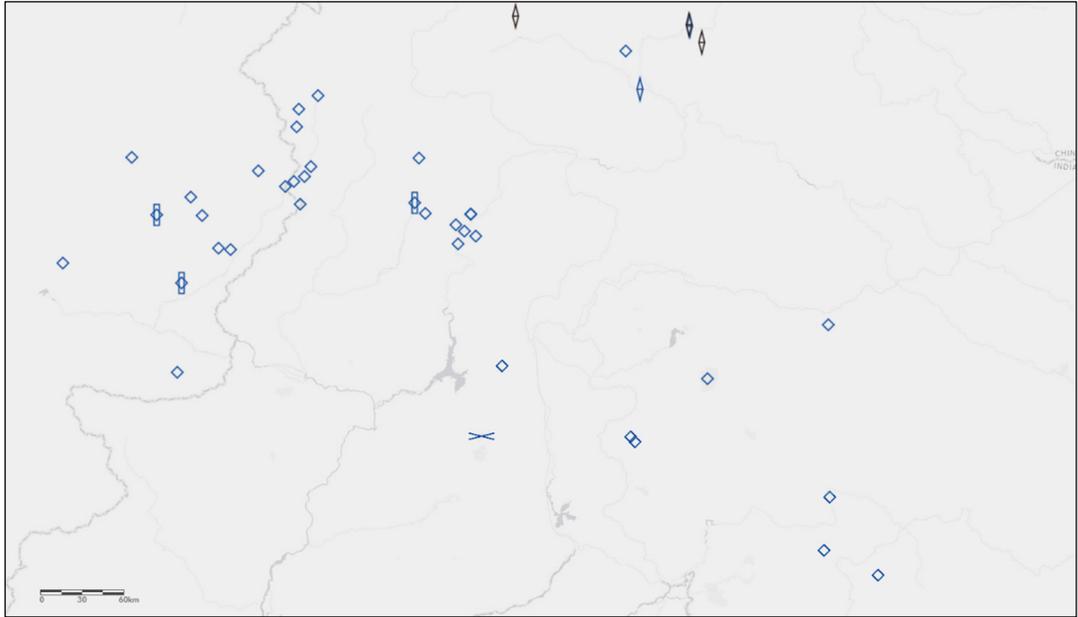
Map 2.14.2: 'Chicken' in northern Pakistan (the area encloed by the rectangle in Figure 2.14.1).
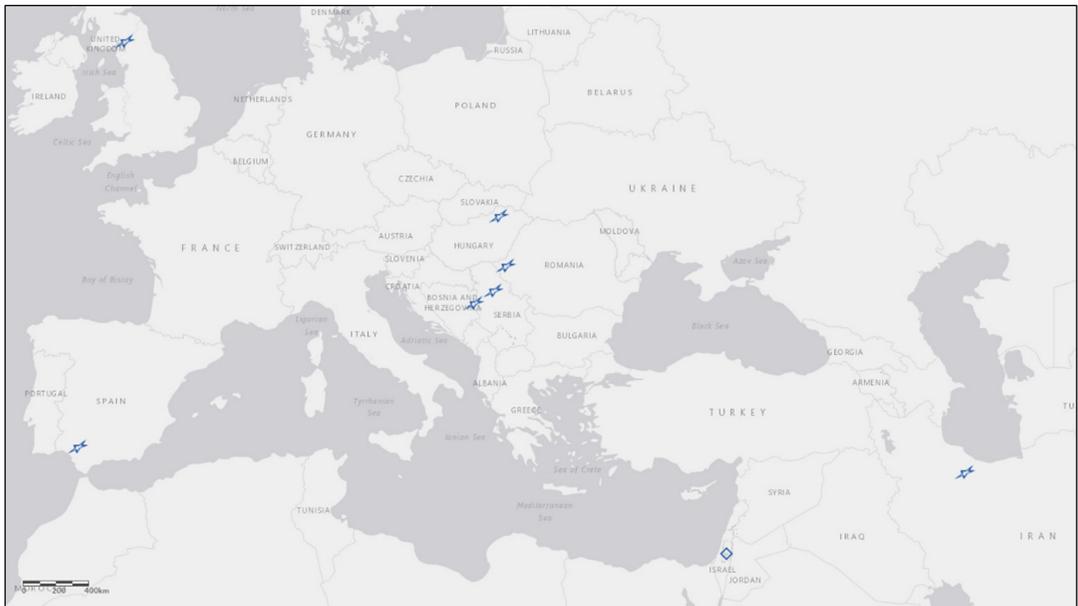
Map 2.14.3: Types for 'Chicken' in Indo-Aryan languages outside South Asia.

'Fowl' in Dravidian

DEDR identifies three etyma for 'fowl, jungle fowl' in Dravidian languages. Reflexes of the etymon KOZI (DEDR #2028) in South Dravidan languages and Telugu have their respective regular reflexes of intervocalic *Z, voiced retroflex fricative. As reflexes of KOR (DEDR #2160) in South Central and Central languages usually have irregular plural forms often with a retroflex alternant of dental R, KOR and KOZI may be considered as cognate.

If all the South, South Central and Central Dravidian words that specifically refer to 'a fowl' are reflexes of a single proto-form, it may serve as an indication that the proto-language was spoken in the habitat of these birds of the genera *Gallus*, South and Southeast Asia.

DEDR (#2013) suggests that its reflexes Kurukh *xe:r* and Malto *qe:ru* may be related to Tamil verb *kēru* 'to crackle (as a hen)' in DEDR #2009. Brahui employs borrowings from Iranian.

(KODAMA Nozomi)

⌄ kōṟi

⌣ koˑli, kōḷi   ◸ qéru, xēr

Ⅴ kōḍi   ◖ xaṛōsk / kakuṛ

⌐ kwï·ḍy, kwï·y

⌒ ko·y, ko:yi

╱ kor, korr, korru, koṟ

+ kozu, kuzu:, koju, kuy, koyu
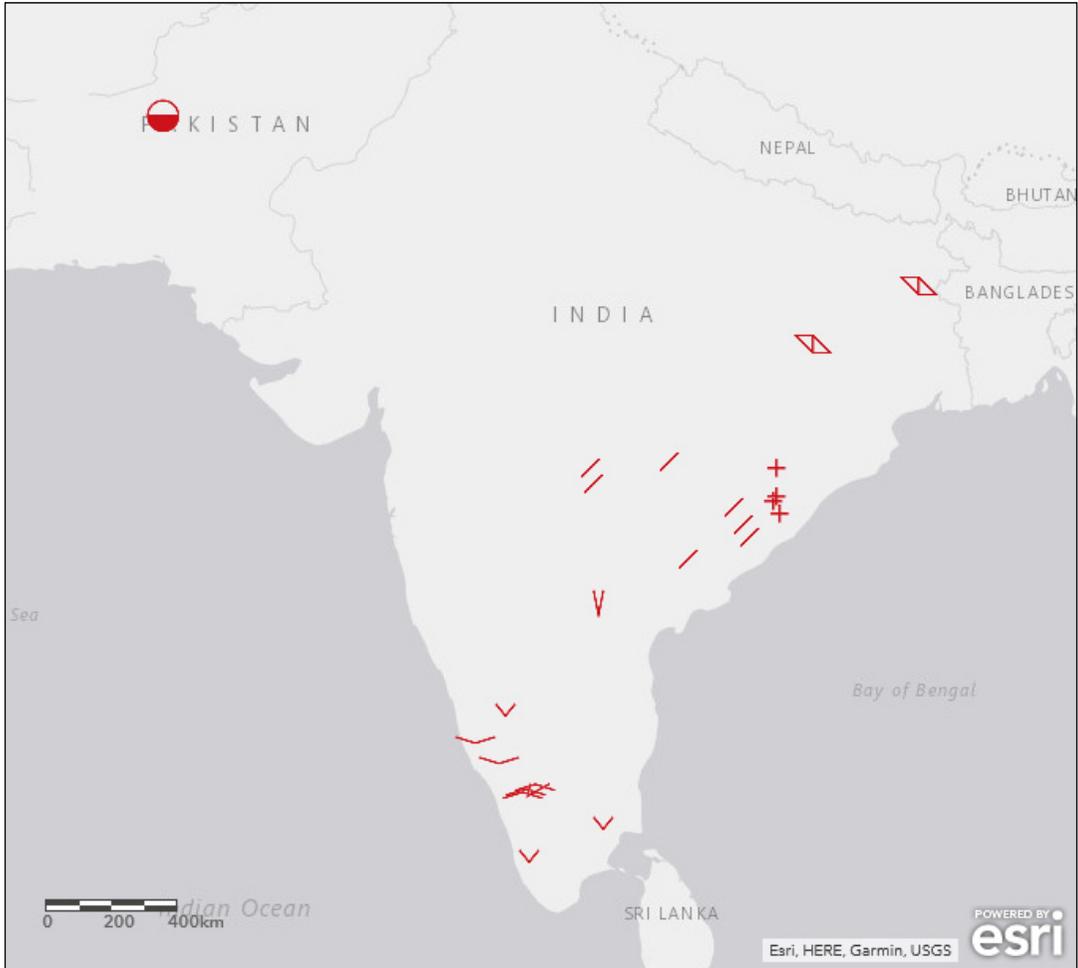
Figure 2.15.1: 'Fowl' in Dravidian.

'Chicken/hen/rooster' in Iranian

In the figure 2.16.1, there are four types of 'chicken'. The most common Type A derived from PIr. *kr̥ka-* 'chicken', which seems to be a onomatopoeic word.

Type B is found mainly in Western Iranian languages. It derived from PIr. *mr̥gá-* 'bird', which originally denoted '(wild) game animals' in Proto-Indo-Iranian. Type C and Type D are observed in Balochi and Sarykoli respectively, whose etymologies are unclear.

As mentioned above, there is a tendency that the words for 'hen' also represent 'chicken'. To avoid confusion, I excluded the languages in which the word for 'hen' is identical to 'chicken'. Type A derived from PIr. *māta-(ka)-* 'female'. It is dominant in Southern Pamir languages and Persian (As explained above, Persian also has word *murɣ* for 'hen, chicken'). Type B has the same Type as Type A in word for 'chicken', but it probably denotes only 'hen'. Type C

gwač, is restricted to Balochi, which is probably from an onomatopoeic word.

There are seven types of forms in the figure 2.16.3. The most dominant Type A and B are distributed mainly in the eastern areas. Type C is scattered in Munji and Eastern Gilaki.

Interestingly, Type D is in Ossetic and in Parachi, which are far away from each other. In Ossetic, it literally means 'who likes to tweet, singer'.

Type E is a loanword from Arabic ديك *[dik]* 'cock', which is found in Kurdish and Zazaki. Type F *pīng*, found only in Ormuri, seemingly experienced a semantic shift from 'dawn' to 'rooster' (cf. Persian بنگ *[pang]* 'dawn, morning'). Type G is distributed only in Northern Talysh, etymology of which is unclear.

(IWASAKI Takamasa)

A: *kerk* type   B: *murɣ* type   C: *tuxi* type   D: *čūrī* type

Figure 2.16.1: 'Chicken' in Iranian.

A: *mak* type   B: *kerk* type   C: *gwač* type
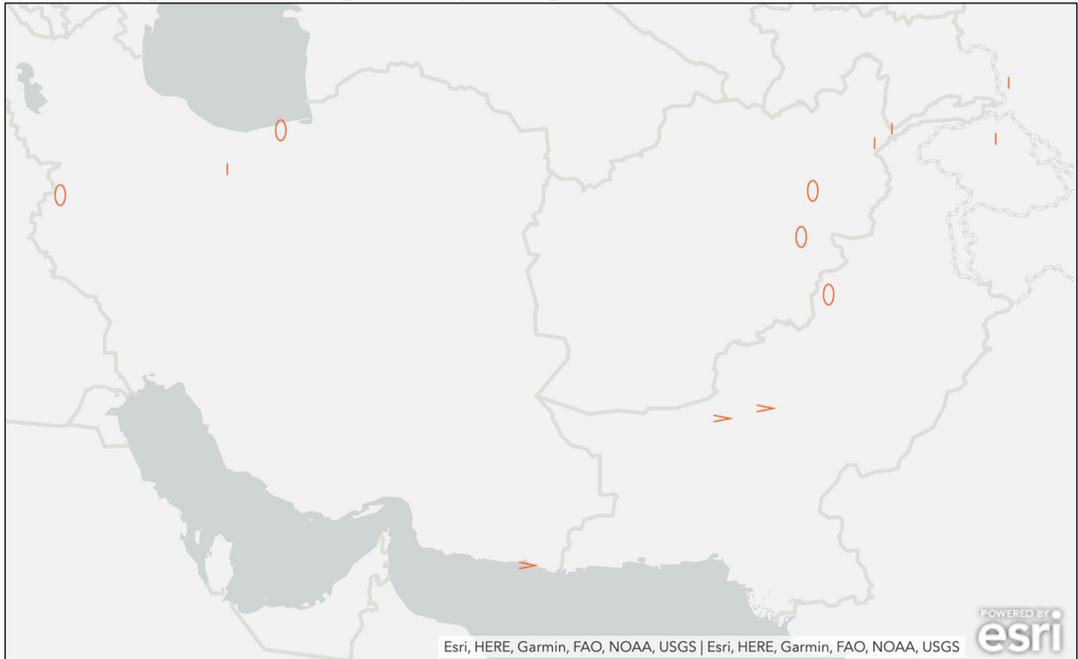
Figure 2.16.2: 'Hen' in Iranian.

A: *xurus* type   B: *kerk* type   C: *tela* type   D: *wasæg* type

E: *dīk* type   F: *pīng*   G *suk* type
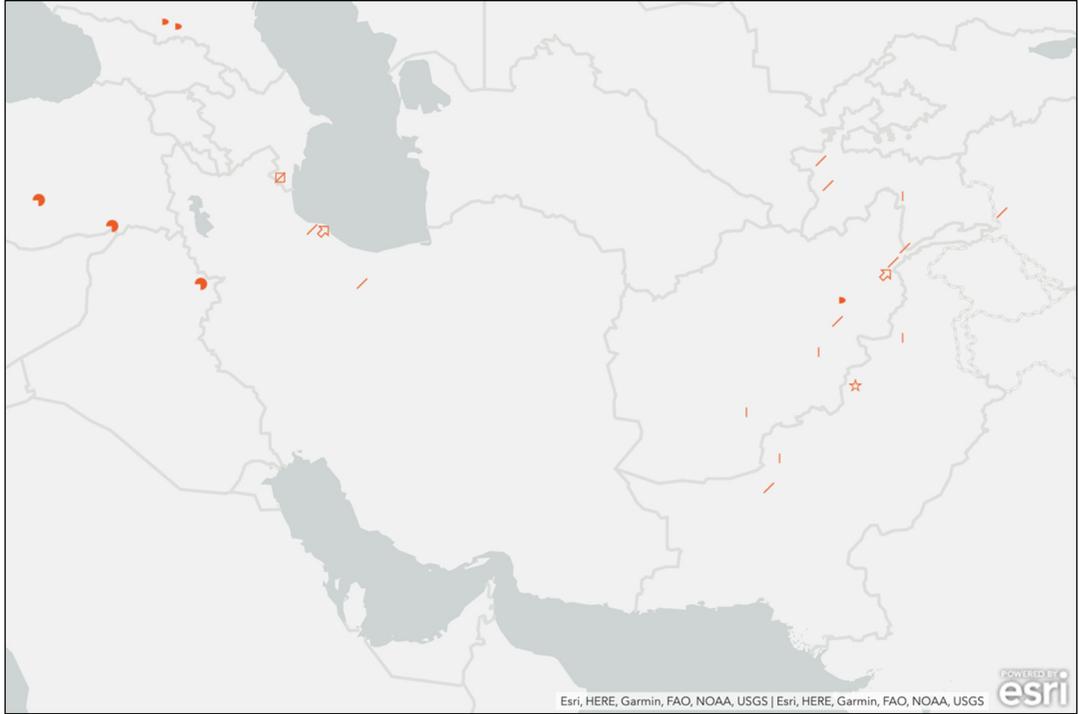
Figure 2.16.3: 'Rooster/cock' in Iranian.

'Cock' in Caucasian languages

Stems for 'chicken' in Caucasian languages distinguish 'cock' from 'hen'. The forms for 'cock' are described here.

Type A appears in Kartvelian languages—Bats, Tsez, and Bezhta—and is derived from the root meaning 'male'. Type B exists in Abkhaz, Abaza, and Nakh languages. Type C appears in the other Abkhazo-Adyghean languages. Type D is found in northern Dagestanian languages, and Type E in southern Dagestanian languages. Type F appears in Kryz and Budukh. Type G, an Azebaidjani loan, is found in Kryz and Khinalug.

Type A is mainly attested in Kartvelian languages and some Nakho-Dagestanian languages in contact with Kartvelian. Hence, we can assume Type A's distribution expanded from the Kartvelian-speaking area to its peripheral areas.

(SUZUKI Hiroyuki)

- A: m-type; *mamal-i*, *mamilu*, etc.
- B: r-type; *a-rbaɣʲ*, *borʁal*, etc.
- C: t-type; *ataqʼe*, *taqa*, etc.
- D: h-type; *heleku*, *haleko*, etc.
- E: d-type; *daʁa*, *dadal*, *datta*, etc.
- F: kp-type; *kpäl*, *kpəl*.
- G: *χuruz*.
- H: others; *kuntab*, *guluči*, *ažari*, *kek*, etc.
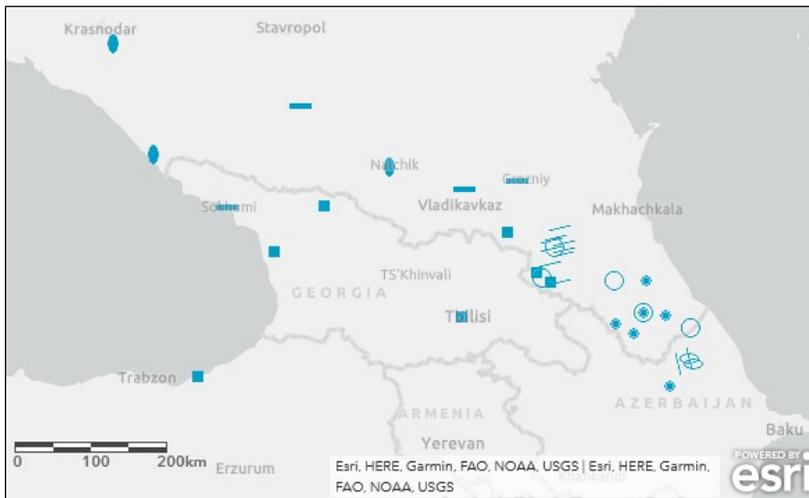
Figure 2.17.1: 'Cock' in Caucasian languages.

'Chicken' in Semitic

A. *dadʒaːdʒ* type (●) is Arabic form, sometimes with the feminine ending *-a, -e, -ih* indicating the unity: *didʒaːdʒ* (Iraq, Yemen); *didʒaːdʒih* (San'a of Yemen); *dʒaːʒe* (Lebanon, Syria); *ʒaːʒ(e)* (*dʒ > ʒ*) in Syria, Palestine, Jerusalem; *dʒadʒa* (Morocco); *dijaːj* (*dʒ > j*) in Gulf; *tidʒiːdʒa* (Maltese Ar.) with voiceless *t <*d* that we do not know why did *d* devoice.

   B. *dʒidaːd* type (○) where metathesis has occured is found in Arabic on the Sudan belt: *dʒidaːd* (Sudan Ar., Chadian Ar. Nubi Ar.); *gidída* (Nubi); *ɟidáda* (Juba Ar.) and some parts of Morocco and Algeria (Behnsted ed. 2011: 309. *Wortatlas*).

   C. *derho* type (▲) is in Ethiopic: *derho* (Tigre), *dərho* (Tigrinya ደርሆ), *doro* (Amharic ዶሮ).

   D. Other type:
Other Arabic forms are *farruːʒ* (◘ < 'chick', Tunisia), *firaːx* (✚<'poultry', Cairo), *ṭeːr* (★ 'bird', Uzbekistan).

Other types are *dəkkiít* (▼ Hobyot of South Arabian), *ksésa* (△ Gzira of Aramaic), *tarnególet* (■ Hebrew תַּרְנְגוֹלֶת).

                                    (NAGATO Youichi)
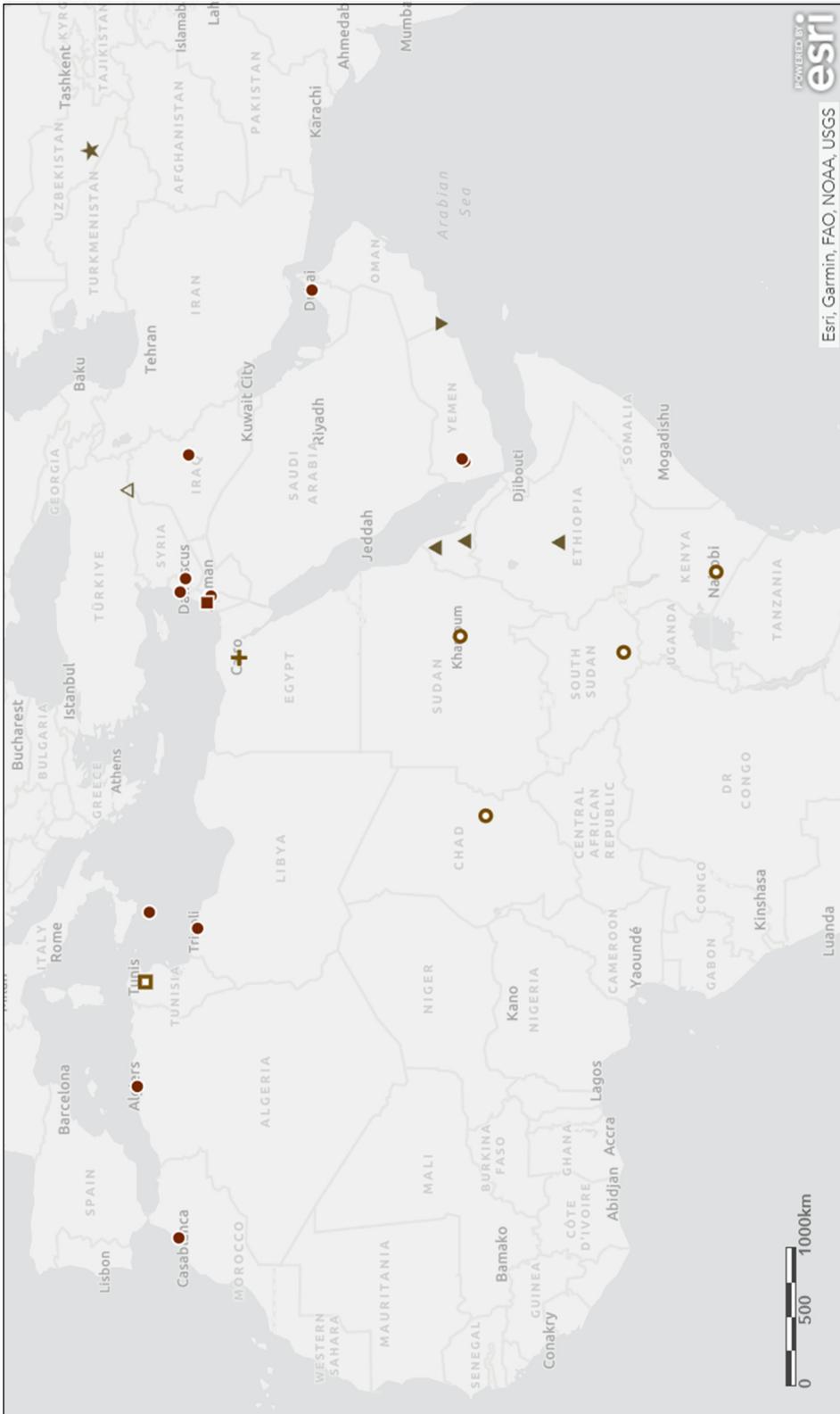
Figure 2.18.1: 'Chicken' in Semitic.

'Chicken' in Nilo-Saharan

What follow are heuristically reconstructed (marked with a hash #) Nilo-Saharan roots for 'chicken' (98 languages surveyed). We simplify diacritics and notations for the [±ATR] feature in the original data. For languages that do not (seem to) distinguish between 'chicken' and 'hen' (or 'fowl'), the term for 'hen' (or 'fowl') is used.

Types A1 #kokor and A2 #koko are the most widespread root across Nilo-Saharan, as evidenced in Daju (Dar Daju Daju *kukurge*), Temein (*kokorok*), Nilotic (Buraadiga *ŋqaqoora*, Dongotono *xoxoro*, Samburu *nkoko*, Masai *e-lukunku*), Surmic (Majang *koogele*), Saharan (Tudaga *kogoya*), Mabang (Maba *kerik*), Kuliak (Ik *ɲokokor*), Kadu (Krongo *kookoro*), Songhay (Zarma *gorŋo*) and less possibly Taman (Assangori *kormot*). These roots are quite widely attested in Africa across different families (Williamson 2000). It would remain questionable if this were due to language contact and diffusion or its onomatopoeic nature (or both).

Types B1 #kanda and B2 #kunza are mostly attested in Central Sudanic (Dongo *kanda*, Yulu *kaandə*, Gula *kunza*, Mbay *kənja* and perhaps Aja *ngbanda*) except in Darfur Daju (*kandane*). Boyeldieu (2000a) reconstructs two proto-forms *k-nd- (= B1) and *k/ng-Rnj- (= B2).

Type C #dirbad, attested only in Nubian (Nobiin *dirbad*, Dongola *durmade*), has been related with the Cushitic isogloss (Bilin *dirwa*, Somali *dooro*, borrowed into Ethio-Semitic, e.g., Amharic *doro*).

Type D #(n)gweno is found in Western Nilotic Lwo (Acoli *gweno*, Shilluk *gyeno*) and Central Sudanic (Bongo *ngono*, Baka *ngono*). Boyeldieu (2000a) reconstructs *ng-n- for the Central Sudanic isogloss. The other types (Type E) include Central Sudanic Aja *ngbanda*, which could be categorized with Type B1 or D.

(NAKAO Shuichiro)
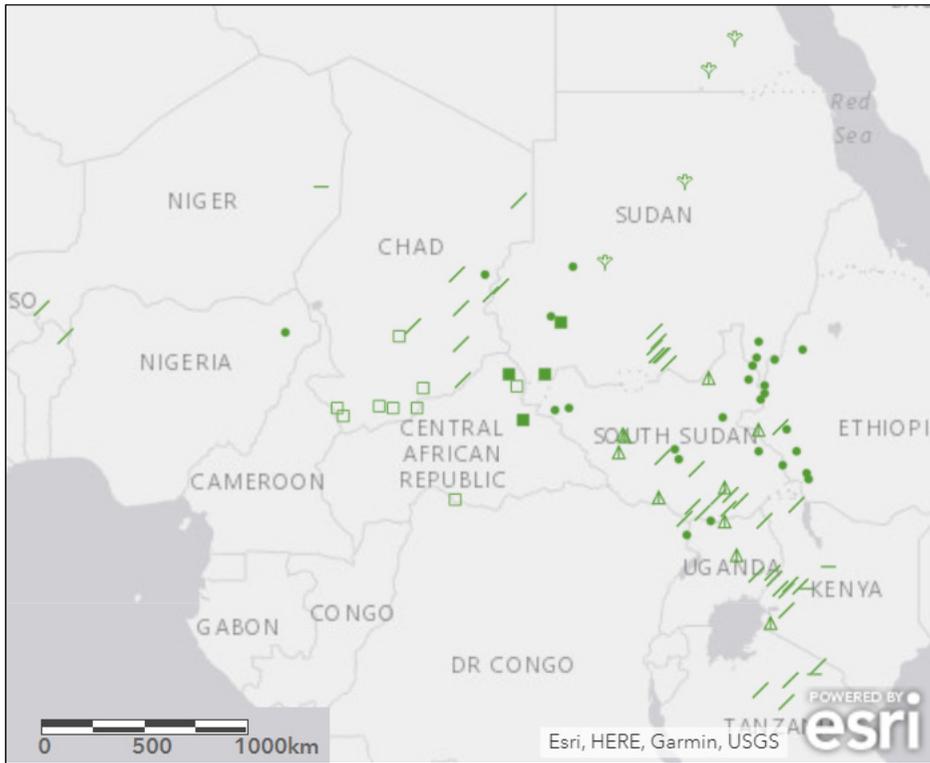
Figure 2.19.1: 'Chicken' in Nilo-Saharan.

✓ A1. *#kokor*
− A2. *#koko*
■ B1. *#kanda*
□ B2. *#kunza*
⚘ C. *#dirbad*
△ D. *#(n)gweno*
● E. The other types
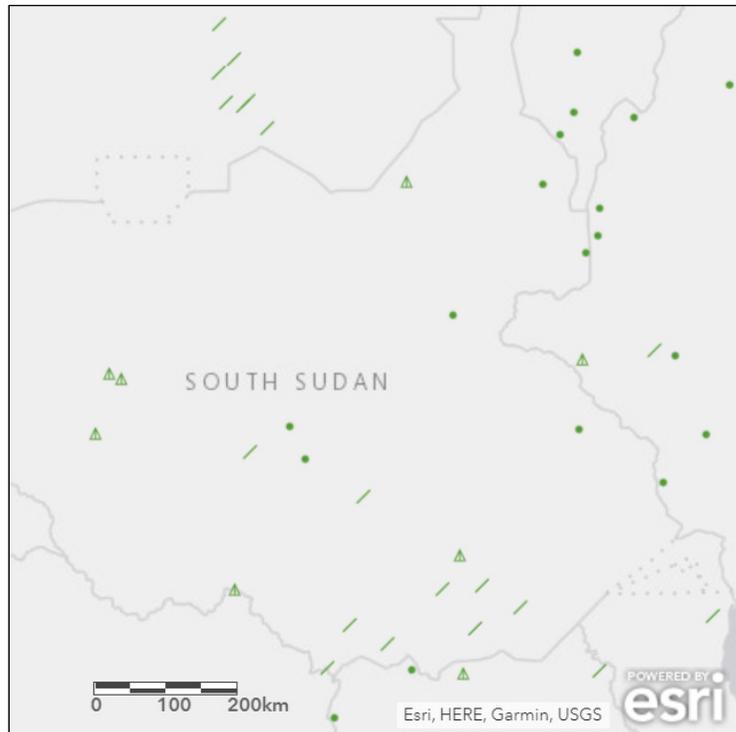
Figure 2.19.2: 'Chicken' in Nilo-Saharan around South Sudan.

'Chicken' in Bantu

In the Proto-Bantu lexicon reconstructed by Meeussen (1969), two distinct forms denoting 'chicken' are listed as a main entry. One is *-kókó (registered as [BLR-MAIN-1904] in the lexical database by Bastin *et al.* (2002)), which has a variant with the second-grade vowel *-kɔ́kɔ́ [BLR-VAR-2020], and the other is *-kúbà [BLR-MAIN-2105].

As shown in the map, forms directly relatable to *-kókó are scattered over northern and central areas including zones A, B, C, D, F, H, J, K, L, and M (for the latest 'Guthrie code' see Hammarström (2019)). However, the center of ditribution seems to be around the Lake Victoria, which is zone J, and the Savannah area especially in zone D. The examples include; Nzadi [B865] ŋkwɔ́, Enya [D14] ŋkɔkɔ, Rundi [JD62] inkoko, Kisu [E31b] eŋgoxo, Sukuma [F21] ngoko, Mbala [H41] koko, Bemba [M42] iŋkoko, Manda [N11] ng'oko, Matumbi [P13] ngɔkɔ.

In our database, which contains more than 150 data points collected from the list of corresponding forms of common lexical items (termed as 'Comparative series') in Guthrie (1967–71) as well as from 'The Tanzania Language Survey', which is a lexical database compiling data collected from Eastern Bantu languages mainly spoken in Tanzania based on the work done in 1970's by Derek Nurse and Gérard Philippson, it is clearly shown that the descendant forms from the variant *-kɔ́kɔ́ are more broadly distributed than *-kókó at least in Eastern Bantu area. The relatable forms are distributed in all zones except for North-Western zones A, B, and C, and the

interlacutrine zone J, where *-kókó is dominantly distributed. The examples include; Gikuyu [E51] ngʊkʊ, Rangi [F33] nkunku, Swahili [G42] kuku, Congo-Yoombi [H16c] visusu, Bungu [JD53] inguku, Luyana [K31] iŋuku, Malila [M24] inkuku, Ndengereko [P11] nguku, Yeyi [R41] uŋkuku, Manyika [S13a] xuku.

The other PB etymon *-kúbà, on the other hand, is exclusively distributed in zones A, B, and C, as in Londo [A11] kuɓa, Duala [A24] wuɓa, Basaá [A43] kóp, Bulu-Bene [A74] kup, Mvumbo [A81] pfuwo, Pol [A92] kuɓɛ, Seki [B21] ŋguɓɔ, Ngom [B22b] kfuɓa, Ngondi [C11] kuɸa.

The 'North-West vs. the rest' pattern of geographical distribution, which is clearly in this case, is of particular interest in light of the historical process of migration of the anscestors of people inheriting a language descended from a single origin, i.e, the Proto-Bantu language. As current studies show (e.g. Grollemund *et al.* 2015), the migration route of the Bantu speaking people, generally known as 'Bantu expansion', would have started at the homeland, which is estimated to be around Sanaga River Valley in Northern Cameroon (Watters 2018: 8) and the first major branching would have taken place between the (ancestral language of) North-Western group and the rest. The distribution pattern demonstrated by lexical forms denoting 'chicken' thus can be regarded as rather faithfully reflecting an early stage of the historical process of migration.

(SHINAGAWA Daisuke and KOMORI Junko)

∧ *-kókó   ○ *-kúba

▽ *-kókó   ψ others
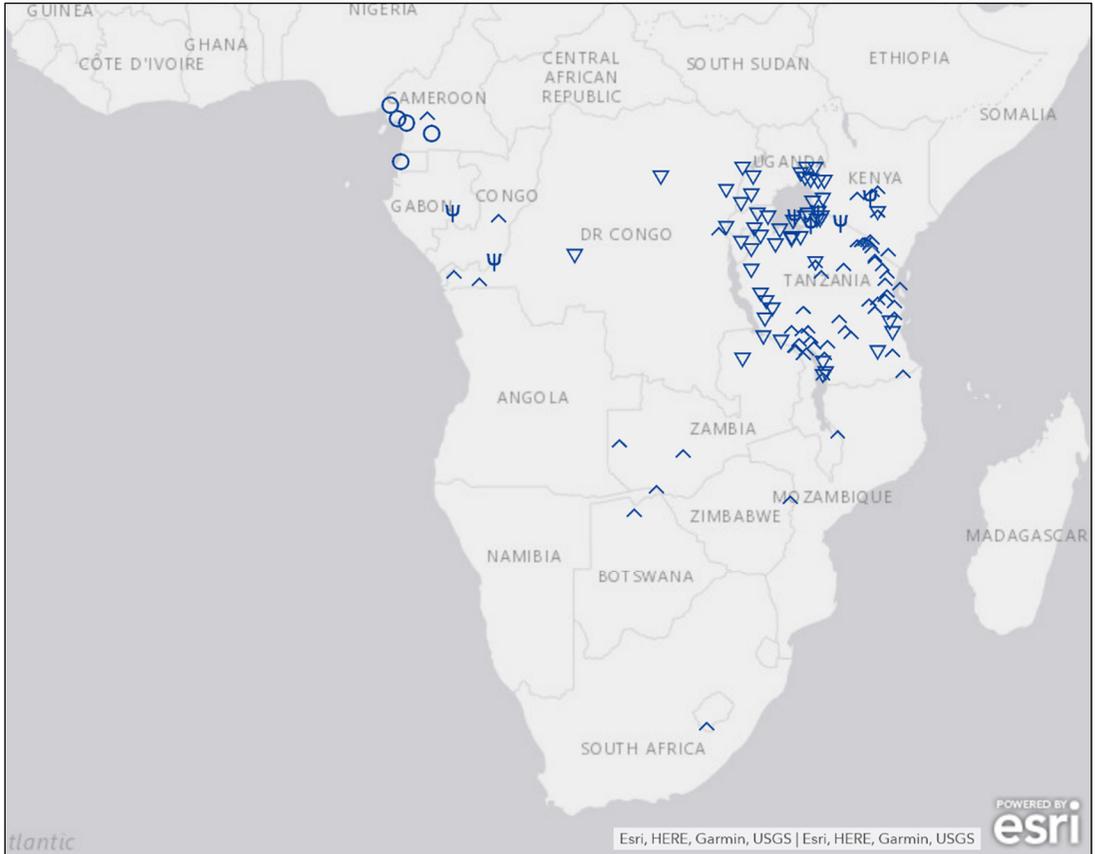
Figure 2.20.1: 'Chicken' in Bantu.
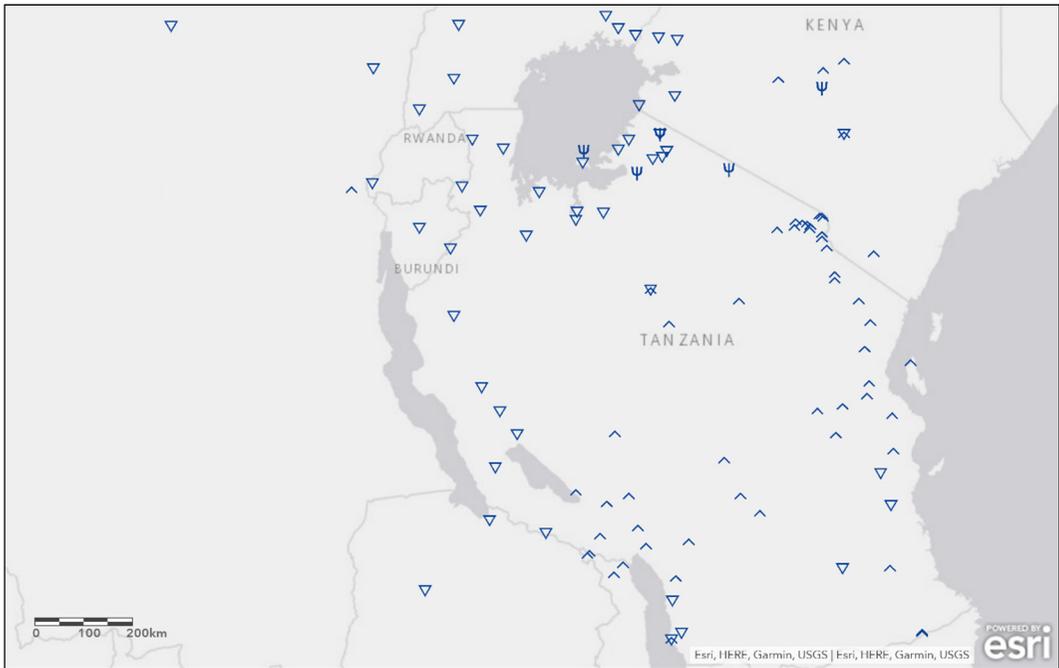
Figure 2.20.2: 'Chicken' in northeastern Bantu zones.

'Chicken' in the Kalahari Basin area

In the Kalahari Basin area (KBA), five types of the word "chicken" appear in 15 sample languages from three language families, Tuu, Kx'a and Khoe-Kwadi. Word types are determined based on the etymological origin of each word form; for example, *hunder* (A1) and *hǒŋgūrì* (A2) in Table 1 share the same origin, thus they are integrated into Type A. Note that three forms included in Type B (B1 *kúu-kúu* in Tuu, B2: *kúúkù* in Kx'a, and B3: *kooko* in Khoe-Kwadi) are borrowed from a Bantu language, presumably Setswana (cf. *koko* "a domestic fowl," Matumo 1993).

Unlike Type B, Types A, C, D, and E are distributed within one of the three language families, and each etymological word type is not distributed across the borders between language families; that is, word forms in

Type A are observed only within the Tuu family, Type C in the Kx'a family, and Types D and E in the Khoe-Kwadi family.

Table 1 Geographical variation of "chicken"

	Tuu	Kx'a	Khoe-Kwadi
A	A1: *hunder* A2: *hǒŋgūrì*		
B	B1: *kúu-kúu*	B2: *kúúkù*	B3: *kooko*
C		C: *kʰúú-kʰúú*	
D			D: *hūkú*
E			E1: *goro* E2: *ʔánǐ*

(KIMURA Kimihiko, NAKAGAWA Hirosi)
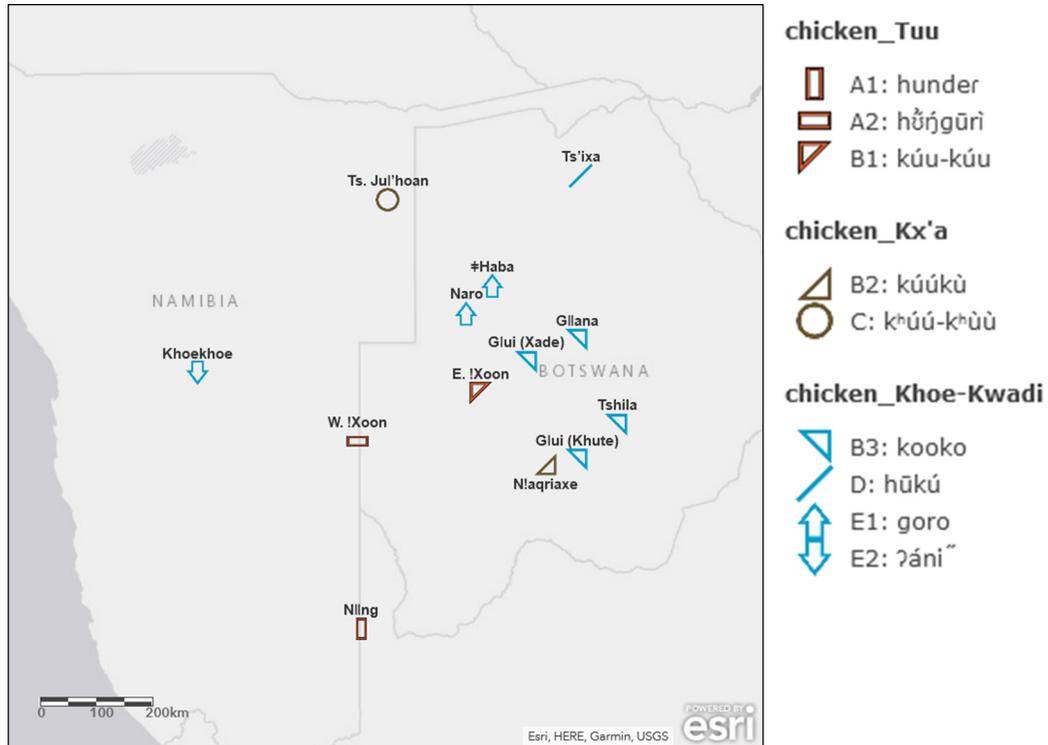
Figure 2.21.1: Geographical variations of "chicken" in KBA.

118

Chapter III

Horse

'Horse' in Asian and African languages

Horses (*Equus ferus caballus*) were domesticated in Central Asia before 3500 BCE (see Matossian 1997; Librado et al. 2021). The worldwide spread of horses was rapid, principally originating from their utilisation with vehicles and chariots.

Table 1 shows the principal forms in our data. No data are available from Chukotko-Kamchatkan languages or those of the Kalahari Basin area.

Table 1: Main word forms for 'horse'.

Languages	Word forms
Ainu	*úmma* (< Japanese *uma*) *cóme* (< Japanese *zyoome*)
Japonic	*uma* 馬
Korean	*mar*
Sinitic	*ma* 马
Hmong-Mien	*mra* *ma*
Tibeto-Burman	PTB **s/m-raŋ* (WrT *'brong*) PTB **r-ta* (WrT *rta*) IE **ghoḍā* PKar **k-sre^T*
Kra-Dai	*ma C* *ka* *ni* *nchau*
Austroasiatic	PMK: **mraŋ* PMK: **ʔseh* Proto Vietic: **m-ŋə:ʔ* Proto Kherwarian: **sadɔm*
Austronesian	hos-type (< English *horse*) kabalyu-type (< Spanish *caballo* or Portugese *cavalo*) ʤ..r...-type
Tungusic	MORI
Uralic	*hevonen* *ló* *val* *juno*
Mongolic	*mori*
Turkic	*at* *sïlgï*
Indo-Aryan	Proto-IA **ghōṭa* Proto-IA **Háśwas* *grast* (< Old Armenian)
Burushaski	*hayúr*
Dravidian	*kutirai*-type

	gurram-type
Iranian	PIr. **átsυa-* PIr **bāraka-*
Caucasian	cxen-type či-type k-typa g-type s/š-type
Semitic	*ħiṣa:n* *ʕawd* *zi:mel* *faras* *su:s*
Nilo-Saharan	*#murta* *#kaj-* *#sunda*
Bantu	*faras*-type (< Arabic) *ɲumbu* *nzoi*

The most striking feature of the word form for 'horse' is the widespread commonality of sounds, including an /m/-consonant, in its forms, which goes beyond language families, especially in eastern Eurasia, including in Ainu, Japonic, Korean, Sino-Tibetan (Sinitic and TB), Kra-Dai, Tungusic and Mongolic (Table 2).

Table 2: Word forms for 'horse' with /m/.

Languages	Word forms
Ainu	*úmma*
Japonic	*uma* 馬
Korean	*mar*
Sinitic	*ma* 马
Hmong-Mien	*mra* *ma*
Tibeto-Burman	PTB **s/m-raŋ* (WrT *'brong*)
Kra-Dai	*maC*
Austroasiatic	PMK: **mraŋ* Proto Vietic: **m-ŋə:ʔ*
Tungusic	MORI
Mongolic	*mori*
Nilo-Saharan	*#murta*

The word for 'horse' in Ainu appears to be a borrowing (from Japonic), as do the

words in Hmong-Mien (from either Sinitic or TB), TB (from Austroasiatic and Indo-Aryan), Austronesian (from various European languages), and other languages. Some Uralic forms are also borrowed, mostly from Germanic languages.

A noteworthy point here is the existence of languages using a non-/m/-initial native word for 'horse'; for example, WrT *rta* appears in Tibetic languages, spoken in the heart of Asia. In these languages, the word form derived from PTB *s/m-raŋ* is reserved for WrT 'brong 'wild yak'. This semantic change also deserves to receive attention when we consider the background of its distribution.

(SUZUKI Hiroyuki)

'Horse' in Ainu

All the terms for 'horse' (*Equus caballus*) in Ainu are borrowed words from either Japanese or Russian languages. Type A (the *úmma* type) originates from *uma* 馬 meaning 'horse' in Japanese. Type B (*cóme* [ʧome, tsome, ʤome, dzome] type) comes from *zyoome* 上馬 meaning 'good horse' in Japanese. The term *rosot* in Type C is borrowed from Russian Лошадь Loshad' meaning 'horse.'

Nakagawa (1989) discussed the discrepancy between the accents of words in Type A. The high pitch during pronunciation of the first syllable, e.g., *úmma* of A-1, was influenced by the Kinki dialect because many merchants from the Ōmi province (Ōmi shōnin) worked in Hokkaido in 1789, when the Matsumae Clan sent the first 20 horses to southern Hokkaido, Muroran. The term *umá* in A-4 in the Sōya dialect was influenced by the Tohoku dialect. In 1807, when horses were sent by the Tokugawa shogunate to Sōya, the Tsugaru and Nambu Clan in the northern province of Japan were working in Hokkaido under the direct control of the shogunate (for more details see Nakagawa 1989).

(FUKAZAWA Mika)

A. *úmma* type
- A-1. *úmma ~ únma~ unma*
- A-2. *úma ~ ḿma*
- A-3. *uuma*
- A-4. *umá*

B. *cóme* type
- B-1. *cóme*
- B-2. *cómen*

C. *rosot*
- *rosot*

Figure 3.2.1: 'Horse' in Ainu.

'Horse' in Japonic

As for the word form of the horse, the UMA types (UMA, MMA, M'MA, …) are distributed in both mainland Japan and the Ryukyu Islands. In addition, the NOUMA type is distributed in Miyako Islands, and DA(UMA) and DOODOO types can also be found in mainland Japan.

NOUMA (*nuuma*), meaning "wild horse" , found in Miyako Islands, contains UMA. There is an opinion that NOUMA comes from *nori-uma 'riding horse', but it is not adopted because the prosody of NOUMA (type C) does not correspond to *nori (type A) and the sound change of *noriuma > *nouma (> *nuuma*) is unnatural in Miyako languages. DA is the Chinese word for "駄", which means "to carry luggage", as it refers to "a horse that carries luggage". DOODOO is a shout when controlling a horse, which itself points to the horse.

In terms of horses, you can see KOMA and GANZYOO used for male horses (stallion), DA(UMA) and ZOOYAKU for mares, and TOONEN(KO) and TOOZAI(KO) for foals, but KOMA originally means "foal"; this became a general term for horses in Old Japanese and for male horses in several dialects. Although the etymology of GANZYOO is uncertain, it is thought that it is derived from Chinese words such as "五調". The mare's DA(UMA) means "駄" (to carry luggage), because the mare played many roles in carrying luggage. ZOOYAKU is a "miscellaneous role" and is derived from

the use of mares for various miscellaneous purposes and tasks for male horses used for riding and military purposes. The foals TOONEN(KO) and TOOZAI(KO) mean "当年・当歳" (current year) and mainly refer to foals born in that year. These are analytical expressions based on usage, etc., and there is no doubt that UMA is the oldest type for "horse" in Japonic languages. The Japonic form of horse in Chinese zodiacs is UMA, even in Miyako, where NOUMA is used for "horse". There are some expressions involving hair color, such as *kage* 'deer's hair', but these expressions are few.

UMA is thought to have been borrowed from forms found in East Asian languages, such as Chinese *mǎ*, Mongolian *mori*, and Korean *mar*. It seems to have been borrowed directly from the Old Chinese (modern form is *mǎ*), but in Japonic languages, there is an epenthetic vowel /u/ at the beginning of word. As the vowel epenthesis is also found in *ume* 'plum' (cf. modern Chinese *méi*), it seems that the Old Chinese /m-/ sounded like /um-/ to Japonic speakers. Some dialectal forms beginning with /mma-/ seem to be old because they do not have epenthetic vowels, but some forms of Ryukyuan languages begin with the glottalized [ʔmma-], which indicates that it started with a vowel (*uma- [ʔuma-]).

(NAKAZAWA Kohei and YOKOYAMA Akiko)

I	UMA	O	MAKO	=	M'MA
0	UMAKO	⌒	NOUMA	≡	MMMA
–	MMA	⌃	DA(UMA)	/	MA
⊖	MMAKO			⬝	M'A
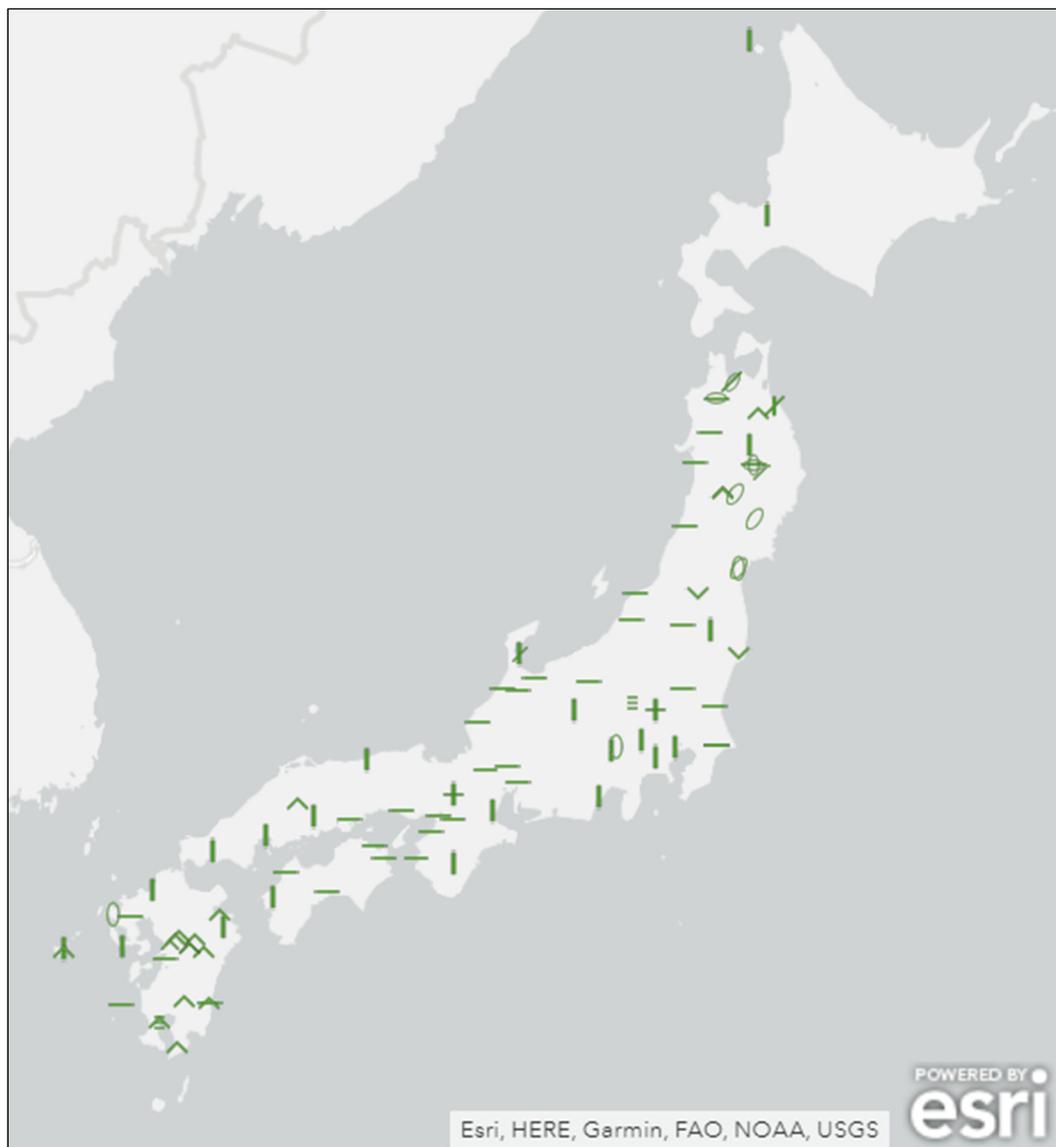
Figure 3.3.1: 'Horse' in mainland Japan.

Figure 3.3.2: 'Horse' in Northern Ryukyu Islands.

Figure 3.3.3: 'Horse' in Southern Ryukyu Islands.

KOMA

KO[UMA]

WO[UMA]

WOSU[UMA]

OTOKO[UMA]

WEKE[UMA]

WONTA[UMA]

GANZYOO

ZOOYAKU

KAGE

Figure 3.3.4: 'Stallion' in mainland Japan.
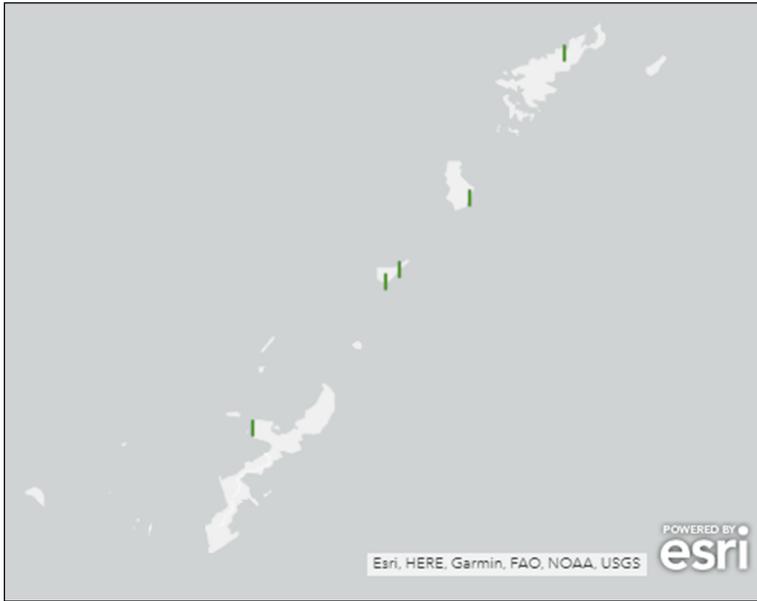
Figure 3.3.5: 'Stallion' in Ryukyu Islands.

Figure 3.3.6: 'Stallion' in Ryukyu Islands.

◇ DA(UMA)   ═ WONNA[UMA]   ⊘ MENTA[UMA]
⌒ ZOOYAKU   ⌣ HIMBA   ⊖ METTA
▮ ME[UMA]   ⫴ MENTA[UMA]   И NR
▮ MESU[UMA] ⌃ GANZYOO
▬ WONAGO[UMA] 0 MEMBA
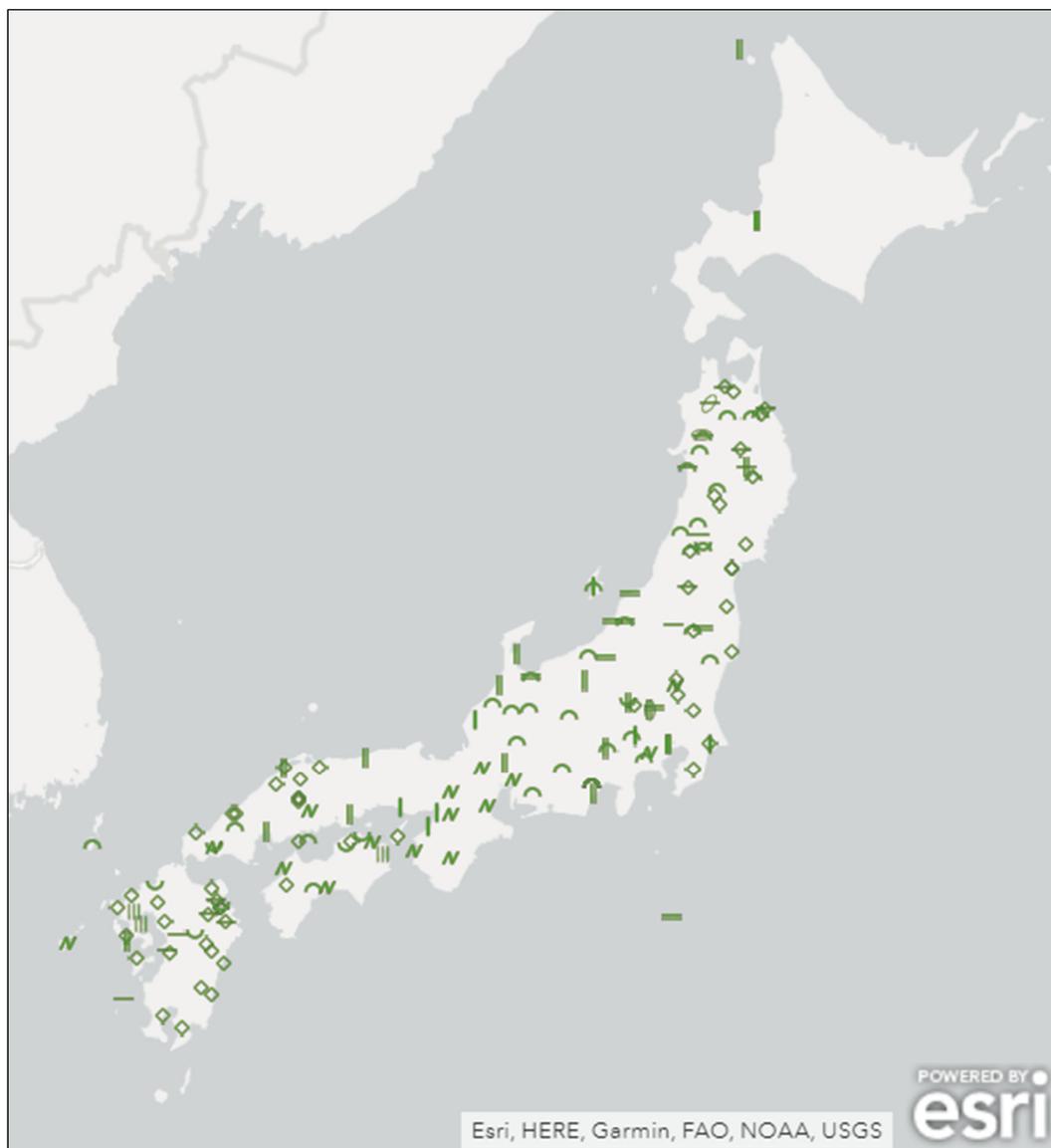
Figure 3.3.7: 'Mare' in mainland Japan.

Figure 3.3.8: 'Mare' in Northern Ryukyu Islands.
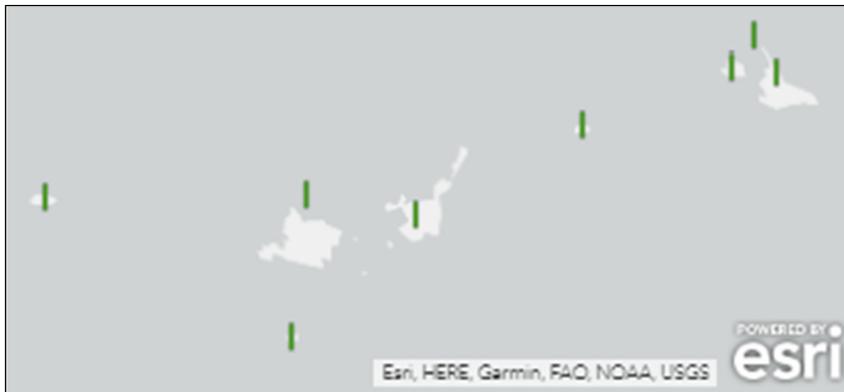
Figure 3.3.9: 'Mare' in Southern Ryukyu Islands.

KO[UMA]   — TOONEN(KO)   ∩ BEKO
KOMA   = TONKO   ⌣ MEKO
KO[UMA]KO   ☰ TONKOME   ⋏ NR
[UMA]KO   ┃ TOOZAI(KO)

Figure 3.3.10: 'Foal' in mainland Japan.

Figure 3.3.11: 'Foal' in Northern Ryukyu Islands.

Figure 3.3.12: 'Foal' in Southern Ryukyu Islands.

'Horse' in Korean

Modern standard form for 'horse' is 'mar [mal]' and Middle Korean form 'mʌr [mʌl]'. Middle Korean is the language spoken from the middle of the 15th century to the end of the 16th century. These two forms are almost the same except for the vowel. The vowel 'ʌ' has lost its phonemic status and merged with the vowel 'a' in many dialects except for the Cheju dialect. In some southern and northern dialects the vowel in question appears as 'o'.

The MK form 'mʌr' has been sometimes considered as a loan word from Manchu or Mongolian 'morin'.

Dialect variation is not so great. We have only a few phonetic varieties such as the following:

A-1 mʌr, A-2 mar, A-3 mor

All these forms derived from the same native etymon shown above.

(FUKUI Rei)
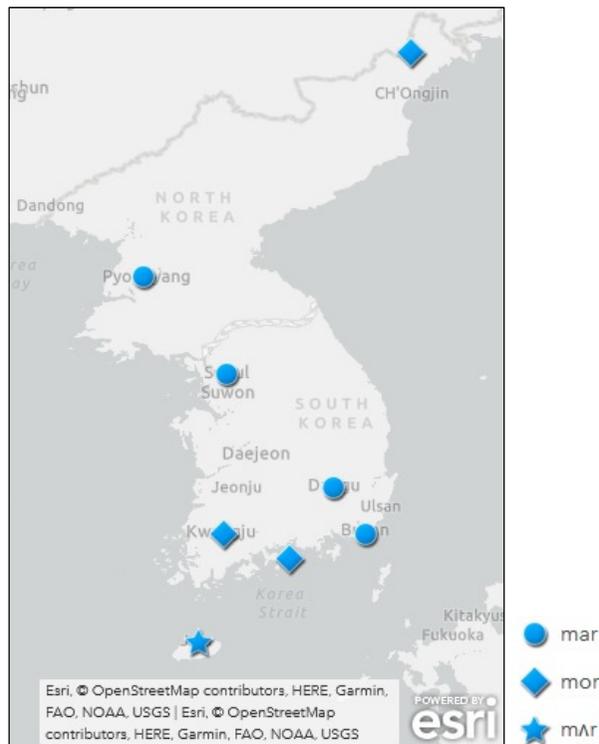
Figure 3.4.1: 'Horse' in Korean.

'Horse' in Sinitic

All of the forms denoting horse in Sinitic have the same stem 马. We classify them based on types of suffix.

A1: 马　　ma (北京)

A2: 马儿　mar (离石)

A3: 马子　ma tɕʅ (诸城)

A4: 马牯　ma ku (新化)

　　Large part of forms denoting horse in Sinitic are monosyllabic 马 (A1 type). Some dialects have forms with suffix, however is very rare.

Reconstructed forms of 马 of Middle Chinese and Old Chinese are shown below.

	1	2	3	4
MC	ma	ma:	mɑeX	-
OC	meɑ	mɔ	*mˤrə	*mrăʔ

1:郭錫良(2010), 2: Karlgren (1957[1997]), 3: Baxter & Sagart (2014), 4: Schuessler (2007)

(YAGI Kenji)

A.　type

　　○　A-1　马ma

　⊕　A-2　马儿mar

▲ A-3　马子ma tɕʅ

△ A-4　马牯ma ku

Figure 3.5.1: 'Horse' in Sinitic.

'Horse' in Hmong-Mien

There are three types in HORSE: A: *mra*; B: *ma*; C: *ljei hɔŋ*.

Both Type A and Type B are loanwords from the external source(s): Sinitic or Tibeto-Burman. Type A is observed in Hmongic, and Type B is in Mienic. Type A and Type B could be ultimately related in Proto-Hmong-Mien. If so, we would have a uniform distribution almost throughout the Hmong-Mien area.

(TAGUCHI Yoshihisa)

Figure 3.6.1: 'Horse' in Hmong-Mien.

'Horse' in Kra-Dai

The most widespread form is type A1 *ma C*, obviously a borrowed word from the Sinitic *ma B*. As well known, Kra-Dai tone C corresponds to the Sinitic tone B. This type is distributed among the Tai, Dong-Shui, and Ong-Be branches in Hainan. Type A2 pə[5] is found in the Li language in Hainan. Moroever, it is possible that this is an isolated word without any relationship to the other forms. Still, it is tentatively classified as a denasalized variety of A1.

Type B has two subtypes with a nasal or denasalized initial consonant. The nasal variety is distributed both in the Kra branch in the continent and in the Li branch on Hainan Island; hence, it should be an old indigenous form in Kra-Dai.

The other types C to I are different from one another and scattered sporadically; therefore, it is difficult to infer their origins and the formation process.

(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)
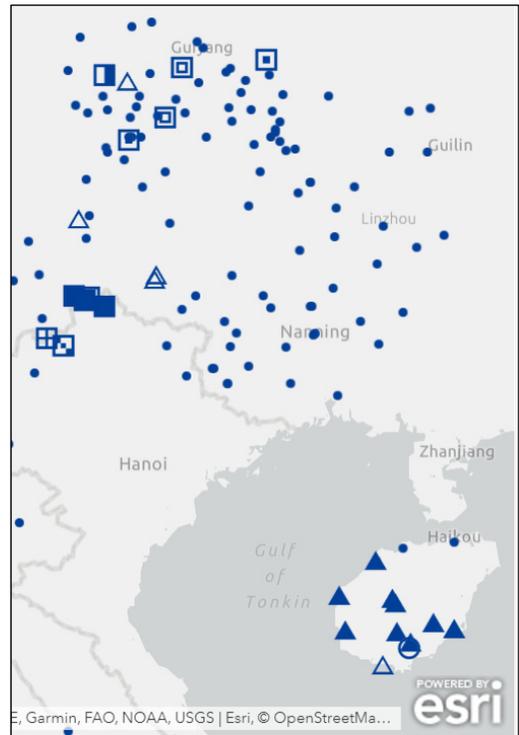
Figure 3.7.1: 'Horse' in Kra-Dai (enlarged).

A. *ma C* type
 • A1: ma:^{C2}, ma^2, ma^{23}, ma^3, ma^4, mā^4, mɒ^4, ma^5, ma^6, maa^3, maa^4, maa^5, maa^6, maʔ^8, mja^4, mjɛ^4, tə ma^4, tə^{0}ma^4
 ○ A2: pə^5
 B. *ka* type
 ▲ B1: ka^3, kɔ^4, kha^3
 △ B2: ŋa^{11}, ŋa^3, ŋou^{55}, ma^{0}ŋa^{11}, ma^{0}ŋa^{32}
 C. *ni* type
 ■ ni^{334}, ɗɔ^{33}ni^{33}, ɗɔ^{33}nĩ^{33}
 D. *ntchau* type

 ◉ ɳʈɕau^{13}, ɳʈɕau^{21}
 E. *nho* type
 ◙ ŋo^{35}, ŋo^{53}
 F. *linn* type
 ⊞ lin^{53}ŋ̩^{44}
 G. *she* type
 ⊠ ʂe^{33}
 H. *u* type
 ▣ ʔu^{45}
 I. *vunhung* type
 ◫ vu^{0}ŋuŋ^{55}

Figure 3.7.2: 'Horse' in Kra-Dai.

'Horse' in Tibeto-Burman

There are four major stems (word roots) for 'horse' in TB. These stems include not only etyma of the proto-level forms of Proto-Tibeto-Burman (PTB) and Proto-Karen (PKar; Luangthongkum 2019), but also loans from neighboring major languages. The TB word forms for 'horse' contain word formations, which consist of a single stem, a stem plus an affix, or compounds of two stems. We classify them first into stem types and then into compound types.

Type A is derived from *s/m-raŋ (HORSE) in the PTB etymon. This root is cognate with Chinese 馬 (OC *ma) (STEDT). Type B is derived from *r-ta (HORSE) in the PTB etymon. Type C is borrowed from Indo-Aryan *ghoḍā (HORSE). Type D is from *k-sre^T (HORSE) in the PKar etymon. This etymon is an Austroasiatic loanword (Luangthongkum 2019: xxviii).

In addition to the four major types, there are some marginal roots, which are labelled as Type X. All of them are etymologically unknown. The word form *lalo* is used as a general term for 'horse', whereas *rta* (Type B) refers exclusively to 'male horse' in some dialects of Amdo Tibetan.

Most of the above-mentioned forms contain a single stem, but we also found one type of a compound form of the C-compound type (*sya-n*+C). The etymology of *sya-n* is *sya-n* ⚔ *sin* (FLESH/ANIMAL/BODY) in the PTB etymon. In some languages and dialects, two word forms coexist: A and X (*alumu* and *lalo*).

Type A is the most widespread across the branches of TB. This type is found in the central-eastern part (rGyalrongic and Qiangic), at the eastern edge (Tujia), in the southern part (Lolo-Burmese), and in the central-western part (Sal, Kuki-Chin, and Deng) of the TB area.

Type B is found in Tibetic and its neighboring languages (Basum and Nung), and in northern Nepal (Tamang, Manang, Thakali, and Gandaki Gurung).

Type C is found in north-eastern India (Central Naga, Angami-Pochuri, Sal, and Tani) and central Nepal (Ghachok Gurung). The C-compound type is found in north-eastern India (Tangkhulic, Meithei, and Kuki-Chin). Type D is found in southern Burma (Karenic) only.

From a chronological perspective, Type A is the oldest because it is widely dispersed from a geographic location and is found in broader language groups. Type B is recognized as one of the characteristic word forms that distinguish Tibetic from the other Himalayish languages (Takeuchi 2021). This point might be related to the fact that horses played important historical roles in trading and carrying official documents in the era of the Tibetan Empire (from the 7th century to the middle of the 9th century), which governed a vast extent of land. This is supported by the fact that Type B is used in languages neighboring Tibetic. The geographical distribution of Type C (a loan word from IA) is also considered to be closely tied to transportation by horses.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

○ A. *s/m-raŋ* type

nbzaŋ, m̥zaŋ^{31}, ⁿbrâ, wə, ɹu, raɴ, ʐu, ɹəq, 'jy, ʁò, guẽ^{13}, buro, byìn, nyang, mɳo^{35}, man, myiɴ^H, sɔrə, mbo ro, mbro, vre, ɦre, etc.; mo^{21}ku^{55}lu^{55}, mu^{21}pɑ^{55}, mɔ^{21}pha^{21}, mo^{44}ŋ^{21}, mjo^{44}tha^{42}, etc. (suffixed); a^{55}m̩(ɰ)^{21}, a^{21}bɔŋ^{21}, a^{55}mo^{31}, i^{35}mv^{53}, gum^{31}raŋ^{31}, gim rang, pɑ^{31}xoŋ^{35} (prefixed); ɑ^{55}mo^{31}zɑ^{31}, etc. (affixed).

╱ B. *r-ta* type

l̥ta, rta, ʂta, ta, ʰta, ʰto, te^{53}, da^1, etc.

□ C. *ghoḍā* type

(N)gohḍaːq, guree, gura, gum^{4}raaŋ^1, kierü, keru, kuri, kor gɯ-ri, etc; sə-kɯ, etc. (prefixed).

⊞ C-compound type

sya-n +C: sa gol, sa kawr, sa-koi, sā kòl, si-kwe.

△ D. *k-sre^T* type

kɤ̃ θe^3, kə^{11}se^{45}, se^{31}, θè, ngshe, etc.

⌄ X. others

lalo, copi, sʰəpùʔ mok, pferi, ɦgɤ̃, etc.
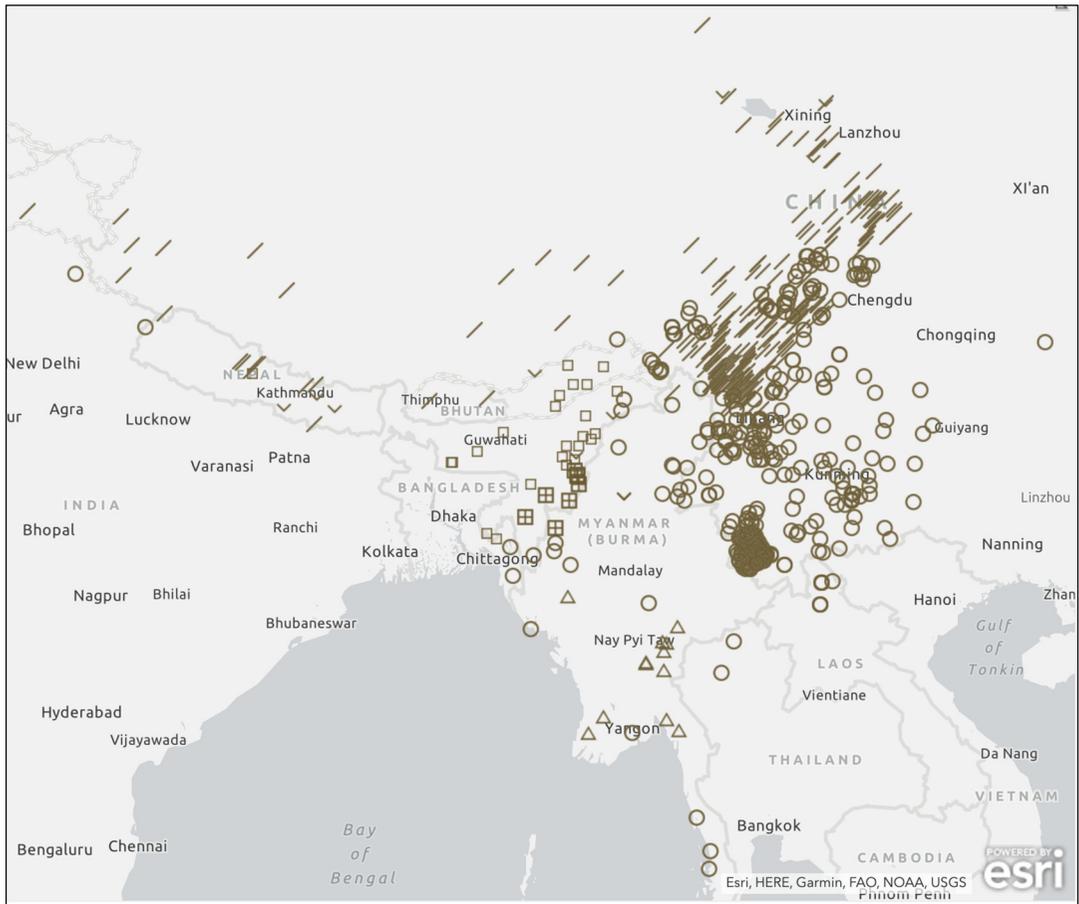
Figure 3.8.1: 'Horse' in Tibeto-Burman.
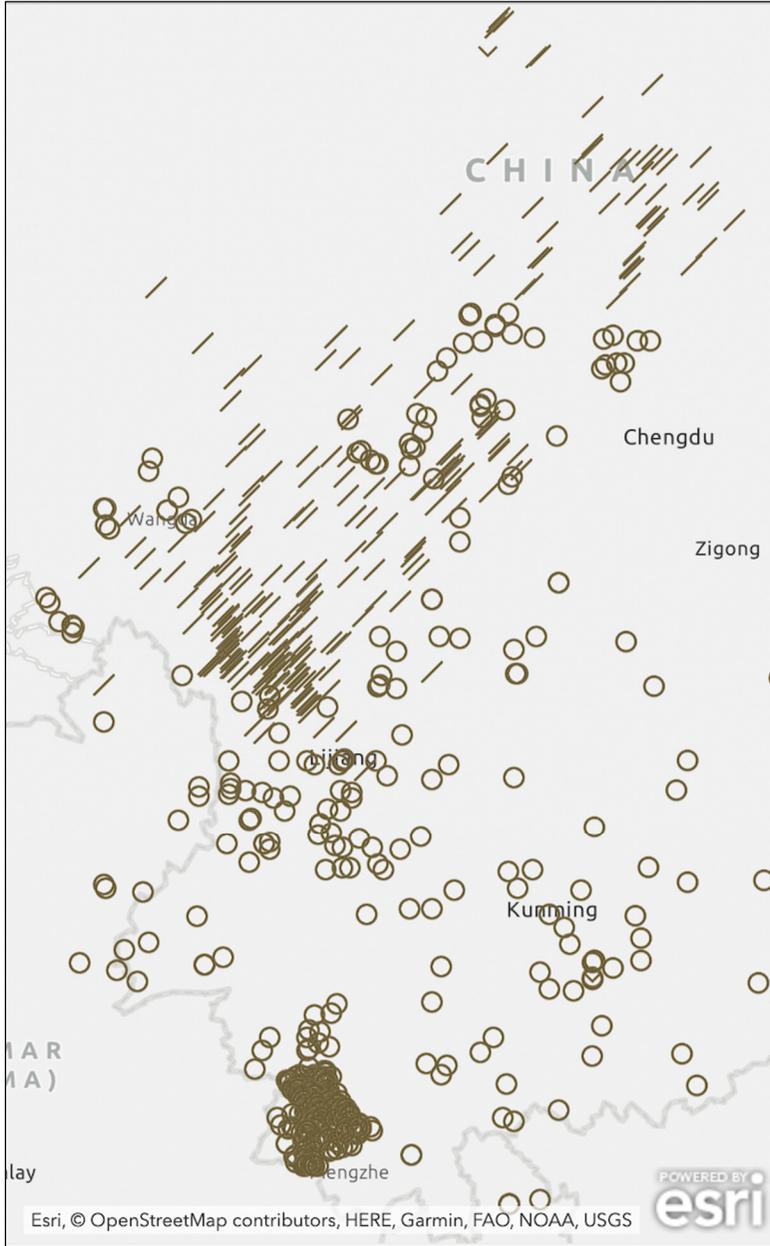
Figure 3.8.2: 'Horse' in Tibeto-Burman (detailed).

'Horse' in Austroasiatic

The word forms meaning "horse" in Austroasiatic are classified into eight types, as follows.

A. *mraŋ type*
Proto MK: *mraŋ (Shorto 2006)

Proto Khmuic: *hmraŋ (Sidwell 2013); hmbraŋ (Khmu [Cuang]), hmraŋ (Khmu)

Proto Palaungic: *mraŋ (Sidwell 2010); maruang (Wa), mbraŋ (Lawa [Bo Luang])

B. *ʔseh type*
Proto MK: *ʔseh (Shorto 2006)

Proto Bahnaric: *ʔəsɛh (Sidwell 2011); asɛh (Bahnar), cheh (Brao), sɛh (Alak)

Proto Katuic: *ʔasɛh (Sidwell 2005); (ʔa)sɛh (Kui), sɛh (Ngeq), ʔasɛh (Bru)

Khmeric: seh (Khmer)

Monic: cheh (Mon)

Pearic: sɛh (Chong)

C. *m-ŋəːʔ type*
Proto Vietic: *m-ŋəːʔ (Ferlus 2007); ngựa (Vietnamese [Hanoi]), ŋɨa[4] (Chứt [Rục])

D. *sadɔm type*
Munda: *sadɔm (Proto Kherwarian: Munda 1968); sadom (Santali [Singhbhum])

E. *kulaːj type*
Khasic: kulaːj (Khasi)

F. *ljiŋ[33] type*
Mangic: ljiŋ[33] (Mang)

G. *koːɾa type*
Nicobarese: koːɾa (Car)

Munda: gora (Bondo), ghoɽa (Juang), ghuɽgi (Korku)

Aslian: kudaʔ (Jahai)

H. *ma: type*
Loan from Chinese: ma: (Phong), mậː (Thavung)

Shorto (2006) reconstructed two forms as proto Mon-Khmer forms for "horse": *mraŋ (A) and *ʔseh (B). The B type *ʔseh*, which is unique to Austroasiatic languages, is distributed in the central and southern area among Bahnaric, Katuic, Khmeric, Monic and Pearic.

The A type *mraŋ*, which is distributed among Khmuic and Palaungic in the northwest, is shared with Tibeto-Burman *s/m-raŋ (Ebihara et al. 2022).

The H type *ma:* distributed in the northeast is obviously borrowed from Chinese.

It is noteworthy that the G type *koːɾa* is widely distributed across the Bay of Bengal and the Andaman Sea among Munda, Nicobarese and Aslian, and that it is shared with the Austronesian form /kuda/ (Utsumi 2021). Kodama (2021) pointed out that the Dravidian forms gōḍā, kōɽā, and ghoɽo are borrowed from Indo-Aryan, probably from which the G type in Austroasiatic originated.

(SHIMIZU Masaaki, MINEGISHI Makoto)
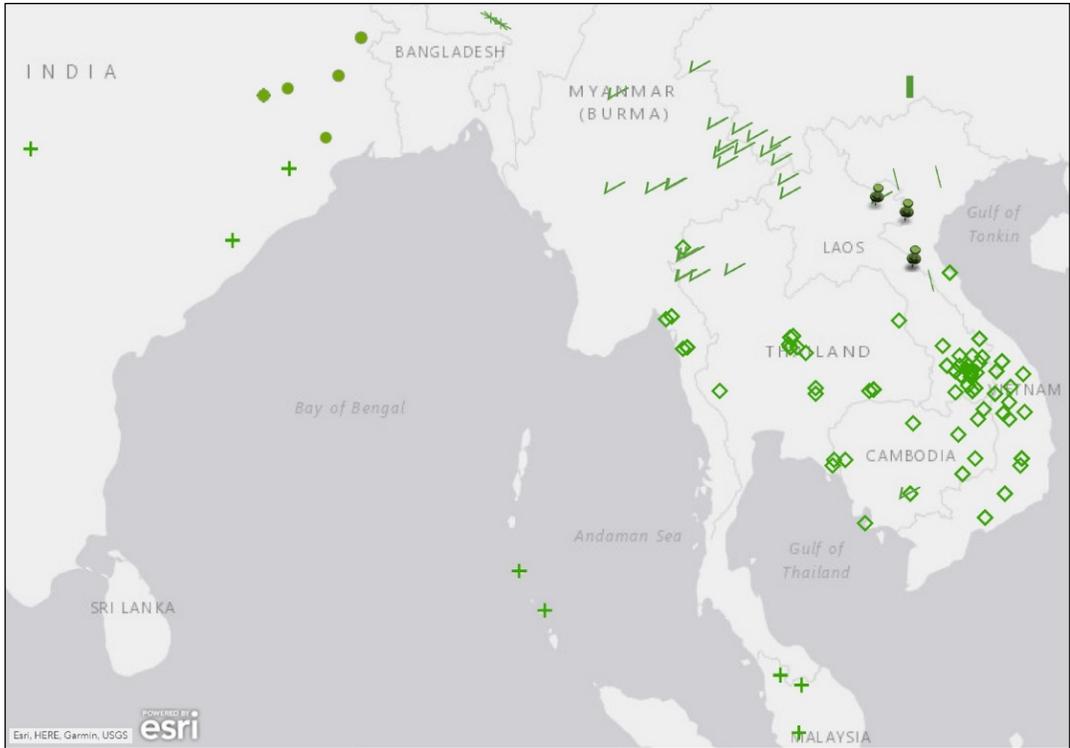
Figure 3.9.1: ‘Horse’ in Austroasiatic.

↙	A	*mraŋ type*
◇	B	*ʔseh type*
\	C	*m-ŋəːʔ type*
●	D	*sadɔm type*
✳	E	*kulaːj type*
∣	F	*ljiŋ³³ type*
+	G	*koːra type*
🐴	H	*maː type*

'Horse' in Austronesian

It appears that the word "horse" is not indigenous to the areas where Austronesian languages are spoken, especially in islands in the Pacific. The form most often found is one that seems to have been derived from English loan word "horse," which is grouped as a type A form. The word forms that belong to type A have word-initial /h/ and/or /s/ sound in the final syllable. Examples of type A include *hos* (Nyindrou, Kaulong, and North Tanna), *hōs* (Paamese), *hɔsu* (Rotuman), *hoosi* (Tongan), *hosi* (Tawala, Motu, Lau, and Negnone), *hose* (Maringe), *os* (Lewo), *osi* (Manam), *ōsi* (Xârâcùù), *ose* (Eastern Fijian), *ohe* (Western Fijian), *ɔsi* (Yabem), *wos* (Takia), *wosa* (Kilivia), *ɣōs* (Buang), *hōɲ* (Nemi), *yōs* (Woleaian), *wōse* (Mele-Fila), and *hoi* (Rapanui). These forms are predominantly found in Papua, Solomon Islands, and small islands that are located in the Pacific.

Type B forms are supposedly derived from the Spanish or Portuguese word for "horse" (either caballo or cavalo) and predominantly have word-initial /k/ or /ʔ/ and word-internal /b/. Examples are *kabalyu* (Isnag), *kabāyu* (Kalinga Limos), *kabāyo* (Tagalog), *kabāyuʔ* (Aklanon), *kobayu* (Palawan), *kabayu* (Kagayanen), *kabaro* (Bantik), and *ʔaballo* (Talaud). Type B spreads in some areas in the Philippines and Sulawesi.

Type C and D do not seem to be loan forms. Type C forms are found in the Sulawesi and Java islands, as well as the lesser Sunda islands. They typically have a word-internal alveolar sound, such as /r/ or /d/ and are disyllabic: *dʒaran* (Javanese), Balinese, and Sasak), *dʒʰaran* (Madurese), *dʒara* (Da'a and Ngada), *dʒaraʔ* (Uma), *aɲɲaraŋ* (Bugis), *dʒaraŋ* (Konjo and Sika), and *adʒara* (Wolio).

Type D forms are found in Western Java and Sumatra, and have word-initial /k/, /g/, or /h/, and word-internal /d/, a typical form being /kuda/: *kudaʔ* (Molbog), *guda* (Acheh), *hoda* (Batak Toba), *kudo* (Minangkabau), *kuda* (Indonesian, Sundanese, Buru, and Sawai).

Type E is also considered to be a loan word from the French word for horse, "cheval." This type is found in *suavali* (Malagasy Merina) and *yovaɾi* (A'jië).

Many other forms do not have common features and are grouped together as type F. Examples are: *ramaiʔ* (Atayal), *ʔua sa-sm-ovri* (Tsou), *rigi* (Rukai), *baka* (Yami), *bulmaot* (Tolai), *háricān* (Cémuhî), *solofanua* (Samoan), and *puʔarehenua* (Tahitian). *Uma* in Paiwan looks like a loan word from the Japanese word /uma/, but there is no other form similar to this, so it is included in type F. They are found in Taiwan, Solomon Islands, and islands in the Pacific.

(UTSUMI Atsuko)

Ⴘ A: *hos, hōs, hɔsu, hoosi, hosi, hose, os, osi, ōsi, ose, ohe, ɔsi, wos, wosa, ɣōs, hōɲ, yōs, wōse, hoi*

◢ B: *kabalyu, kabāyu, kabāyo, kabāyuʔ, kobayu, kabayu, kabaro, ʔaballo*

⌃ C: *ʤaran, ʤʰaran, ʤara, ʤaraʔ, aŋŋaraŋ, ʤaraŋ,* and *aʤara*

◗ D: *kuda, kudaʔ, guda, hoda, kudo*

◡ E: *suavali, yovaɾi*

✚ F Other types: *ramaiʔ, ʔua sa-sm-ovri, rigi, baka, bulmaot, háricān, solofanua, puʔarehenua, uma*
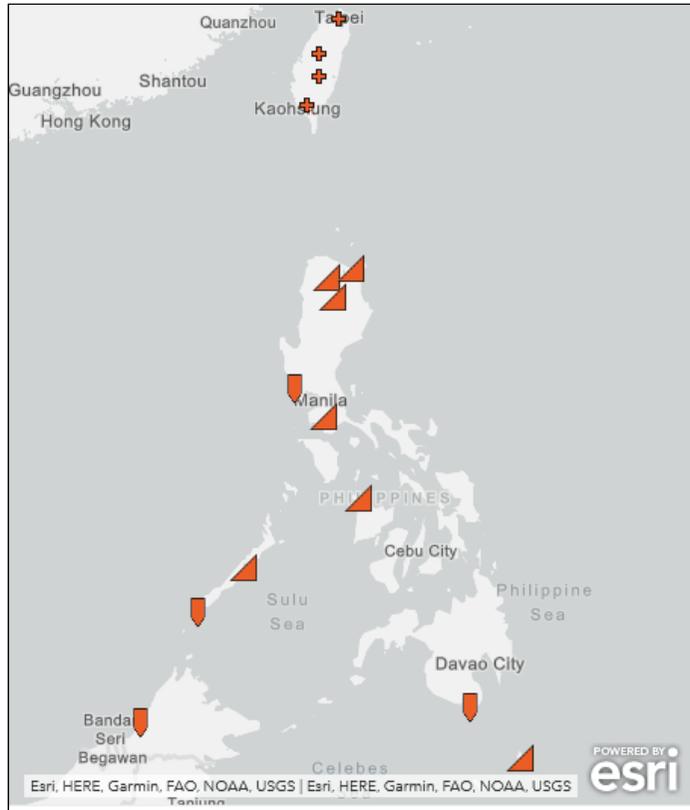
Figure 3.10.1: 'Horse' in Taiwan and the Phillipines.

Figure 3.10.2: 'Horse' in Indonesia.
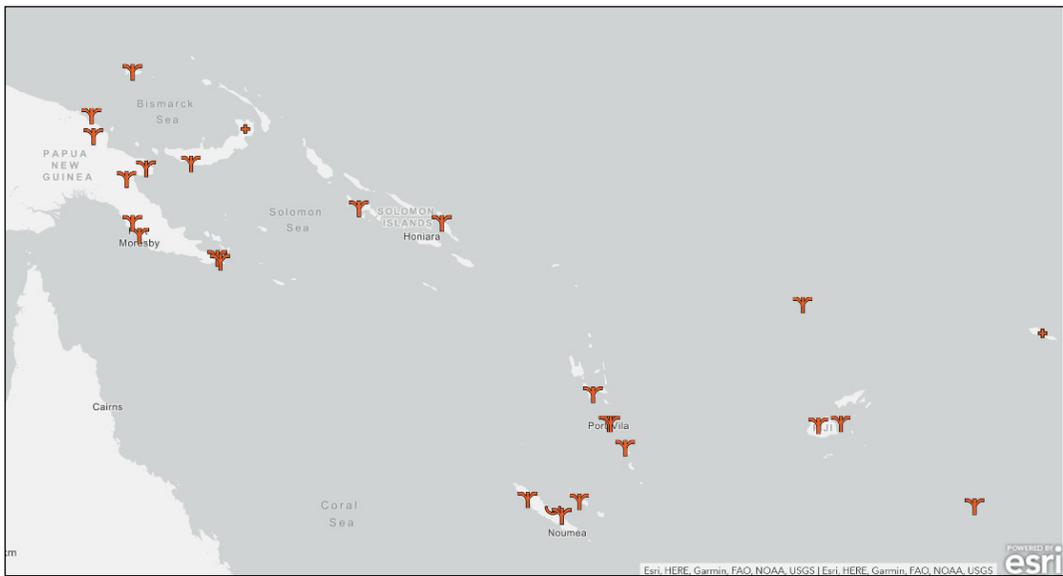
Figure 3.10.3: 'Horse' in Papua and the Pacific.

'Horse' in Tungusic

In Tungusic the word form for 'horse' is only MORI, which is a well-known borrowing from Mongolian.

(MATSUMOTO Ryo)

▮ A-type   MORI

Figure 3.11.1: 'Horse' in Tungusic.

'Horse' in Uralic

In Uralic there are various forms for 'horse'. Here I classified then just according to the sound forms, it could not be referred to the reason why such many forms they have.

A Finnish *hevonen*, Karelian *hebo*, Veps *hebo*, Estonian *hobune*, Ingrian *heppoin*, Votic *opõn*, Livonian *ibbi*

B Sami *heasta*

C Mordvin *alašan'*

D Mari *imne*

E Hungarina *ló*, Mansi *low*, Khanty *tow*, Forest Nenets *law*

F Udmurt *val*, Komi *vöv*

G Tundra Nenets *juno*, Enets *djuda*

H Selkup *čunty*

It could be possible as one of the reasons why such various forms are observed that the horse culture have affected each other with other surrounding cultures such as Scythian, Turkic, Mongolic and Germanic so on, regardless of whether Uralic people had the original horse culture. For example, Nenets people live in far north, for them reindeer is more important and indispensable for their life. Among them Forest Nenets are living in the more southern area very near to Khanty and Mansi. It seems that they borrowed *law* from Ugric people in place of Tundra Nenets *juno*.

(MATSUMOTO Ryo)

	A hevonen, hebo, opon, ibbi		E ló, low, tow, law
	B heasta		F val, vöv
	C alašan'		G juno, djuda
	D imne		H čunty
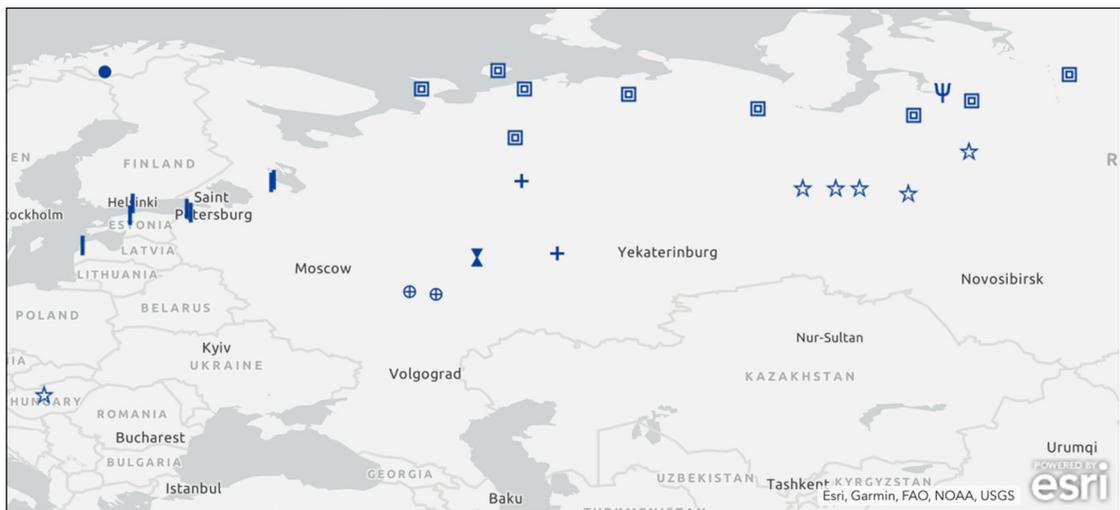
Figure 3.12.1: 'Horse' in Uralic.

'Horse' in Mongolic and Turkic

1. Mongolic
All Mongolic languages have a *mori*-type word for 'horse.'

2. Turkic
Most Turkic languages use *at*-type words.

Some languages in northeastern Siberia have *sïlgï* (Sakha) and *čïlɣï* (Chulym), which may be from *salgā at*, 'a restive horse' (cf. Clauson 1972: 826).

Chuvash has the form *laša* for 'horse.' It is related to *alaša* (Tatar, Bashkir, Kumyk) and *alasa at* (Nogay), which mean 'castrated horse.'

(SAITÔ Yoshio)

A. *mori* type
- mori, mœrʲ, mœr, morə, mɔrin, mørn, moriŋ, mōrə

B. *at* type
- at, āt, hat, aht, a't, ut

C. *sïlgï* type
- sïlgï, čïlɣï

D. laša
- laša

Figure 3.13.1: 'Horse' in Mongolic and Turkic.
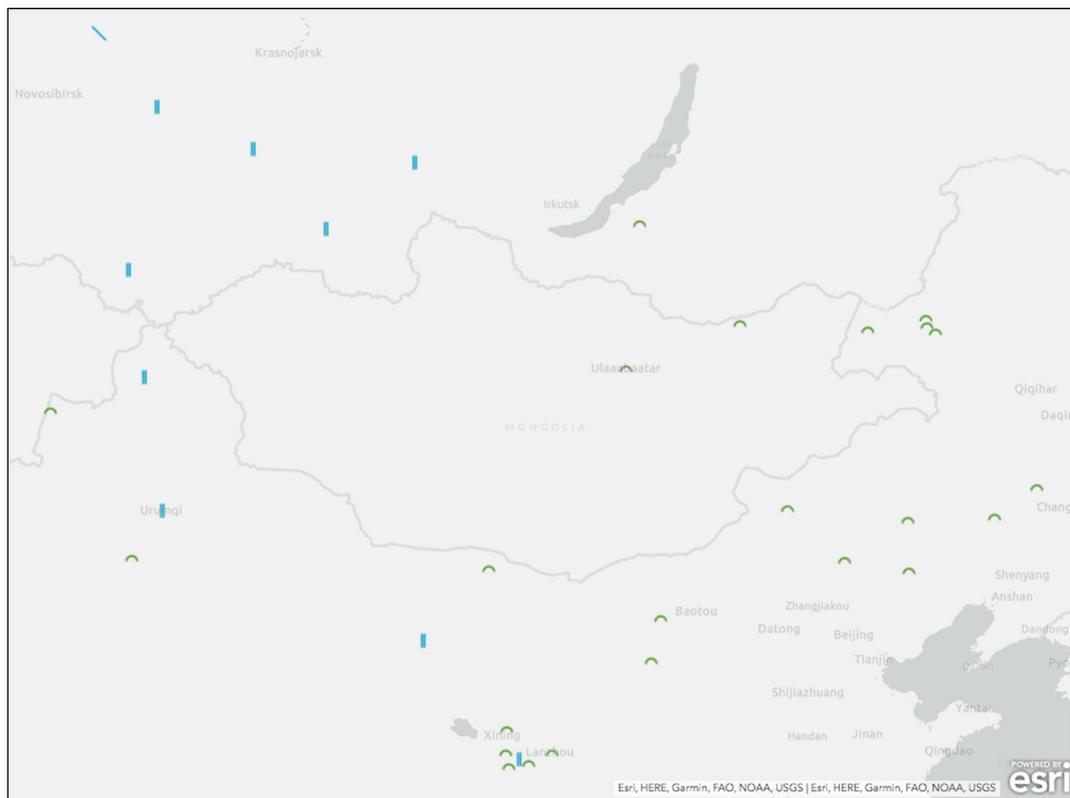
Figure 3.13.2: 'Horse' in Mongolic and Turkic (The Mongolian Plateau and its vicinity magnified).
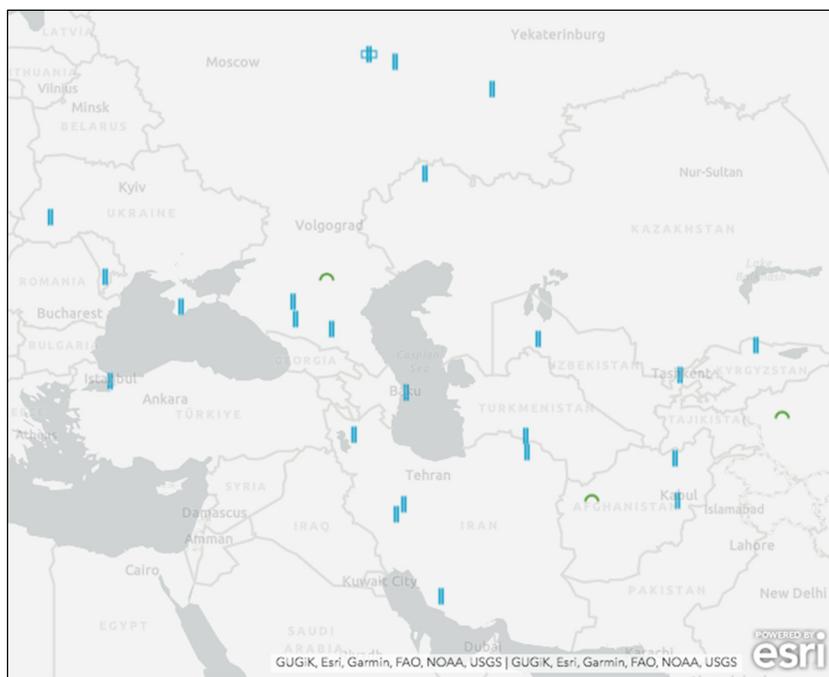
Figure 3.13.3: 'Horse' in Mongolic and Turkic (Central Asia to East Europe magnified).

'Horse' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. When a language has several words for horses (horse (stallion, gelding), mare, colt, filly, foal, or something), I targeted 'male adult horse'.

The distribution of 'horse' words is relatively simple. On the one hand, Indo-Aryan language all over the Indian subcontinent, the Middle East, Caucasus, and German employ the types A. On the other hand, Indo-Aryan languages in the periphery, namely the islands and the northwest mountains employ B type. Type C is for Romani in east and south Europe, and type D is for Burushaski.

The most major type is *ghōṭa*. This type is derived from Sanskrit *ghōṭa* घोट 'horse' (or *ghōṭaka* घोटक), and there are some scholars suggesting that the word is not Indo-European origin but may be from Dravidian (Proto-Dravidian **kHutt-*, Southworth 2006: 143). Forms of this type are used by Indo-Aryan and Nuristani languages. Many languages have lost the /gh/ [gʱ] sound and the word forms changed in variety. For example, Panjabi has changed all voiced aspirate stops into voiceless unaspirate stops as developing tonal distinction, and so Sanskrit *ghōṭa* has become *kòṛā* ਘੋੜਾ/گھوڑا with a falling tone in Panjabi, as well as *kòṛo* in Gojri.

The *áśva* type appears in Indo-Aryan and Nuristani languages, which are concentrated in the area of Himalaya, Karakoram, and Hindukush mountain ranges and on the islands (Sri Lanka and Maldives). The original Sanskrit form *áśva* अश्व refers to 'horse', directly from PIA **Háśwas* < PII **Háćwas* < PIE **h₁éḱwos* '(domestic) horse' (as opposed to **márkos* 'wild horse', being for Sanskrit *marya* मर्य 'stallion' and English *mare* among others). So this type is cognate with Latin *equus* and Ancient Greek *híppos* ἵππος. There are various modern forms of this type, of which the Brokskat form *āpš* ཨཱཔྴ is the most interesting, as it shows a metathesis.

The third one, a minor type grast can be seen in Romanic languages in Europe and Iran, while Sinti Romani in German has *khuro* of the type A. This type is in the vein of Old Armenian *grast* գռաստ 'pack animal', and the origin of the word has not been unvailed yet.

Next, the *hayúr* type is used only in Burushaski. This form might be inherited from Sanskrit *ghōṭa* (cf. *ghōṛā* گھوڑا 'male horse' in Urdu), which is of the type A, though Berger (1998: 185) guesses as from Turkish *ayğır* 'horse'.

Looking at the remaining spradic 'horse' vocabulary in South Asia, *ṭār* টার in Bengali is obviously derived from Sanskrit *ṭāra* टार 'horse'. In Gujarati they use *vari*, which is originated in Sanskrit *vārakin* वारकिन् 'dappled horse'. Khowar *istōr* 'horse' (while *koistāni* 'male horse' and *madiā́n* 'mare') is a rare descendant of Sanskrit *sthōra* स्थोर 'beast of burden'. The Nihali word *māv* is similar to Telugu *māvu* 'horse', so it can be from a Dravidian origin. Unusually, the Vedda word *hoṭava-lumaññaa* bears no resemblance to the Sinhala *as* අස් or *áśvayā* අශ්වයා. From the

sound, it may possibly be related to the   (YOSHIOKA Noboru)
Proto-Dravidian *kHutt-*.

A. *ghōṭa* type (58) ⌒

    ghoḍā, ghŏ̃ḍŏ̃, ghŏ̃ṛǎ̃, ghɔṛa, ghŏ̃ṛŏ̃, ghuṛo, ghoṛū, ghūṛu, ghōṛɨ, gŏ̃ṛǎ̃, gōṛo, gōṛa, gŏ̄ṛu, guṛa, guṛɔ, guṛə, gǎ̃ṛ, gōṛ, khoṛā, kòṛā, kòṛo, kuṛ, ghõṛa, gharā, gōri, gǒ̃r, gur, khuro, khori, ghɔ̃ř̃ʌ, gǫřɔ, ghunni, ghǒ̃, ghǎ̊, gōa, īri

B. *áśva* type (12) ⏐

    ašp, ã̃šup, ašvayā, ʌšpo, ušpa, ušup, wušup, āpš, aṣpa, hã̃š, as

C. *grast* type (6) ◯

    grast, gerāst, gras, gri

D. *haɣúr* type (3) ⟋

    haɣúr, haɣór

E. others

    ṭār, jīrmu, vari, istór, tokli, panāvavrō, māv, hoṭavalumaññaa
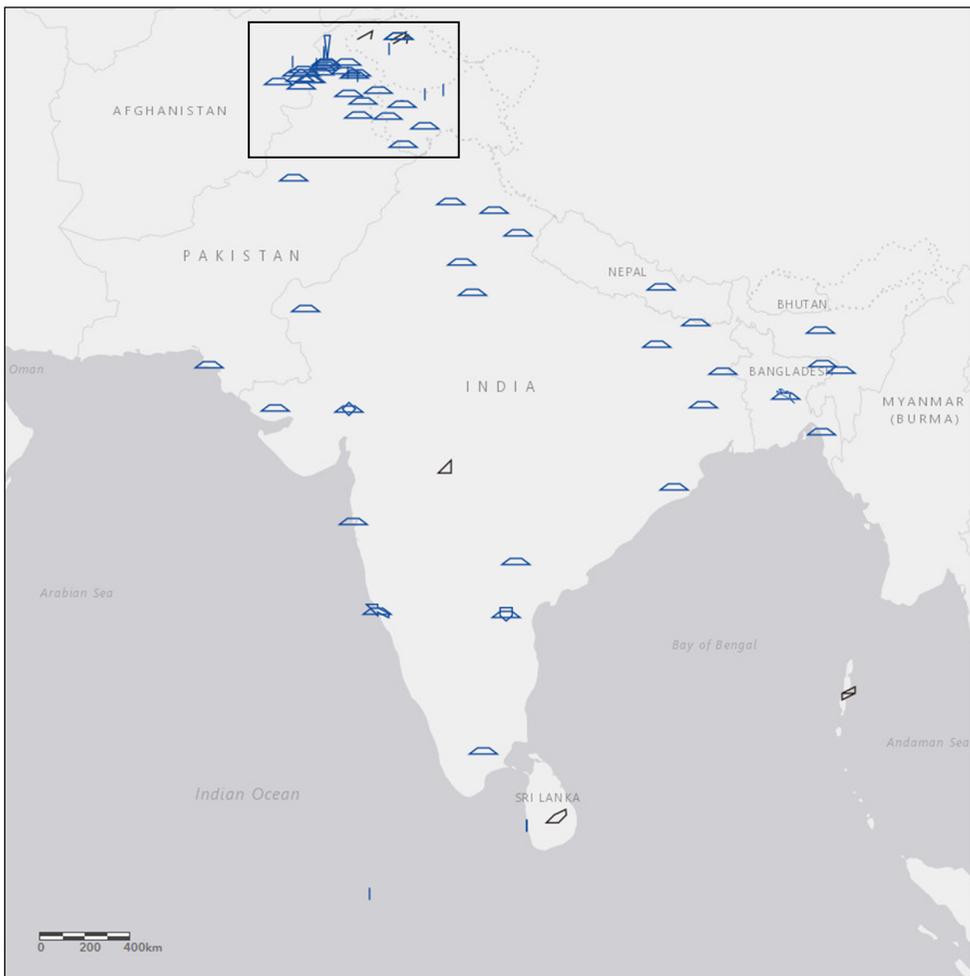
Map 3.14.1: 'Horse' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).
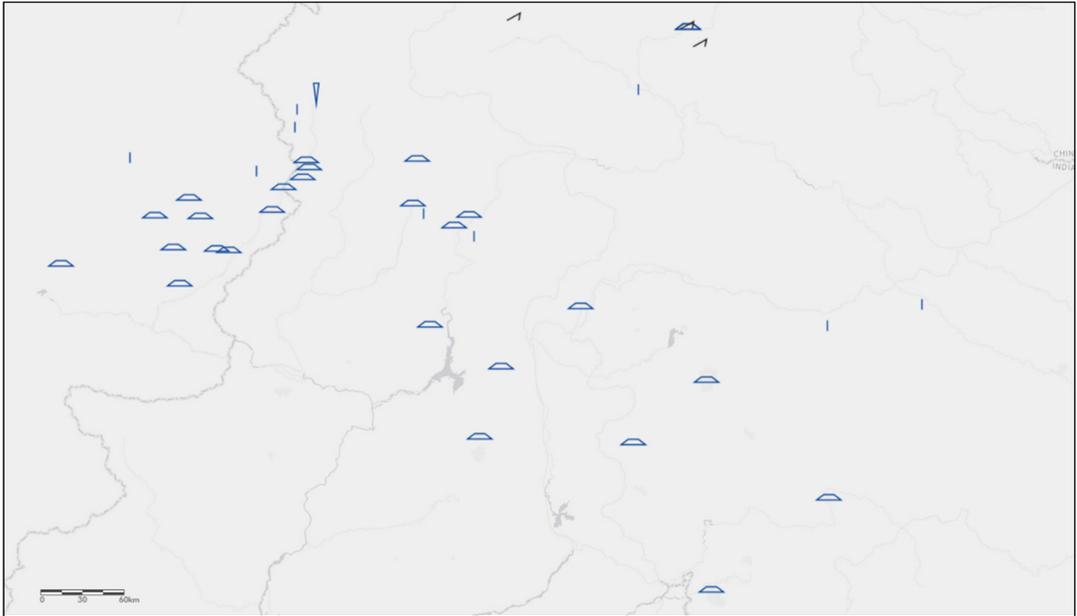
Map 3.14.2: 'Horse' in northern Pakistan (the area encloed by the rectangle in Figure 3.14.1).
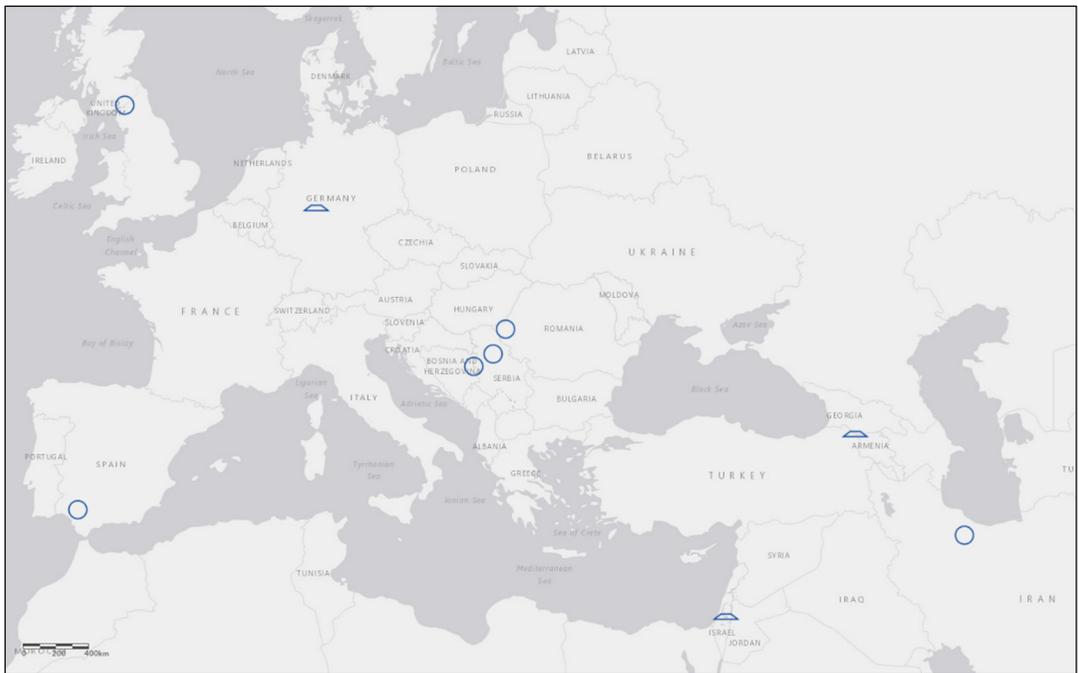
Map 3.14.3: Types for 'Horse' in Indo-Aryan languages outside South Asia.

'Horse' in Dravidian

Horses, *Equus caballus*, are generally assumed to have been brought into the subcontinent in the Late or Post-Harappan period, probably by Indo-Iranian speakers. If reconstruction of a proto-form of 'a horse' is plausible for the Dravidian family as a whole or some of its genealogical subclades, it would entail that the diffusion of the language family or the subclade started on the subcontinent after the introduction of horses there.

If the isolated Brahui word *(h)ullī* 'a horse' is cognate with the Classical Tamil word *ivuḷi* id. as proposed in DEDR #500, or *uḷai* 'mane' as proposed by Emeneau (1997), these would be considered as reflexes of the Proto-Dravidian 'horse', although there are no other cognates retained in the family. In fact, except for the four literary languages, Tamil, Malayalam, Kannada and Telugu, the words for 'a horse' appear to have been borrowed rather than inherited.

The South Dravidian (and Telugu) etymon KUTIRAI (DEDR #1711a) is related in DEDR to the verb etymon KUTI (DEDR #1705) 'to leap, jump'. DEDR separates the Telugu form *gurram* as #1711b, which was borrowed by South Central and Central Dravidian languages, as a possible borrowing from Skt *ghōṭa*, which in turn is assumed by *A comparative dictionary of Indo-Aryan languages* (Turner 1985; CDIAL) to be of Non-Aryan, probably Dravidian(!), origin.

Modern Indo-Aryan reflexes of *ghōṭa* are borrowed by Kurukh, Malto, Kui and Gondi.

Besides, Kobayashi & Tirkey (2017) recorded Kurukh *hakuṛ* 'horse, mare'. If this isolated word is analyzed as suffixed *ha-kuṛ*, *ha* could possibly be related to Skt. *áśva* > Pkt. *aha*, the old Indo-European etymon which was replaced by *ghōṭa* in most Indo-Aryan languages.

(KODAMA Nozomi)

- ⌄ kutirai, kudira, kudire, kudure
- ⌣ kïθïr, kudyr
- △ gurram (?<kudaram), guṟam, gurromi, gūrumi, ghurram, kurmam, gurrol
- ✳ ullī
- ☆ ívuḻi
- ◐ gōḍā, kōṟā, ghoṟo:
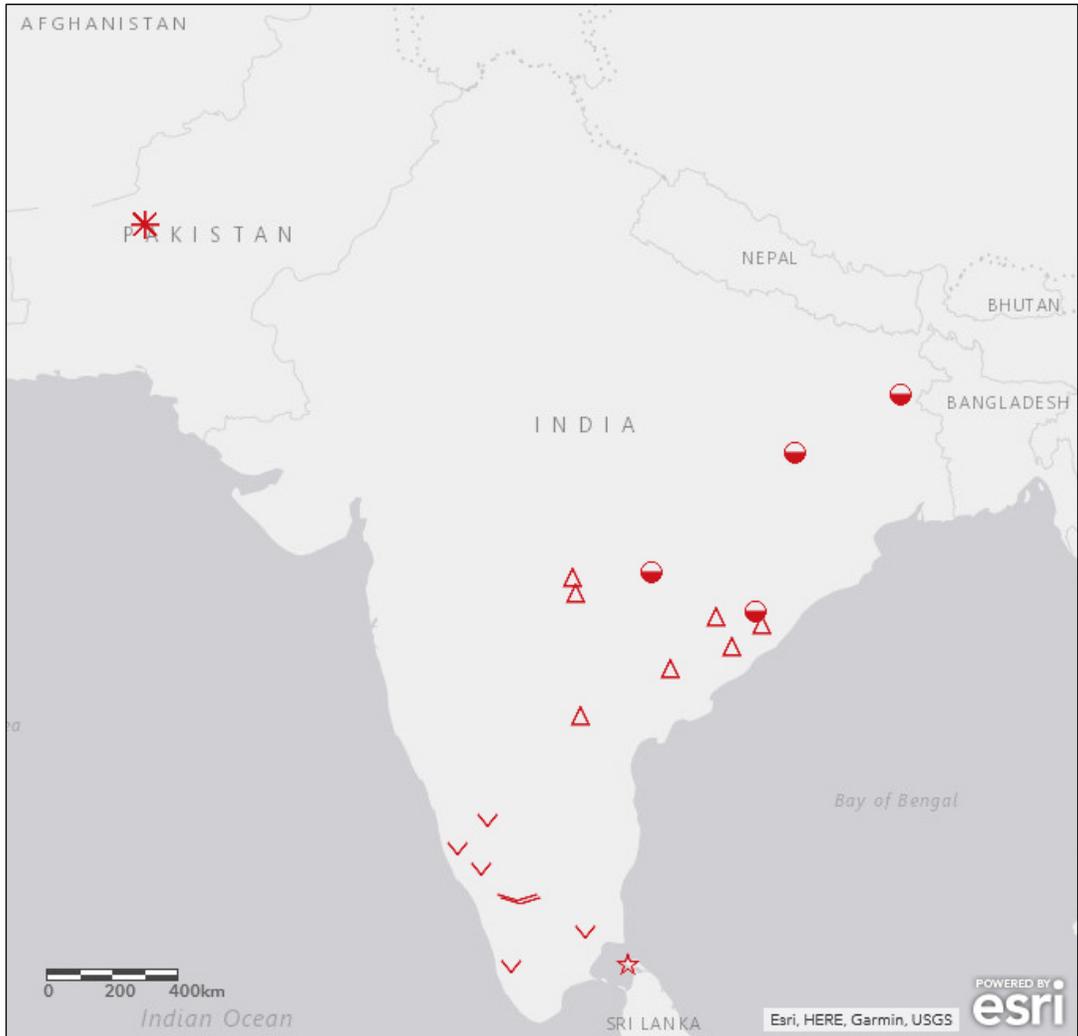
Figure 3.15.1: 'Horse' in Dravidian.

'Horse' in Iranian

Most of the Iranian languages have Type A words, except for the majority of Pamir languages, and Ossetic. This word traces back to PIr. *átsua- (< PIE *ek̂ʷos). Type B derived from the PIr *bāraka- 'sumpter animal, packing animal'. All the Pamir languages, except for Wakhi (Type A), have this type of word.

Only Type C is a loanword (probably from Nakh languages) among this category. Both Ossetic dialects fall into this type. Although their ancestors are considered to be equestrian people (Alan, Schythian), they borrowed from the foreign word denoting 'horse', and limited its semantic ranges of inherited words for 'horse'. Note that Ossetian has words cognate with Type A and B, but their meaning is limited (the former type is *jæfs/ æfsæ* for 'mare', the latter is *bajrag* for 'foal').

(IWASAKI Takamasa)

- ─ A: *asp* Type
- ◊ B: *verk* Type
- ◗ C: *bæx* Type
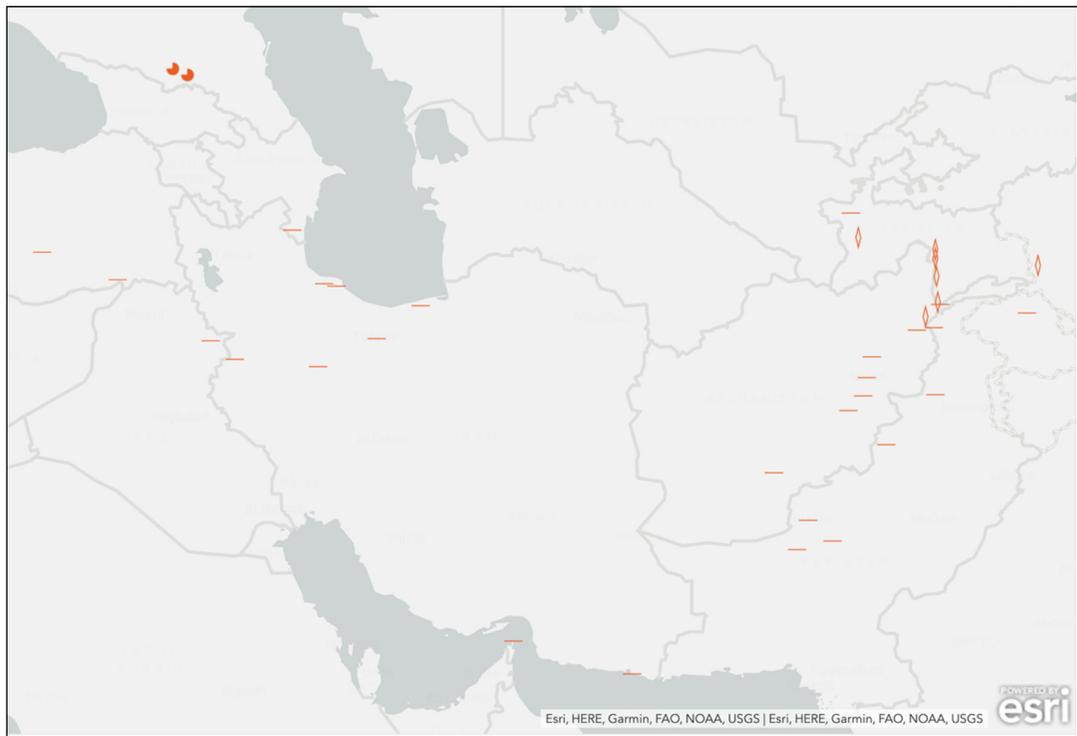
Figure 3.16.1: 'Horse' in Iranian.

'Horse' in Caucasian languages

Most Caucasian languages use different forms for 'male horse' and 'female horse'. The present description is of the former.

Type A, derived from a common root *cxen-, appears in Kartvelian languages except for Svan. Type B is found in Abkhazo-Adyghean languages and Svan. Type C exists in Nakh languages (Chechen and Ingush), as well as Avar. Types D and E are distributed in the north of the Dagestanian-speaking area. Types F and G are distributed at the periphery of the southern Dagestanian-speaking area, and Type H appears in its central region.

Klimov & Khalilov (2003:234) state that Type H is derived from 'animal'. Based on its distribution, Type H can be a newly established form, compared to Types F and G. The origin of Type I, including various forms, is unknown.

(SUZUKI Hiroyuki)

A: cxen-type; *cxeni*.
B: čɨ-type; *čɨ, čän, šɨ*, etc.
C: g-type; *gowr, gulu*.
D: k-type; *kotu, katu, čatw*, etc.
E: χ-type; *χχwani, χʷani*.

F: s/š-type; *soro, šügo-sojra, pši*, etc.
G: b-type; *balkan, balčan*.
H: h-type; *hajwan, hijwan*.
I: others; *noš, dõ, ek*.

Figure 3.17.1: 'Horse' in Caucasian languages.

'Horse' in Semitic

A. *ħiṣaːn* type (▲) is Arabic form (Classical Ar. حصان). Phonetic varieties are *ħuṣaːn* (Egypt, Sudan, Libya of Africa and Sanʼa in Yemen); *ħṣaːn* (Syria, Palestine, Jerusalem, Lebanon, Iraq, Gulf); *ħṣaːne* (Tunisia) with a feminine ending *-e* indicating the unity 'a horse'.

kuusán (◇ Nubi) is related to *ħiṣaːn*. *hasun* (△ Jibbali) may also be related to it.

B. *ʕawd* type (◆) is found in Algerian Ar. *ʕawd* and Moroccan Ar. *ʕəwd*.

C. *ziːmel* (o Maltese Ar. żiemel) goes back to *zaːmila* (زاملة) 'beast of burden' (cf. Dic. of Hava 1899: 288).

D. *faras* type (★) is found in Arabic and Ethiopic: *faras* (Najdi Ar.), *fərəs* (ፈረስ Tigrinya, ፈረስ Amharic).

E. *suːs* type (●) is Hebrew and Aramaic form: *suːs* (Hebrew סוס), *suːsa* (Koy Sanjaq Aram.), *susa* (Hertevin, Jilu Aram.), *səsjo* (Turoyo Aram.).

F. *dʒawaːd* type (◘) is found in Chad and Nigeria: *dʒuwaːd* (Chadian Ar.); *dʒawáad* (Nigerian Ar.) cf. *dʒuwaːd* 'courser, race horse' in Classical Ar.

G. *farħiin* (Hobyot), *ferhajn* (Mehri) in South Arabian (▼) .

H. *dabba* (✕) (Bukhari Ar.) may be derived from *daːbba* 'riding animal' (Cl. Ar.). *ɣdiːʃ* (■ Cypriot Ar.) is derived from *kadiːʃ* 'cart horse' (Cl. Ar.).

(NAGATO Youichi)
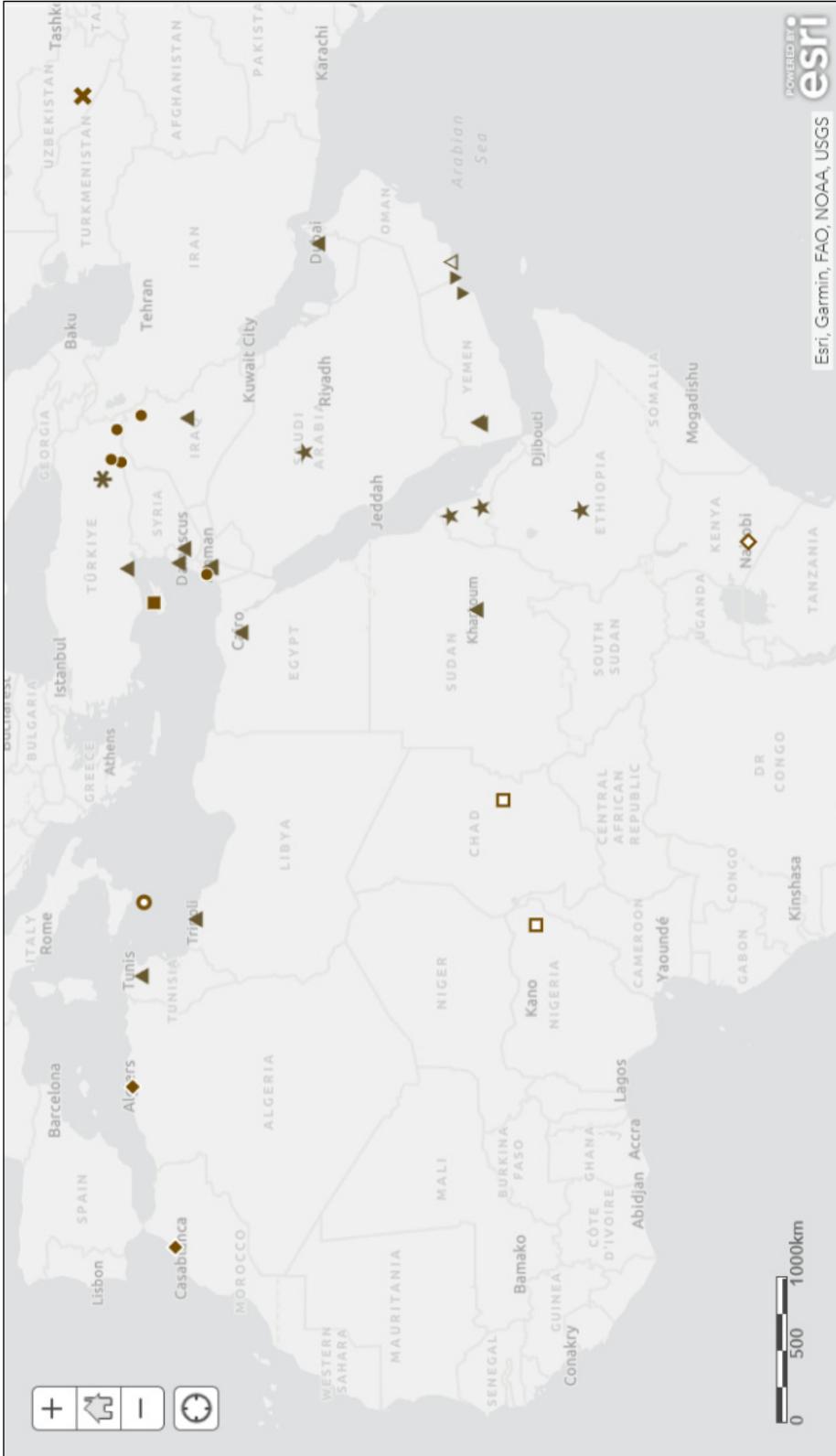
Figure 1.18.1: 'Horse' in Semitic.

'Horse' in Nilo-Saharan

What follow are heuristically reconstructed (marked with a hash #) Nilo-Saharan roots for 'horse' (83 languages surveyed). We simplify diacritics and notations for the [±ATR] feature in the original data.

The most widespread root among Nilo-Saharan (known since MacMichael 1918) is A1 *#murta* and its possible variants A2 *#(m)burta*, A3 *#mutta/munta*, A4 *#furta* and A5 *#far/bar*, which are attested in all Eastern Sudanic branches, i.e., Nyimang (*morta*), Nubian (Nobiin *murti*), Taman (Misiirii *furta*), Nara (*fara*), Berta (*murtha*), Jebel (Gaam *mosor*), Daju (Nyala *murtane*), Temein (*manta*), Surmic (Koegu *parda*, Suri *hartey*) and Nilotic (Kipsigis *baraisit*, Maa *em-barta*, Mayak *morcoŋ*), as well as all the other major Nilo-Saharan branches, Saharan (Beria *hirde*, Berti *burto*), Mabang (Masalit *beiro*), Fur (*murta*), Kunama (*burasa*), Central Sudanic (Yulu *moot*, Kresh *moroto*, Furu *mbarata*), Songhay (Zarma *bari*), Gumuz (Daats'iin *marta*), Koman (Komo *parſa*) and Kadu (*muttu*).

Given its genetic and geographical extent, this root could be proposed as the proto-Nilo-Saharan for 'horse' if any, but at least some of these isoglosses could be a result of later diffusion by Nubian speakers in the Nile Valley (cf. O'Fahey 1980: 96, Quint 2013). Note that similar roots are attested in many genetically irrelevant languages, e.g., Ubangian (Banda *berta*, MacMichael 1918; Sango *mbarata*, Bouquiaux et al. 1978), Kordofanian (Koalib *mortta*, Quint 2013; Tima *mɘrtaa*, Dimmendaal 2019) and Afroasiatic (Wolaytta *para*, Oromo *farda*, Yaaku *barta*, Amharic *färäs* 'horse', Arabic *faras* 'mare'; Heine 1975, Blench 2008a),

although this could be due to coincidence (cf. German *Pferd*). The distribution of its possible variants does not fully meet genetic criteria. On the other hand, since it is historically impossible to attribute the East African Nilotic cognates to Nubian influence, *#murta* could be suggested for a proto-Eastern Sudanic root at latest.

Type B *#kaj(-nV)* is attested by some Nubian (Kenzi *kaj*, Birgid *kisi*) and Western Nilotic (Agar Dinka *akaja*, Shilluk *kyeny*) languages. This root is attested with the sense of 'donkey' in diverse Nilo-Saharan branches, such as Eastern Nilotic (Mandari *kayina*), Daju (Nyala *kacane*), Central Sudanic (Kara *kacini*, Kresh *keſe*, Bongo *akaca*) and Kadu (Katcha *kisine*), in addition to some other Nubian (Nobiin *kaj*) and Western Nilotic (Rek Dinka *akaja*, Luo *kanyina*) languages.

Many Sara-Bongo-Bagirmi languages of the Central Sudanic branch share a unique root reconstructed as Type C *#sunda* (cf. **s-nd-* by Boyeldieu 2000a).

A few (South) Sudanese Nilo-Saharan languages attest a loanword from Sudanese (Type D) *ḥuṣān* or *jawād* 'horse' (Madi *kusani*, Lotuho *akusan*, Aja *jowata*) or *baġal* 'mule' (Bari *bakala*, Uduk *bagal*).

There are some other isolated cases categorized as Type E, including Kuliak Ik *nyaŋole* with Eastern Nilotic Karimojong *angole* and Turkana *angole* (all meaning 'horse'), and Me'en (Surmic) *sigiro*, which is related to an isogloss for 'donkey' in adjacent Eastern Sudanic languages, e.g., Surmic (Murle) *dhigir*, Berta *fiŋir*, Nilotic (Lotuko *asigira*, Nandi *sigiriet*).

(NAKAO Shuichiro)

∕ A1. #*murta*
– A2. #*(m)burta*
– A3. #*mutta*/*munta*
∕ A4. #*furta*
∣ A5. #*bar*/*far*

■ B. #*kaj(-nV)* < 'donkey'
△ C. #*sunda* (Sara-Bongo-Bagirmi)
● D. Sudanese Arabic loanwords
○ E. The other types
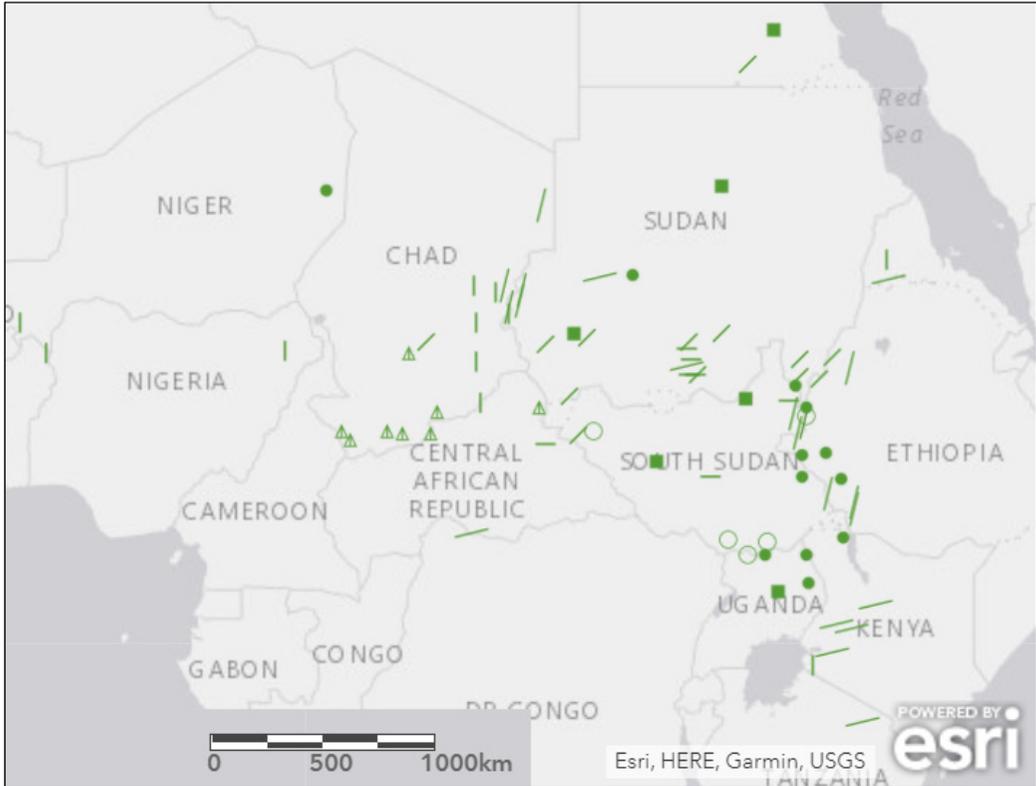
Figure 3.19.1: 'Horse' in Nilo-Saharan.

'Horse' in Bantu

Reflecting the scarcity of indigenous species of horses in Sub-Saharan Africa, no reconstruction for 'horse' is available in the Proto-Bantu lexicon. As shown in the map, in almost the entire area of Eastern Bantu zones 'horse' is a loan word introduced from Arabic *faras*, e.g., in Nguungulu [B72a] *ifalasi*, Luhyia [E32] *farasi*, Tharaka [E54] *mbarathi*, Rombo [E623] *ilifwalasi*, Wungu [F25] *ichibharasi*, Sangu [G61] *afarasi*, Ganda [JE15] *mpalasi*, Nata [JE45] *faashi*, Nyiha [M23] *farasi*, Ndengereko [P11] *embarasi*.

As far as it concerns, 'horse' in Bantu seems to be lexically non-diverse. However, if we focus on local variation, the situation looks different. Figure 2 shows that, in the north eastern part of Tanzania, especially in Kilimanjaro Bantu (Chaga) languages [E60] and the interlacustrine zone J languages, a handful of distinctive lexical forms are observed; e.g. *<ɲumbu>* as in Rombo-Useri [E623A] *nyumbu*, Nyaturu [F32] *nyumba*, Kwaya [JE251] *inyumbu*, and Nyamwanga [M22] *inyumbu*; *<nzoi>*

found exclusively in Central Kilimanjaro Bantu languages including Mochi [E622A] *tsoi*, and Lema [E622C] *nzoi*; *<dogobe>* in Hangaza [JD65] *indogobe*, and in Wanji [G66] *dogovi*; *<tikiri>* in Maragoli [JE41] and in Shubi [JD64] *inturege*; *<ʃuŋgu>* also exclusively found in Westen Kilimanjaro including Machame [E621B] and Siha [E621C] *nshungu*; *<bwisi>* in Kilegi [E25] *imbwisi*.

Another interesting phenomenon is the process of semantic change where a lexical form that originally had a different meaning has been converted to refer to 'horse'. One of the striking examples is observed in Lingala [C30B] *púnda* and Rufiji [P12] *mbonda*. While these forms are a clear descendant of the Proto-Bantu *-pʊnda* [BLR-MAIN-4397] (as registered in Bastin *et al.* (2002)), which means 'donkey', they have converted to be used as a word for 'horse'.

(SHINAGAWA Daisuke and
KOMORI Junko)

Common forms
- ○ farasi
- ╱ ɲumbu
- ╲ nzoi
- ⌄ tikiri
- ⌃ dogobe
- ‖ punda
- ѱ ʃuŋgu
- ■ bwisi
- ⊟ horse
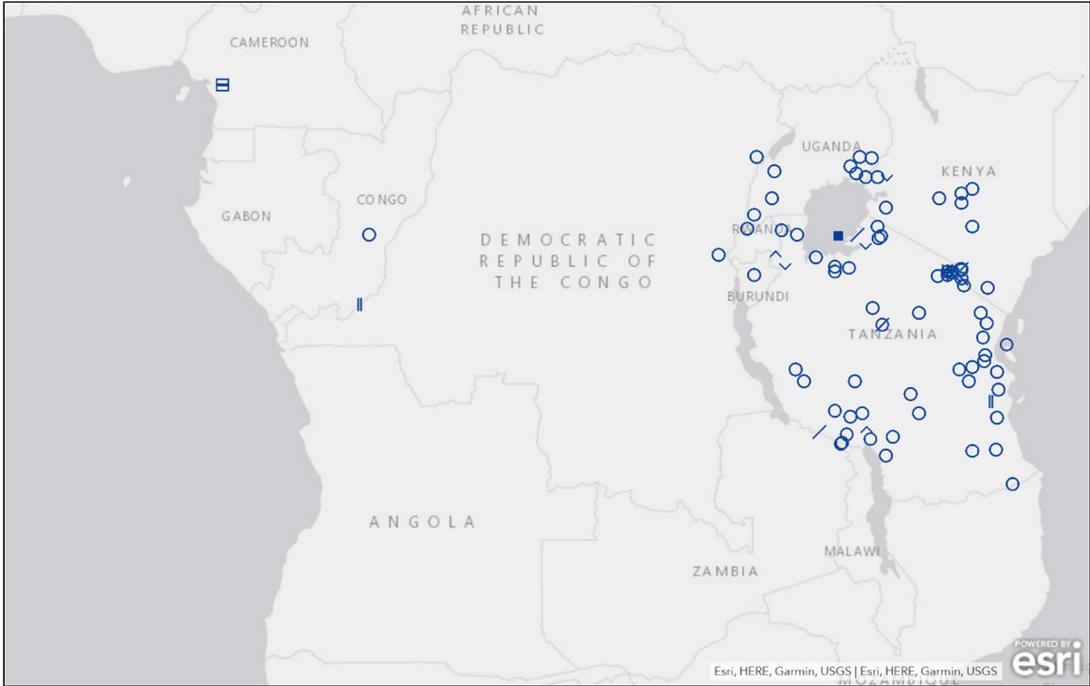
Figure 3.20.1: 'Horse' in Bantu.
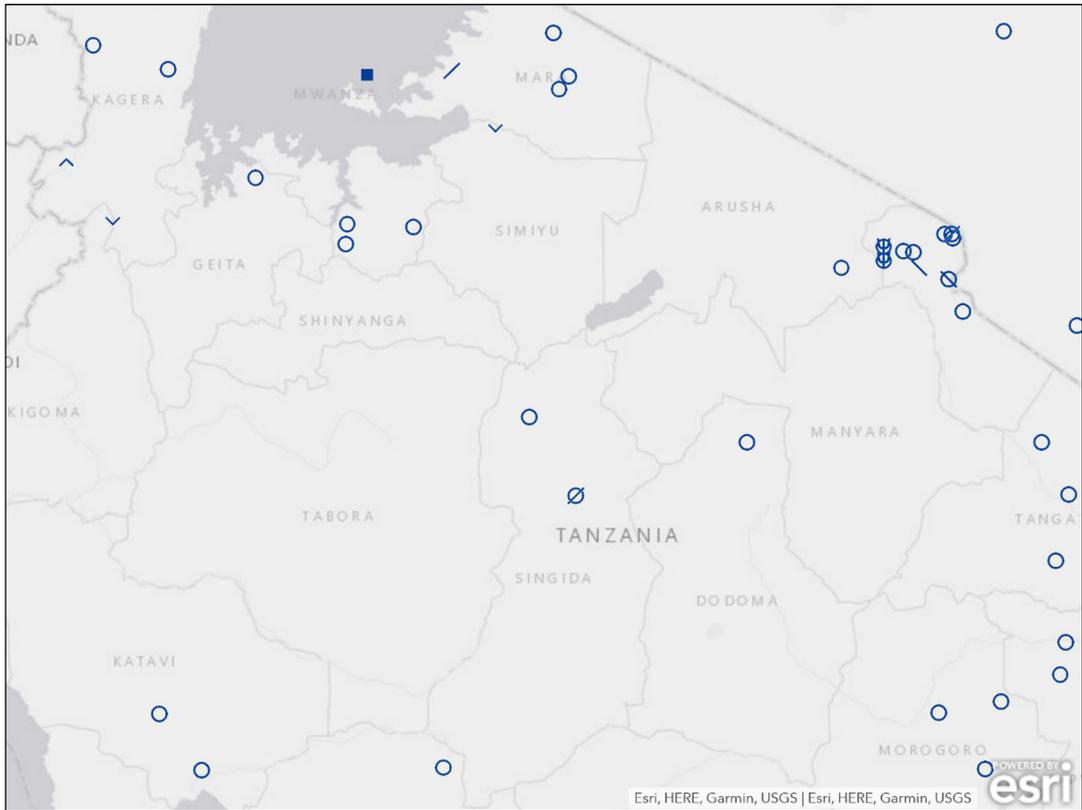
Figure 3.20.2: 'Horse' in northeastern Bantu zones.

Chapter IV

Dog

'Dog' in Asian and African languages

The dog (*Canis lupus familiaris*) is one of the earliest animals to be domesticated from its ancestor the wolf (*Canis lupus*). The place where dogs were first domesticated remains an open question, but it seems safe to say that it occurred somewhere in Eurasia (see Irving-Pease 2018 and Perri et al. 2021). Dogs are widely distributed world-wide, and most languages have a word form for 'dog'.

Data for words for 'dog' are available for all language families and all language groups. The word form, stem and root for 'dog' vary depending on language families and groups, as shown in Table 1.

Table 1: Main word forms for 'dog'.

Languages	Word forms
Chukotko-Kamtchatkan	*ʕətʕən* *qosχ*
Ainu	*setá*
Japonic	*inu* 犬
Korean	*kɛ:*
Sinitic	*quan* 犬 *gou* 狗
Hmong-Mien	*klu* *ljaŋ*
Tibeto-Burman	PTB **d-kʷəy-n* (WrT *khyi*) PTB **m-par* ⋊ *pra* (WrT *'phar*) PTB **na*
Kra-Dai	*maA* *pa* *hang*
Austroasiatic	PMK: **cɔ:ʔ* Proto-Khasic: **ksəw* PMK: **cgəy* Proto-Waic: **mrok*
Austronesian	aCu-type k/hu-type a...ŋ –type
Tungusic	NGINA- *jɔnχun*
Uralic	*koira* *pine* *weh*
Mongolic	*noxoi*

Turkic	*it* *köpek* *adaj*
Indo-Aryan	*Proto-IA **śwā́* *kutta / kuttira* *kurkurá*
Burushaski	*huk*
Dravidian	*nay* *alla* *kukka*
Iranian	*span* *kut* *ɣalv* *tuta*
Caucasian	Proto-Kartvelian **dz₁aɣl-la*-type h-typa χw-type
Semitic	*kalb* *wïʃʃa*
Nilo-Saharan	#bis(i), #bi, #is(i) #gok, #ŋok, #goŋ #kal, #kan, #kud #bel/ber
Bantu	Proto form **-bóà*, #bu *kuri*
KBA	ǂqʰài gǂʰóé ʔaba

Some languages, such as Japonic and TB, also use the root for 'dog' to denote 'wolf'; see Chapter V WOLF. Some other word forms are related to a word for 'jackal'.

Indo-Aryan and Iranian are sister language groups and share some etymons, such as Proto-IA **śwā́* and *span*.

In Sinitic, the form *quan* is a cognate with PTB **d-kʷəy-n*, and *gou* is considered to have originated from Proto-Hmong-Mien **klu*[B] (Ostapirat 2016), which is further related to the Austroasiatic substratum (Akitani et al. 2022:271-273). The Indo-Aryan form *kurkurá* was further borrowed in Dravidian as *kukka*. Generally speaking, the borrowing among different language groups and families do not occur frequently.

Although several stems are attested in a single language family or group, such as TB, Kra-Dai, Austroasiatic, Uralic and Nilo-Saharan, it is unlikely that this morphological variation reflects subspecies or relationships with taboo words. Rather, the forms originally had a range of lexical forms. A hypothetical conclusion would be that several word forms are derived from onomatopoeia of the dog's bark; see Nakazawa and Yokoyama (this volume) for details.

(SUZUKI Hiroyuki)

'Dog' in Chukotko-Kamchatkan

Dog is *ʔatʔən* in Chukchi and *ʕatʕən* in Alutor and Koryak (Kurebito et al. 2001). There is a glottal-pharyngeal corres-pondence /ʔ/-/ʕ/.

In Itelmen, dog is *qosχ* in both dialects (Kurebito et al. 2001).

According to Fortescue (2005), Chukchi-Koryak-Alutor *ʔatʔən~ʕatʕən* and Itelmen *qosχ* are congnate.

(ONO Chikako)

A. *ʔatʔən~ʕatʕən* type
■ A-1 *ʔatʔən*
□ A-2 *ʕatʕən*

● B. *qosχ* type
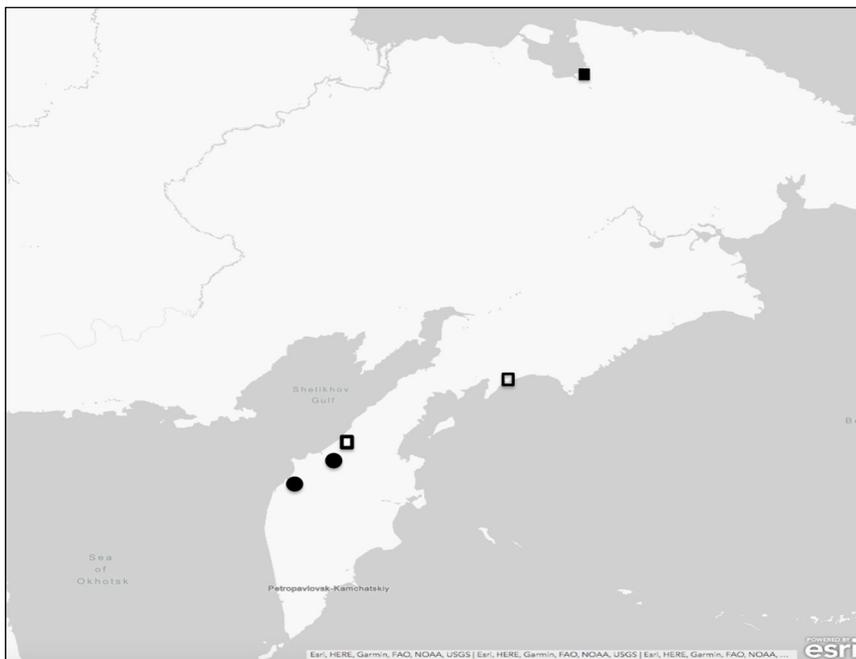
Figure 4.1.1: 'Dog' in Chukotko-Kamchatkan.

'Dog' in Ainu

The term for 'dog' (*Canis familiaris*) can be regarded as a monotonous *setá* type in Ainu. According to Kindaichi (1993 [1937]) and Chiri (2001 [1952]), the Matagi people — the traditional hunters who live deep in the mountains of the Tohoku region of northern Japan — have *seta* セタ, *ʃeda* シェダ, *setta* セッタ and *heda* ヘダ in their language, which mean 'dog' or occasionally 'dog meat.' In the old Ainu lifestyle, dogs must remain close to the people, for hunting or pulling sleds. More terms exist, depending on the role, appearance, sex, and age of the dog, such as the name of *apácapunki* 'entrance keeper' (Bihoro dialect) and *késoseta* 'dog with mottled patches.' In addition, dogs are worshipped as *reyép kamuy* or 'crawling spirit-deities' because they crawl when they approach people that they would like to fawn over (Chiri 1976 [1962]: 137).

(FUKAZAWA Mika)

A. *setá type*
   A-1. *setá ~ seta*   A-2. *sitá ~ sita*

Figure 4.2.1: 'Dog' in Ainu.

'Dog' in Japonic

As for the words for dog in Japonic, the INU types (IN, INU, INKO, INUKO, and INNOKO) are widely found in mainland Japan and Ryukyus, and some other types, such as KANKAN (*ganga*, *gangaa*, *gangan*), KOROKORO (*korokoro*, *koyokoyo*, *korokorokoro*), and WAUWAU (*wawa*, *waawaa*, *wauwa*, *wawako*, *wanko*, *wanwako*, *wanwanko*) are also found.

KO in INKO, INUKO and INNOKO types is a diminutive derived from *ko ~ kwaa* 'child' and is also found in animals other than dogs such as UMAKO (*umakko*, *makko*, *mako*, *umaNko*) 'horse' and TORIKO (*toriko*, *torikko*) 'bird, chicken', but dogs are different in that forms with diminutives such as *ʔingaa* and *ʔinnukwa* are also found in Ryukyuan languages. This is probably because the dog is a more familiar animal than the others. KANKAN and WAUWAU probably changed from the dog's bark. In other animals, the words derived from the cry, such as NYAANYAA for cats and MOOMOO for cows/oxen, are widespread in baby talk. INU itself may have originated from the sound of barking (Otsuki 1932: 330), possibly from *gen-u. This can be seen in other languages such as the Chinese *quǎn* (< MC *kʰwen*) (Todo 1978: 817) and Latin *canis* (< Proto-Indo-European *ḱwṓ*). In any case, the oldest attested form for the dog is *inu* in Old Japanese, and the proto-Japonic form for dog is presumed to be *inu, which is similar to Tungusic forms such as Even *ӈин* /ŋɪn/ and Nanai *инда* /ịnda/, from Proto-Tungusic *ŋinakin ~ *ŋine. Since all of these forms might be coined from dog's barking, the resemblance is due to onomatopoeic motivation. The proto-Japonic form for dog in Chinese zodiacs "戌" is also *inu.

(NAKAZAWA Kohei and YOKOYAMA Akiko)

	IN
	INU
0	INKO
◠	INUKO
0	INNOKO
⌒	WAUWAU
△	CIN
∧	KANKAN
∨	KOROKORO
□	TOOTOO
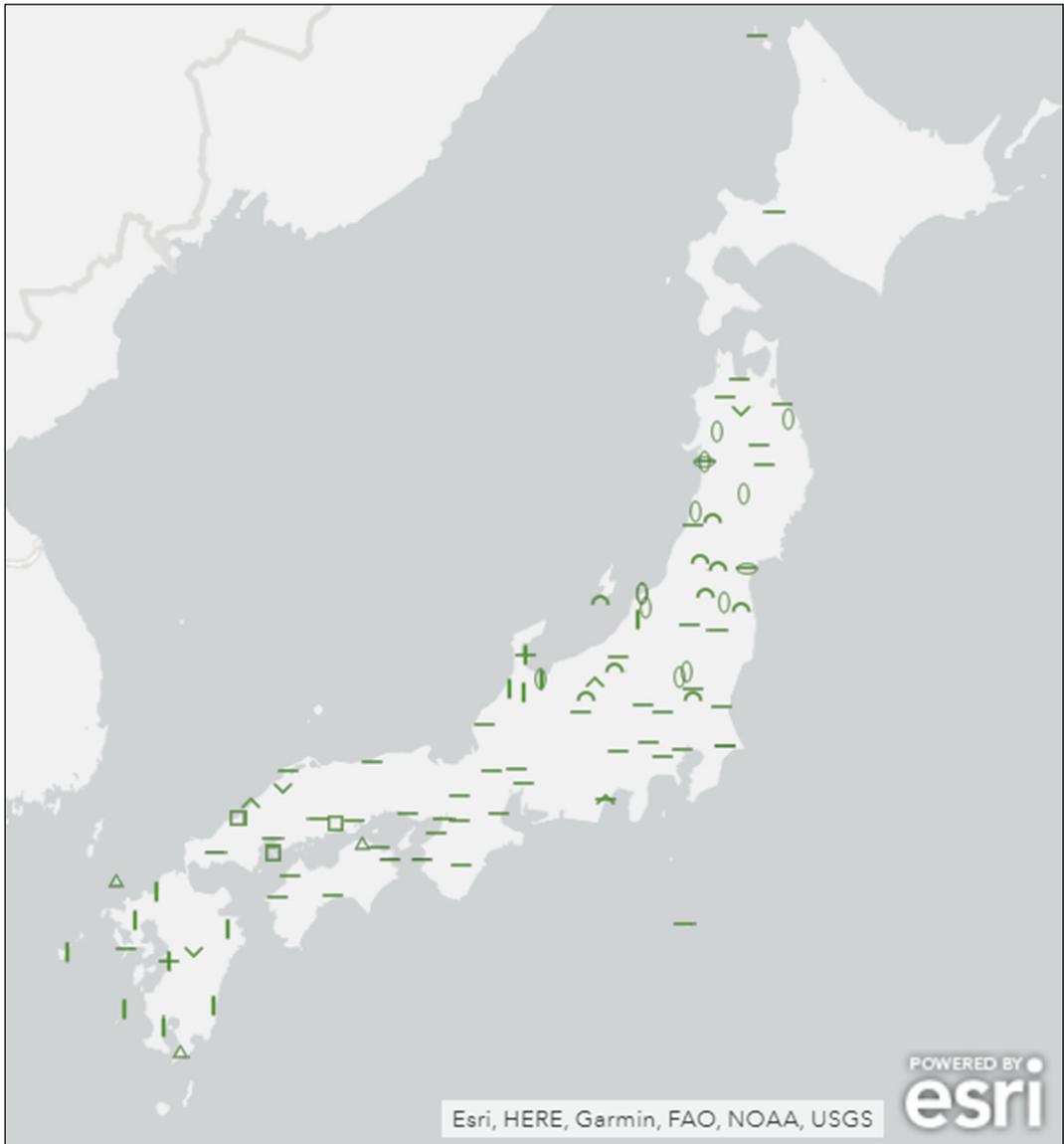
Figure 4.3.1: 'Dog' in mainland Japan.

Figure 4.3.2: 'Dog' in Northern Ryukyu Islands.

Figure 4.3.3: 'Dog' in Southern Ryukyu Islands.

'Dog' in Korean

Modern standard form for 'dog' is 'kε:' and Middle Korean form 'kahi'. Middle Korean is the language spoken from the middle of the 15th century to the end of the 16th century. The modern form is the result of two phonetic changes: the loss of medial 'h' and the contraction of the the diphthong 'ai' to a monophthong 'ε'.

Dialect variation is not so great. Aside from forms made up by adding various suffixes (such forms can have a pejorative meaning), we have only a few phonetic varieties such as the following:

A-1 kε:, A-2 kai

A2 seems a little bit similar to the Middle Korean form but its historical background is not so clear.

(FUKUI Rei)
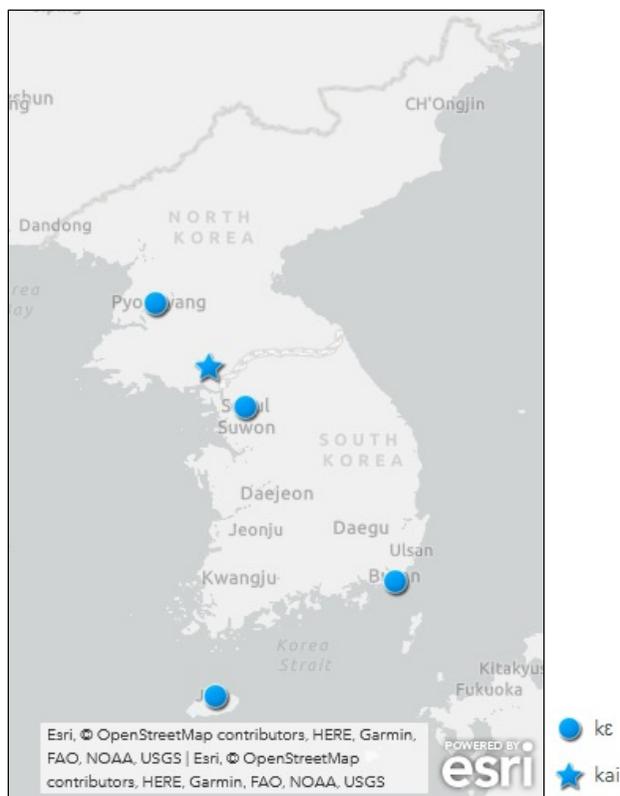
Figure 4.4.1: 'Dog' in Korean.

'Dog' in Sinitic

We classify the words based on stem types, then subclassify them based on types of suffix.

A-1: 犬 kʰɛiŋ (福州)

B-1: 狗 kou214 (北京) kɤ55 (上海) kau35 (恩州) kai (湘鄉) ki (績溪) tɕiɤ (婁底) tɕiɑu (岳阳柏祥)

B-2: 狗儿kəu53 ər11(太原) 狗儿 kɐur (离石)

B-3: 狗子kou tsɿ(武漢) ke tsæʔ (丹陽) kou tə (牟平) kieu tsɿ (于都) kau a (台北) e tsiɛ (建瓯)

B-4: 狗娃子kəu va tsəʔ (大同)

B-5: 狗囝 kau (k-)iaN (枫亭)

C: Others 来福 (武汉) 地羊 ti22 iɐŋ21 (南寧)

Monosyllabic 狗 are distributed in the whole China, and 狗 plus suffix types (B2~B5) are distributed mainly in central China.

A-1 犬 type is distributed in some southern dialects, and is considered as the older type.

Reconstructed forms of 犬 and 狗 for Middle Chinese and Old Chinese are shown below.

犬	1	2	3	4
MC	khiwen	khiwən:	khwenX	-
OC	khiwan	khiwən	*[k]ʷʰˁ[e][n]ʔ	*khwĭnʔ
狗	1	2	3	4
MC	kəu	kəu:	kuwX	-
OC	ko	ku	*Cə.kˁroʔ	*kŏʔ < *kloʔ

1:郭錫良(2010), 2: Karlgren (1957[1997]), 3: Baxter & Sagart (2014), 4: Schuessler (2007)

(YAGI Kenji)

A.
☆ A-1 犬kʰɛiŋ
B.
○ B-1 狗kou, kɤ, kau, kai, ki, tɕiɤ, tɕiɑu
● B-2 狗儿kəu ər, 狗儿kɐur
□ B-3 狗子kou tsɿ, ke tsæʔ, kieu tsɿ, kau a, e tsiɛ

□ B-4 狗娃子kəu va tsəʔ
△ B-5 狗囝kau (k-)iaN
C.
✹ C Others: 来福, 地羊ti iɐŋ
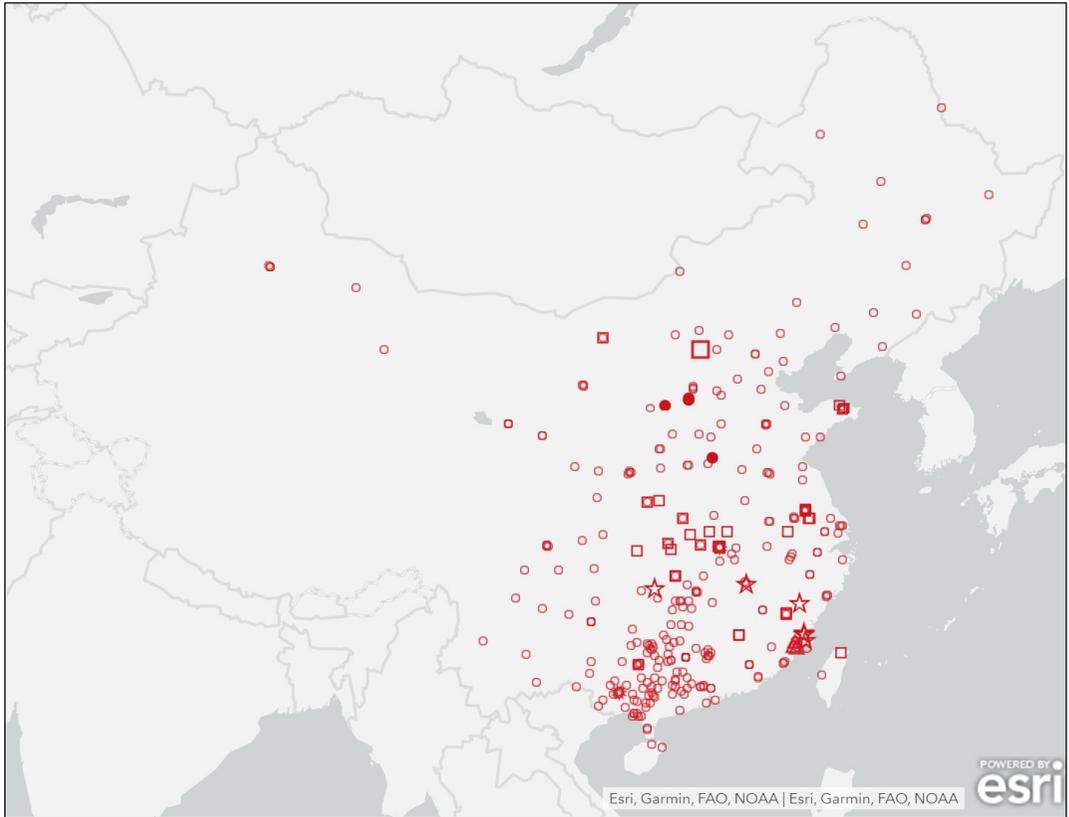
Figure 4.5.1: 'Dog' in Sinitic.
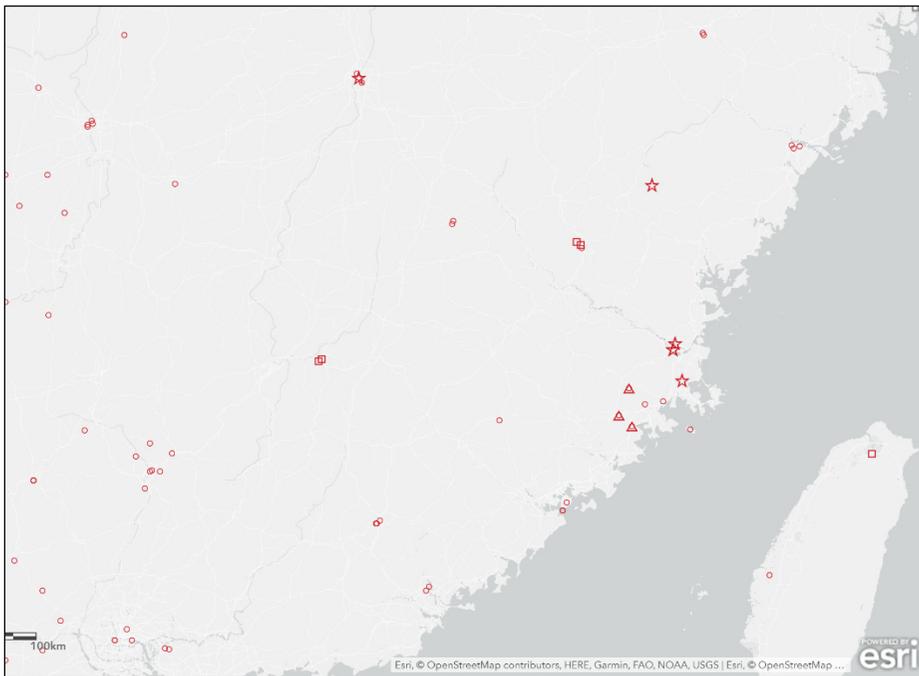
Figure 4.5.2: 'Dog' in Sinitic (Fujian province).

'Dog' in Hmong-Mien

There are two types in DOG: A: *klu*; B: *ljaŋ*.

There is only one major type in this entry: Type A. Type B is only observed in lects belonging to Pa Hng, a Hmongic language. As Type A is relatively similar to one of the Sinitic words denoting DOG, some scholars assume that this term is a loanword from Sinitic. If Type A is a loanword from Sinitic, Type B might represent the more archaic state. However, as the direction of borrowing is debatable, we cannot ascertain which type is more archaic.

(TAGUCHI Yoshihisa)
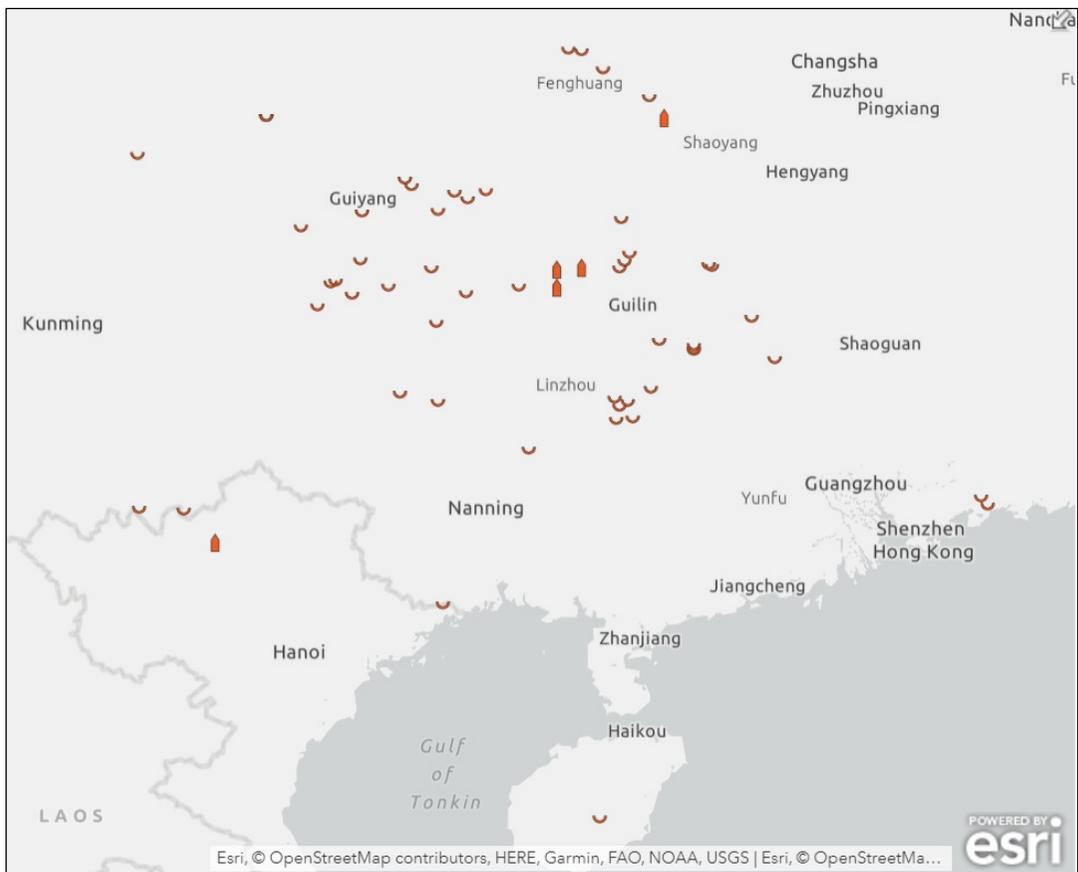
Figure 4.6.1: 'Dog' in Hmong-Mien.

'Dog' in Kra-Dai

Type A *ma* is widespread among all branches. The proto-Tai form by Li (1977) is *hma A1 preserved in Dong-Shui as type A2. Voiceless nasal initial consonants changed to ordinary voiced nasal ones in the Tai and Ong-Be branches. It is perfectly denasalized as p- or ph- of type A3 in Li. Types A4 to A10 show differences with a vowel or without a vowel.

Types B, C, and D have initial velar consonants. It is possible that they have a relationship with type A. Types E, F, and G have a velar initial. Type H is located in an isolated place in Hainan. Types J and K are distributed closely; thus, they can be treated as one type. Type I is similar to types J and K in terms of sound shape: however, it is located in a distant place; hence, their relationship is difficult to infer.

Figure 4.7.1: 'Dog' in Kra-Dai (enlarged).

(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)

A. *ma* type
- A1: m^{55}, $ma{:}^{A1}$, ma^1, $mɒ^1$, ma^{31}, $mā^6$, maa^1, maa^2, maa^5, mo^1
- A2: $m̥a^{53}$, $m̥a^1$, $tə\ m̥a^1$
- A3: pa^1, pa^4, pha^4, pou^4
- A4: $m̥ɯ^{31}$
- A5: $mpau^{33}$
- A6: mu^3
- A7: $dɔ^{33}hm̃^{46}$
- A8: $lɔ^{33}hm̥^{55}$
- A9: $ljou^{53}m̥^{44}$
- A10: $nu^3m̥^{45}$, num^{45}

- B. $ŋəɯ^{31}$
- C. $ŋwa^{1'}$
- D. $ŋ̊wa^1$
- E. $naŋ^{24}$
- F. $haŋ^{44}$
- G. $hɯŋ^{35}$
- H. $khak^5$
- I. $khuə̃{:}i^1$
- J. qoi^{24}
- K. $ʔu{:}i^{33}$

Figure 4.7.2: 'Dog' in Kra-Dai.

'Dog' in Tibeto-Burman

There are four major stems (word roots) for 'dog' in Tibeto-Burman (TB). Among these, Type A is remarkably widespread in terms of the diversity of its branches. Three of these stems are etyma of the proto-level forms of Proto-Tibeto-Burman (PTB; see STEDT). They contain word formations, which consist of a single stem, a stem plus an affix, or two compound stems. We first classify the TB word forms for 'dog' into stem types and then into their compound types.

The etymology of Type A is derived from *d-kʷəy-n (DOG) in the PTB etymon. This root is recognized as cognate with the Chinese 狗 *gou* (OC *ku) and 犬 *quan* (OC *k'iwən) (Matisoff 2003).

The etymology of Type B is *m-par ⪥ pra (WILD DOG/WOLF) in the PTB etymon. The etymology of Type C is *na (DOG) in the PTB etymon. This etymon contains other meanings in addition to 'dog' (WILD DOG/WOLF).

The above-mentioned PTB etyma also contain several meanings in addition to 'dog ', including *m-par ⪥ pra (WILD DOG/WOLF).

Type D *hapa* is borrowed from the modern Chinese 哈巴 (狗) *haba(gou)*, which means 'Pekingese' (a kind of dog species).

In addition to the four major types, there are some marginal roots, labelled as Type X. All of them are etymologically unknown. Most of the above-mentioned forms consist of a single stem (with an affix), but we also found two types of compound forms: A+C and C+A.

In some languages and dialects, two word forms coexist: A and C as well as A and D. In some dialects of Amdo Tibetan that have both A and D, A refers to relatively big dogs such as the Tibetan mastiff and its crossbreeds, whereas Type D is used for smaller ones.

Type A is the most widespread across the branches of TB. It is found in the northern and central-eastern parts of the TB area (Tibetic, rGyalrongic, and Qiangic groups), as well as in the southern part (Lolo-Burmese). Moreover, compounds with Type A (A+C and C+A) are found in rGyalrongic and Qiangic, in addition to some languages in Nepal (Manang, Tamang, and Thakali).

Type B is only found in the southwestern part of Yunnan province, in China (Lahu), and Type C is found in the northern part of Yunnan (Lisu, Yi, and Lipo) and in Nepal (Gurung). Type D is distributed in Qinghai, Gansu, and Sichuan provinces in China (mainly in Amdo Tibetan).

In terms of the relationship between 'dog' and 'wolf', there are no languages or dialects that synchronically colexify 'dog' and 'wolf' in our data. However, there are examples of modern languages/dialects where the proto-forms of Types A (*d-kʷəy-n 'DOG'), B (*m-par ⪥ pra 'WILD DOG/ WOLF'), and C (*na 'DOG') are all used for 'wolf'.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

▬ A. *d-kʷəy-n type

cʰi⁵³, ci⁵⁵, tɕʰə, khɯ³³, tshi³³, tshʐ̩¹¹khɤ³³, khi³³, khui⁴⁴, phɯ⁵³, kui³¹, tefü, kei, sĩ, ʔúy, ci, gùy, thwī, hi, thi¹¹, ²³džɪ, šəy, hui, í, ŭi, ʃĩ³³, fü za, thwì, wī, gui hen, ʔûy, ʃoi, hi, etc.; tɕʰəru, chi⁵⁵bo⁵⁵, khɯ³¹gɑ³¹, khɯ³³jo³³, khli-tśa, khi:bu, fhurro, khira, khinu, etc. (suffixed); ɔ³¹khɯ³¹, azü, mɔ⁵⁵khɯ²¹, etsü, lə³¹kha³⁵, ³ɲukyu, əgi³¹, atsü, etc. (prefixed)

◁ A+C: khe³³ne⁵⁵, khəna, khuɛ⁵³ŋi³³

△ C+A: 'nɔ kju, nā ki

╱ B. *m-par ⋊ pra type

phɯ³, phɤ³, phi³, etc.

○ C. *na type

nagi, nā ki , 'nakyn, etc. (suffixed); a³³no²¹, etc. (prefixed)

□ D. hapa type

haba, xapa, hapa, etc.

⌄ X. Others

wathi, kotsho, a-chak, wɔk, boh, nyi⁴ mɯː², tsuʔyu, etc.
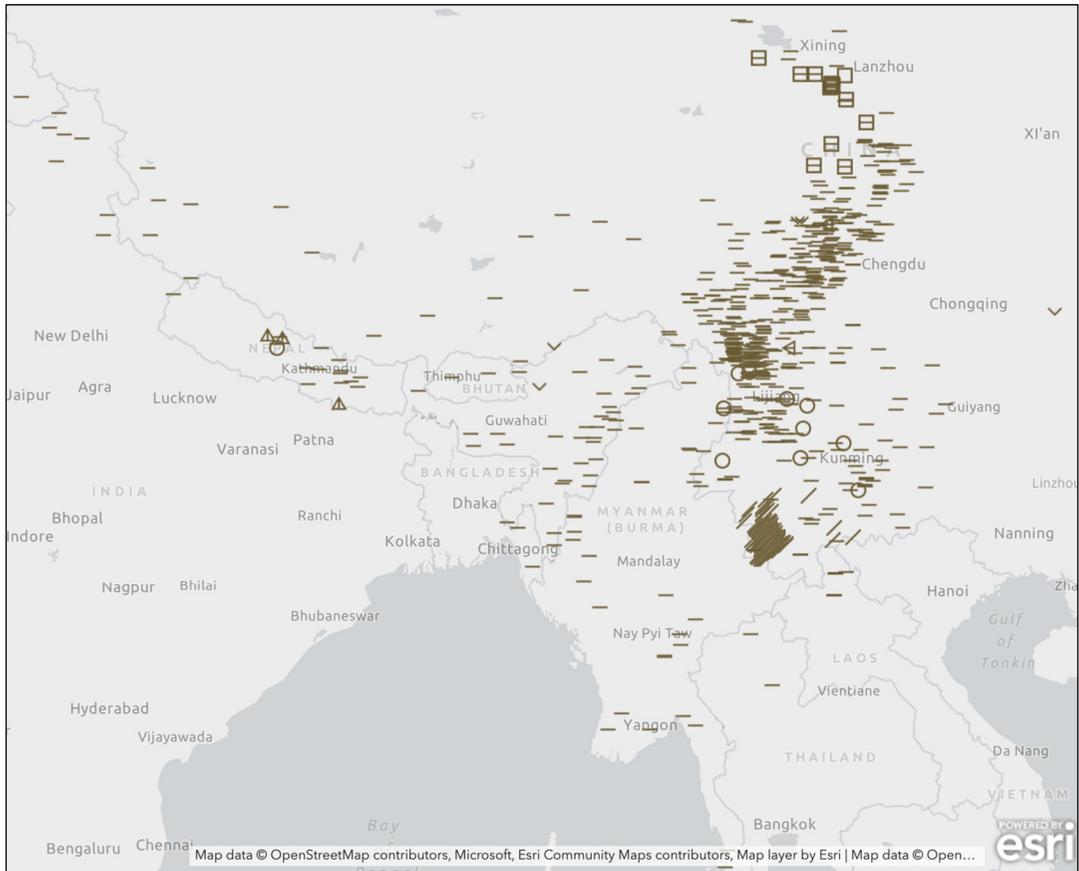
Figure 4.8.1: 'Dog' in Tibeto-Burman.
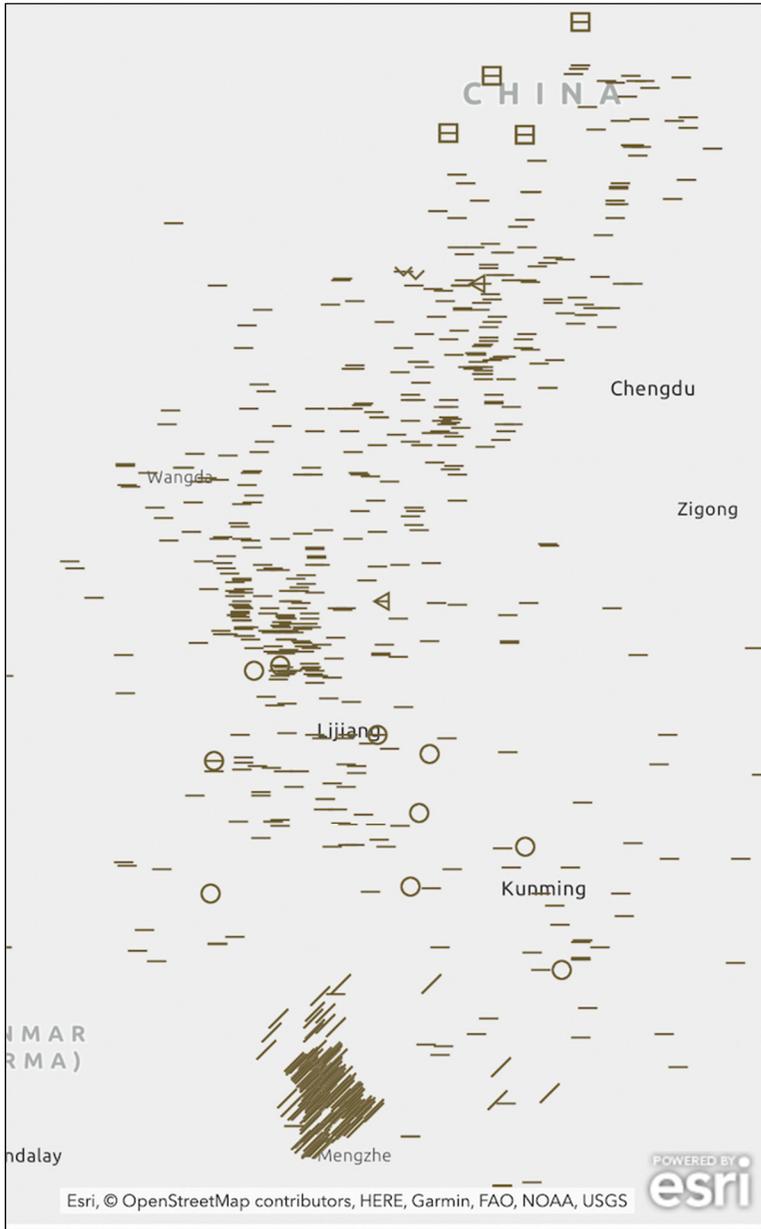
Figure 4.8.2: 'Dog' in Tibeto-Burman (detailed).

'Dog' in Austroasiatic

The word forms meaning "dog" in Austroasiatic are classified into eight types, as follows.

A1. *cɔːʔ type*

Proto MK: *cɔːʔ (Shorto 2006)

Proto Bahnaric: *cɔːʔ (Sidwell 2011); choo (Tampuan)

Proto Khmuic: *cɔːʔ (Sidwell 2013); chɔɔʔ (Mlabri)

Proto Palaungic: *cɔːʔ (Sidwell 2010); tsoʔ[1,3] (Danaw)

Proto Pearic *c(ɔ)ʔ (Headley 1985); chɔɔ (Chong)

Mangic: tsu[53] (Mang)

Proto Katuic: *ʔacɔɔ (Sidwell 2005); (ʔa)cɑː (Kui), ʔa.cɔː (Pacoh)

Proto Vietic: *ʔa-cɔːʔ (Ferlus 2007); acɔː[3] (Chứt [Rục]), chó (Vietnamese [Hanoi])

Aslian: cɔ (Semelai)

A2. *ksəw type*

Proto Khasic: *ksəw (Sidwell 2012); ksaw (Pnar [Jowai]), ksia (War [Amwi])

Munda: gusoʔ (Bondo)

Proto Waic: *sɔʔ (Diffloth 1980); so (En), so (Kentung-Wa)

Bahnaric: so (Sre)

B. *cgəy type*

Proto MK: *cgəy (Shorto 2006)

Khmeric: ckəe (Khmer)

Vietic: cầy (Vietnamese [Hanoi])

C. *mrok type*

Proto Waic: *mrok (Diffloth 1980); maruk (Wa), mbrɔk (Lawa [North])

D. *haɖi(ɟ) type*

Munda: *haɖi(ɟ) (Proto Kherwarian: Munda 1968); haɽiɟ (Santali)

E. *sɛta type*

Munda *sɛta (Proto Kherwarian: Munda 1968); seta (Santali [Singhbhum])

F. *ʔam type*

Nicobaric: ʔam (Car)

G. *chúr type*

Monic: chúr (Nyah Kur)

H. *klə type*

Monic: klə (Mon)

Shorto (2006) reconstructed two proto MK forms (A1 and B), in which the A1 form *cɔːʔ is succeeded by Bahnaric, Khmuic, Palaungic, Pearic, Mangic, Katuic, Vietic and Aslian, while the B form *cgəy is succeeded by Khmeric and Vietic.

The A2 form *ksəw* has two subtypes: *ksəw (Pnar [Jowai] and War [Amwi]) and *sɔʔ (En and Kentung-Wa). The Munda form gusoʔ is intermediate between these two forms, since it preserves the first element of the initial consonant cluster ks- of Khasic forms and the vowel plus glottal stop -ɔʔ of Waic forms.

It is highly possible that the A2 form *ksəw* is older than the A1 form *cɔːʔ*, since the sesquisyllabic *ksəw* is distributed from the mainland of Southeast Asia to the Indian East coast, while the monosyllabic *cɔːʔ is distibuted between the Northern and the Southern areas of the mainland of Southeast Asia..

The C type *mrok* is common to the Khmuic words for "wolf," which contain *-bru(ə)k (Proto Khmuic, Sidwell 2013), hmbrok (Khmu [Cuang]), mpjuak (T'in [Mal]) and phluk (Khsing-Mul).

(SHIMIZU Masaaki, MINEGISHI Makoto)
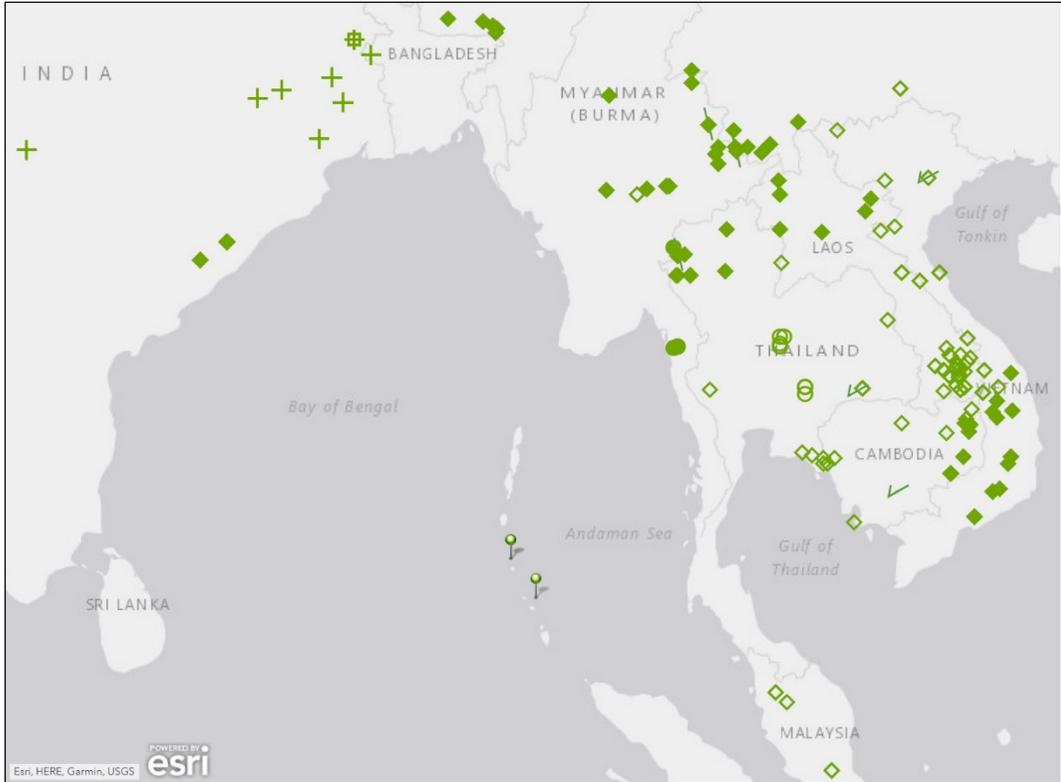
Figure 4.9.1: 'Dog' in Austroasiatic.

◇	A1	*cɔːʔ type*
◆	A2	*ksəw type*
∠	B	*cgəy type*
\	C	*mrok type*
▫	D	*haɖi(ɟ) type*
+	E	*sɛta type*
⚑	F	*ʔam type*
◓	G	*chúr type*
●	H	*klə type*

'Dog' in Austronesian

The lexical item for "dog" in Austronesian exhibits a rich variety, and it is difficult to find similarities within one third of the languages studied here.

Types A and B occur frequently. One example of a typical Type A form is /asu/, which begins with /a/ and is disyllabic. The onset of the second syllable is an alveolar consonant and is typically /s/, which is often substituted by /h/ and sometimes by /t/ in Austronesian languages. Examples of this type include *avʔu* (Tsou), *vatu* (Paiwan), *ātu* (Isnang), *āsu* (Kalinga Limos), *āso* (Tagalog), *asɛə* (Acheh), *asu* (Javanese, Da'a, Bugis, Konjo, and Buru), *atʃu* (Manggarai), *ahu* (Sika), and *anɔɔ* (Buang). Type B consists of words that begin with /k/ or /h/ and are mostly disyllabic, such as: *kui* (Lau), *kuʔi* (Kwaio), *huli* (Paamese), *kuri* (North Tanna and Kawamera), *kiru* (Marshallese), *kiti* (Ponapean), *kolī* (Easter Fijian), *tʃui* (Western Fijian), *kulī* (Tongan), *korī* (Mele-Fila), and *ʔurī* (Tahitian), *lokuli* (Lewo). Type A is spread over a large area, including Taiwan, the Philippines, Sumatra, Java, Sulawesi, and Maluku, whereas type B is exclusively found in the Solomon Islands and Pacific Islands.

Type C forms typically begin with /a/ and have a word-final /ŋ/. Sundanese exhibits a typical form, which is /aɲʤiəŋ/. Other forms of this type include *biaŋ* (Batak Toba), *aɲʤiəŋ* (Minangkabau, Sundanese), *anʤiŋ* (Indonesian), *tʃitʃiŋ* (Balinese), and *atʃoŋ* (Sasak). Type C appears in Indonesia, especially on the islands of Sumatra and Java.

Type D forms have a word-initial /k/ but are distinct from Type B forms: *kaukʷa* (Kilivia), *kadewa* (Tawala), *kʰuma* (Maringe), *te kamea* (Kiribati), and *kɔmia* (Rotuman). This type occurs in Papua and the Solomon Islands. Type E words begin with a vowel and end with /yam/, such as *ayam* (Kagayanen), *ayəm* (Sarangani Blaan), *āyam* (Aklanon), and *iyam* (Adzera). This is found in the Philippines and Papua. Type F has trisyllabic forms that begin with /k/ or /ʔ/: *ʔapula* (Gorontalo), *kapuna* (Bantik and Ratahan), and *ʔappunna* (Talaud). This type mostly spreads in North Sulawesi.

Type G: Words with a word-internal /d/. such as *indaŋ* (Palawan), *idoŋ* (Molbog), and *edoʔ* (Bangingi Sama). They are found in southern Philippines. Type H consists of words with a word-initial /g/: *goun* (Takia) and *gãũ* (Dami). They are found in Papua.

Other forms are categorized into Type I. Examples are: *xuyił* (Atayal), *tawpuŋo* (Rukai), *ino* (Yammi), *ukuʔ* (Murut), *alika* (Malagasy Merina), *patiʔ* (Madurese), *dikeʔ* (Uma), *mantoa* (Wolio), *lako* (Ngada), *busa* (Roti), *kʷoyar* (Dobel), *ɸunə* (Irarutu), *yaw* (Sawai), *mʷi* (Nyindrou), *ʔeu* (Manam), *me* (Mbula), *keam* (Yabem), *elβa* (Kaulong), *pap* (Tolai), *sisi* (Motu), *amuʔe* (Mekeo), *siki* (Roviana), *vʷiriu* (Raga), *lipax* (Port Sandwich), *cawek* (Nemi), *wɔta* (Cémuhî), *lōva* (A'jië), *taiki* (Xârâcùù), *pailai* (Nengone), *xerɔɔxü* (Woleaian), *maile* (Samoan), and *paiheŋa* (Rapanui).

(UTSUMI Atsuko)

⌣   A /asu/ type: *avʔu, vatu, ātu, āsu, āso, asɛə, asu, atʃu, ahu, anɔɔ*

▷ F: *ʔapula, kapuna, ʔappunna*

◢ B: *kui, kuʔi, huli, kuri, kiru, kiti, kolī, ʧui, kulī, korī, ʔurī, lokuli*

◨ G: *indaŋ, idoŋ, edoʔ*

∧ C /aɲdʒiəŋ/ type: *biaŋ, aɲdʒiəŋ, aɲdʒiŋ, ʧiʧiŋ, aʃoŋ*

▬ H: *goun, gãũ*

◤ D: *kaukʷa, kadewa, kʰuma, te kamea, kɔmia*

✚ *xuyiɫ, tawpuŋo, ino, ukuʔ, alika, patiʔ, dikeʔ, mantoa, lako, busa, kʷoyar, ɸunə, yaw, mʷi, ʔeu, me, keam, elβa, pap, sisi, amuʔe, siki, vʷiriu, lipax, cawek, wɔta, lōva, taiki, pailai, xerɔɔxü, maile, paiheŋa*

◉ E: *ayam, ayəm, āyam, iyam*

Figure 4.10.1: 'Dog' in Taiwan and the Phillipines.

Figure 4.10.2: 'Dog' in Indonesia.

Figure 4.10.3: 'Dog' in Papua and the Pacific.

'Dog' in Tungusic

Word forms for 'dog' would be classified in two types as below:

A NGIN-:Ewen *ŋin,* Evenki *ŋinakin* Ewenke *ninakin ~ ninaxin,* Hezhe *inaki,* Nanay *inda etc*

B Sibe *jɔnχun*

As shown most of Tungusic have common word for dog, *ŋin ~ nin.* Evenki uses *ŋinakin* adding the diminutive suffix *-kin* to that form. In some other languages the initial consonant should have changed to *n-* fron *ŋ-.* Only Sibe has the different form, which would be borrowed from the neighboring languages.

Between the words for wolf and dog, any similarity is not observed.

(MATSUMOTO Ryo)

	A NGIN-
●	B *jɔnχun*
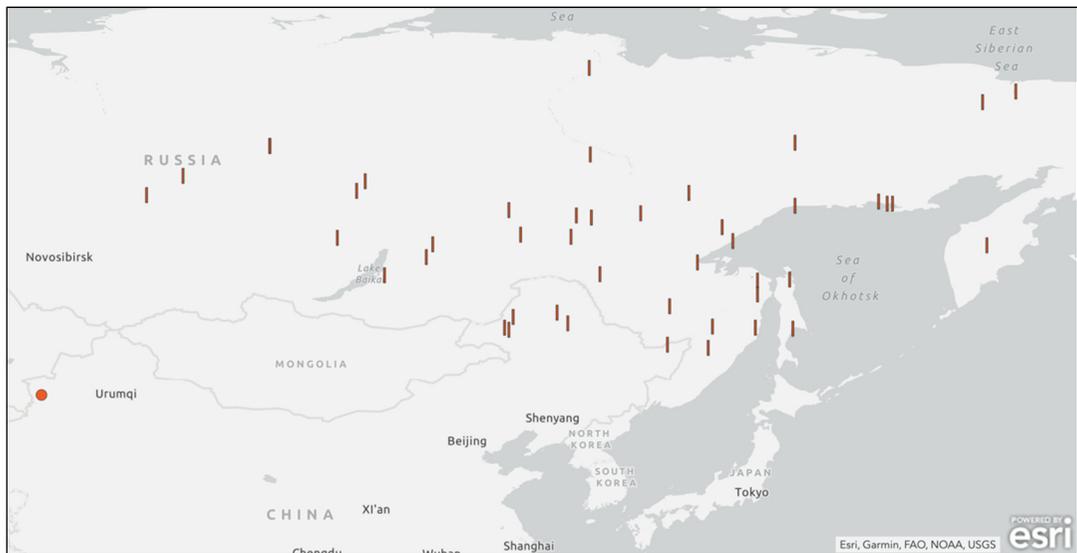
Figure 4.11.1: 'Dog' in Tungusic.

'Dog' in Uralic

In Uralic there are various forms for 'dog'. Here I classified according to the sound forms, it could not be referred to the reason why such many forms they have.

A KOIRA: Finnish *koira*, Karelian *koiru*, Ingrian *koira*, Votic *koira*, Estonian *koer*, Veps *koir*

B *P-: Mordvin *pine*, Komi *pon*, Sami *beana*, Livonian *piņ*, Mari *pij*, Udmurt *puny* Enets *bunyk*

C Hungarian *kutya*

D Khanty/Mansi *amp*

E WE: Tundra Nenets *weh*, Forest Nenets *wedjaku*

F Selkup *kanak*

Type A is only observed in Balto-Finnic languages, and regarding the area as the center B type is distributed widely around them. It would be presumed that Type B is more archaic form and was displaced by Type A. The other forms are characteristic in each language, they would have their own origin – borrowing, onomatopoeia, etc.

(MATSUMOTO Ryo)

╲	A KOIRA		✚	D *amp*
◎	B *P-		☐	E WE
☆	C *kutya*		Ψ	F *kanak*
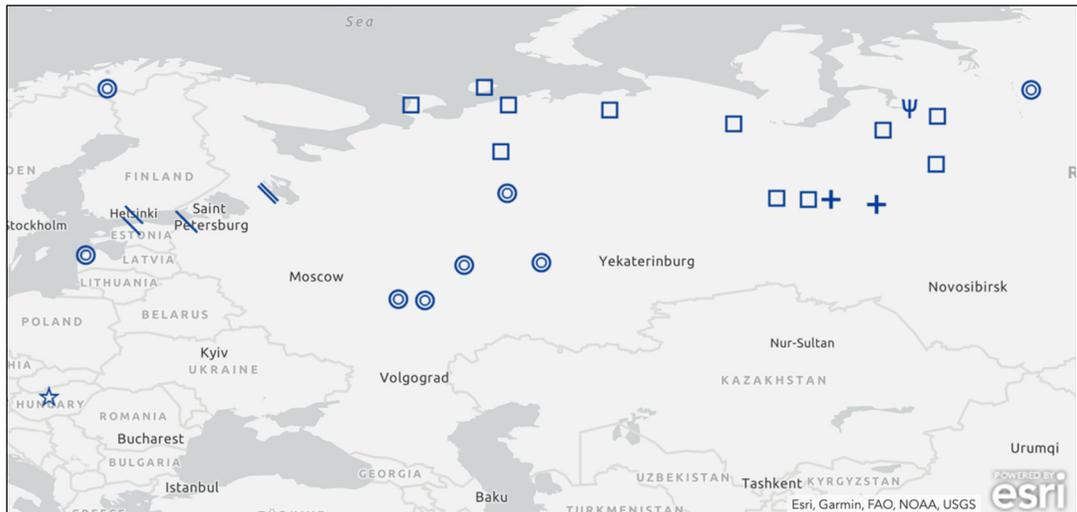
Figure 4.12.1: 'Dog' in Uralic.

'Dog' in Mongolic and Turkic

1. Mongolic
All Mongolic languages use *noxoi*-type words.

2. Turkic
In Turkic, *it*-type words are widespread.

The form *köpek* is found in the southwestern region (Turkish, Azeri, Gagauz, etc.).

The form *adaj* is used in southern Siberia (Chulym, Khakas, and Shor).

(SAITÔ Yoshio)

A. *noxoi* type

⌒ noxoi, nɔxō, noxœ̄, nœʼœ̄, noxā, noχɢui, noɣəi, noɢoi, noxuai, nokai, noɣʷ, nox, nokoi

B. *it* type

∠ it, īt, ït, ɪ̄t, et, itʼ, ijt, išt, əšt, id, əht, ïʼt, jïtă

C. köpek

▭ köpek

D. adaj

ʌ adaj

Figure 4.13.1: 'Dog' in Mongolic and Turkic.
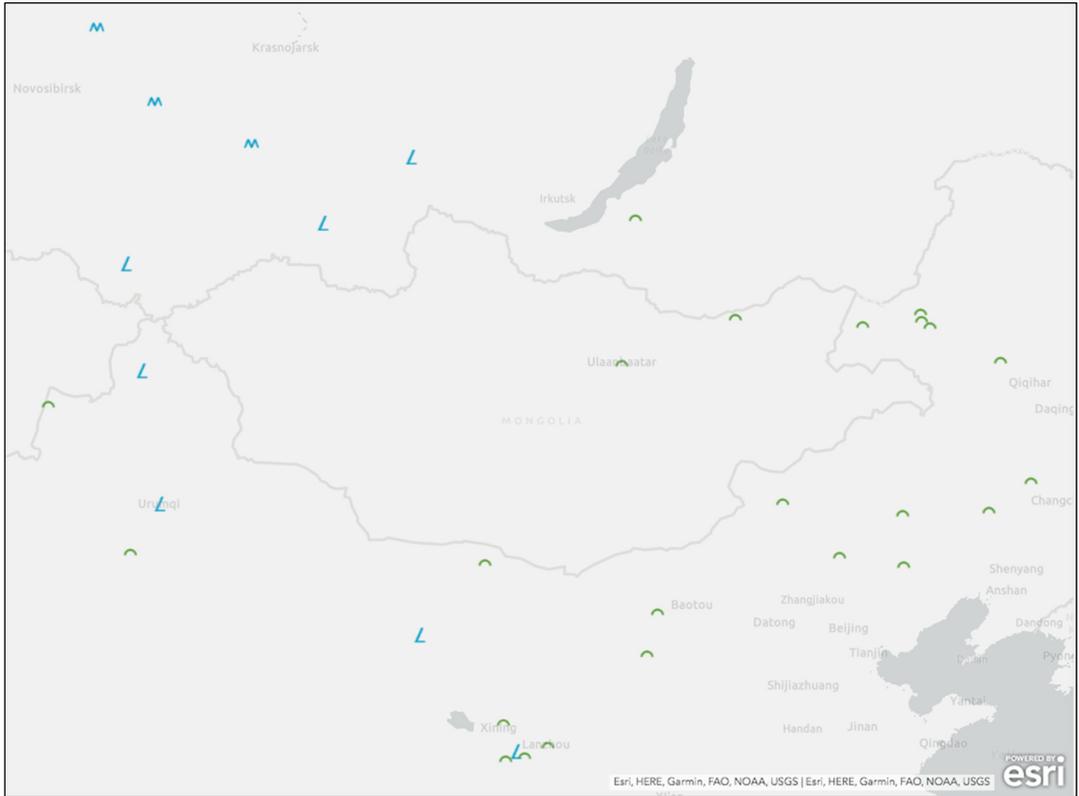
Figure 4.13.2: 'Dog' in Mongolic and Turkic (The Mongolian Plateau and its vicinity magnified).
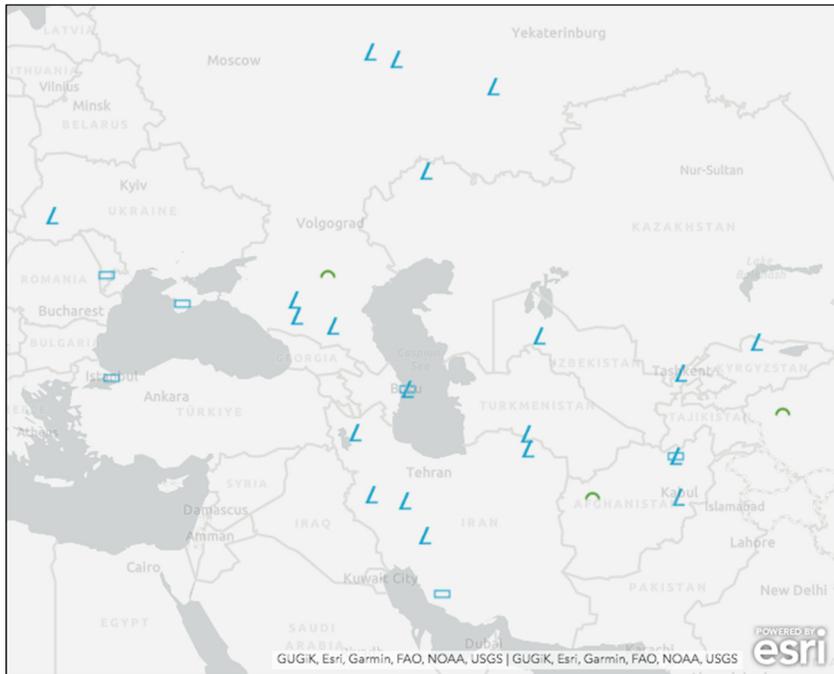
Figure 4.13.3: 'Dog' in Mongolic and Turkic (Central Asia to East Europe magnified).

'Dog' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. When a language has several words for dogs, I targeted 'male adult dog'.

IA and Nuristani languages in northern mountains, the west coast of India, Sri Lanka, Caucasus, and Jerusalem employ the type A. Type B can be seen in whole South Asia. The type C is detected in Vedda and IA languages in central Pakistan, northern India, Nepal, Bangladesh, Sri Lanka, and Aleppo. Type E is for European Romani lects and Type F is only for Burushaski lects.

The most major type is *śvan*. This type is derived either from Sanskrit *śvan* श्वन्, *śvaka* श्वक, or *śuna* शुन 'dog' (< *PIA *$św\acute{a}$ < PII *$ćw\acute{a}$ < PIE *$\acute{k}w\acute{o}$ 'dog' < pre-PIE *$\acute{k}wóns$). So this type is cognate with Latin *canis*, Ancient Greek *kúōn* κύων, English *hound*, and German *Hund*. The forms of most languages of IA are derived from *śuna*, while Waigali, a Nuristani language, has both from *śuna* and *śvaka*, namely *cũ* and *späi* respectively. The latter form is surely a loanword from Pashto, a Iranian language, in which they say *spay* سپی 'dog' (and *spəy* سپی 'bitch').

The *kutta* type appears only in IA languages, but the distribution is enough wide, from the north of Pakistan to Maldives, and from the south of Pakistan to the southeast of Bangladesh. Both the reconstructed Sanskrit form **kutta* *कुत्त and **kuttira* *कुत्तिर mean 'dog', derived from PIA **kúttas*, and PII **kúttas* (it can not go back to PIE). The types B, C, and D are

considered as onomatopoeic, as well as the PIE origin type A.

The third major type *kurkurá* can be seen widely in South Asia and Syria but not in the mountains. Next, the *kuccura* type i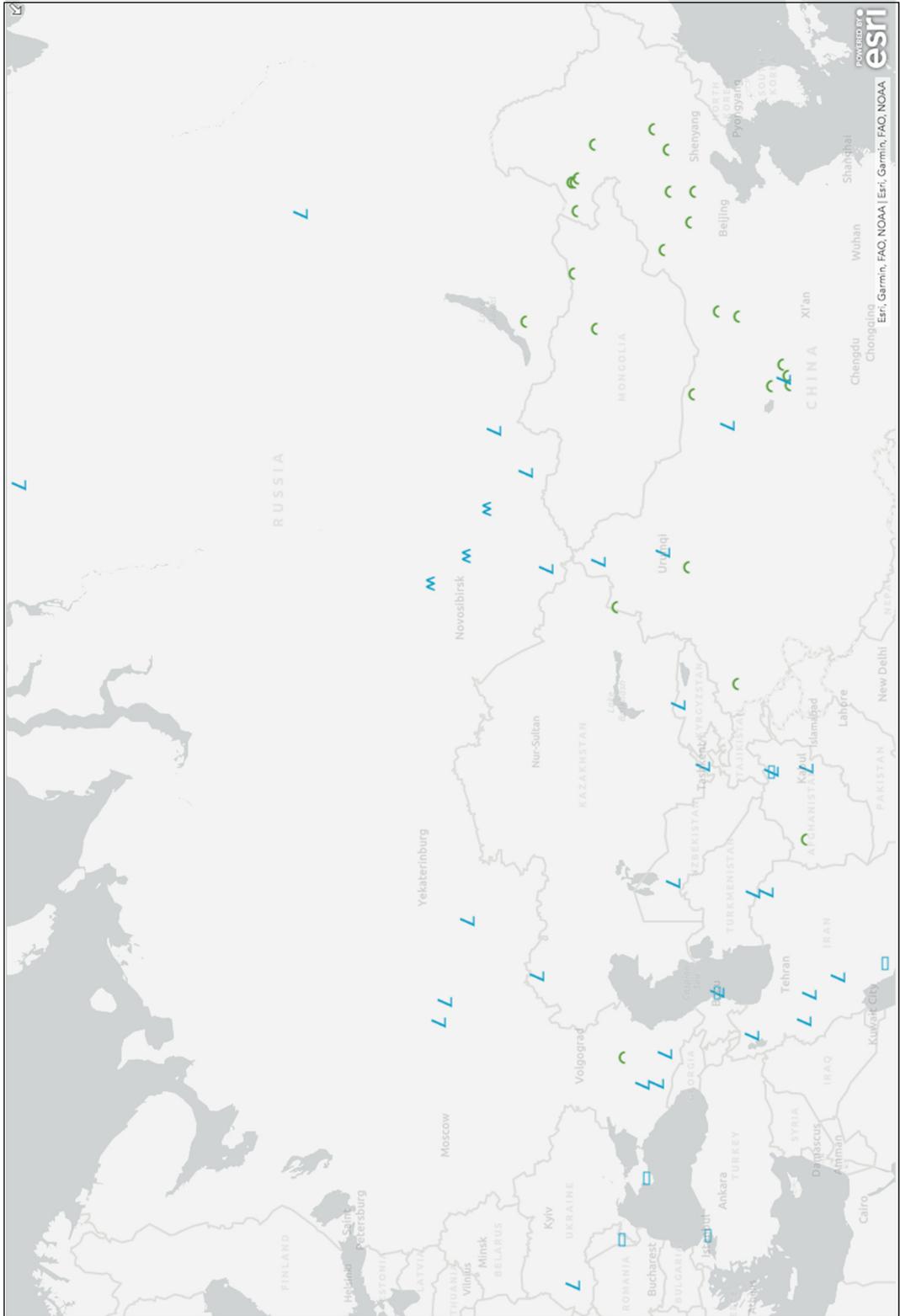s used sporadically in Bangladesh and around the India-Nepal border, and convergently in the east of Afghanistan and the northwest of Pakistan.

The *zhukel* type E is detected in Romanic languages. It is unclear what word of what language is the origin of this type. There are three major hypotheses for it, the first is the vein of Persian *jāhel* جاهل 'ignorant' (< Arabic *jāhil* جاهل 'id.'), the second is of Sanskrit *jukuṭa* जुकुट 'dog', and the last is of Proto-Kartvelian *$ʒ_1aɣl$- 'dog'. Anyway the Romani word of *zhukel* type has spread to other surrounding languages in Europe, such as Dutch *joekel* '(slang) dog', Macedonian *džukela* џукела 'mutt, cur, mongrel', and Croatian *džùkela* 'dog, mongrel'.

The type F of *huk* is for Burushaski. Only *huk* 'dog' and *urk* 'wolf' (borrowed from Iranian) have in common, and exclusively, the plural suffix *-ái*, while there are so many kinds of plural suffixes in Burusahski.

Besides them, languages in South Asia have some more words for 'wolf'. Sinhala *bállā* බල්ලා and Dhivehi *baḷu* ބަޅު are derived from Sanskrit *bhaluha* भलुह 'dog', which is made by *bhalluka* भल्लुक 'bear-like'. *bibi* is found in Jarawa and Bea in the South Island of the Great Andaman in common, while the Mixed Great Andamanese language employ a quite different form *čao*. Nihali *nāy* seems to be from Proto-

Dravidian *naH-ay* 'dog'. And Khowar   (YOSHIOKA Noboru)
reéni is inherited from Sanskrit *rāyaṇika*
रयणिक 'barking'.

A. *śvan / śvaka / śuna* type (28) 〇

 [*śuna*]

   šū́nā, šunó, šinó, šṹṛə, šuríŋ, šoṛíŋ, šḗṛ, šṍa, šūŋg, šŭ, ṣuŋ, ṣŭ, ṣva, sūnā, sənā, suno, sūṇe, suṇẽ, sunakhayā, senuta, snṍta, cunā́, cŭ̃, hūn

 [*śvan*] suvan; [*śvaka*] späi

B. *kutta / kuttira* type (26) ◁

 [*kutta*]

   kuttǎ, kuttɔ, kutǎ, kuto, kutawā

 [*kuttira*]

   kuttar, kutar, kutrā, kutru, kutro, kūtryō

C. *kurkurá* type (21) ➛

   kukurā, kukurɔ, kukyry, kukro, kŭkur, kŭkar, kūkr, kukkaa, kṳ̃r, xugùr

D. *kuccura* type (20) ⴸ

   kucuro, kucurɔ, kucuru, kucur, kucor, kučurŏ, kučur, kusur, kūsar, kuyu, kũir, kụi, koi, kuṛi, kuři, křui, kürog

E. *zhukel* type (6) ⟩

   žukel, džukel, żukel, žukhlo, žukal, čukel

F. *huk* type (3) ⤸

   huk

G. others

   [*bhaluha* type (2)] bállā, baḷu; [*bibi* type (2)] bibi; aagai, čao, nāy, reéni, selɔk
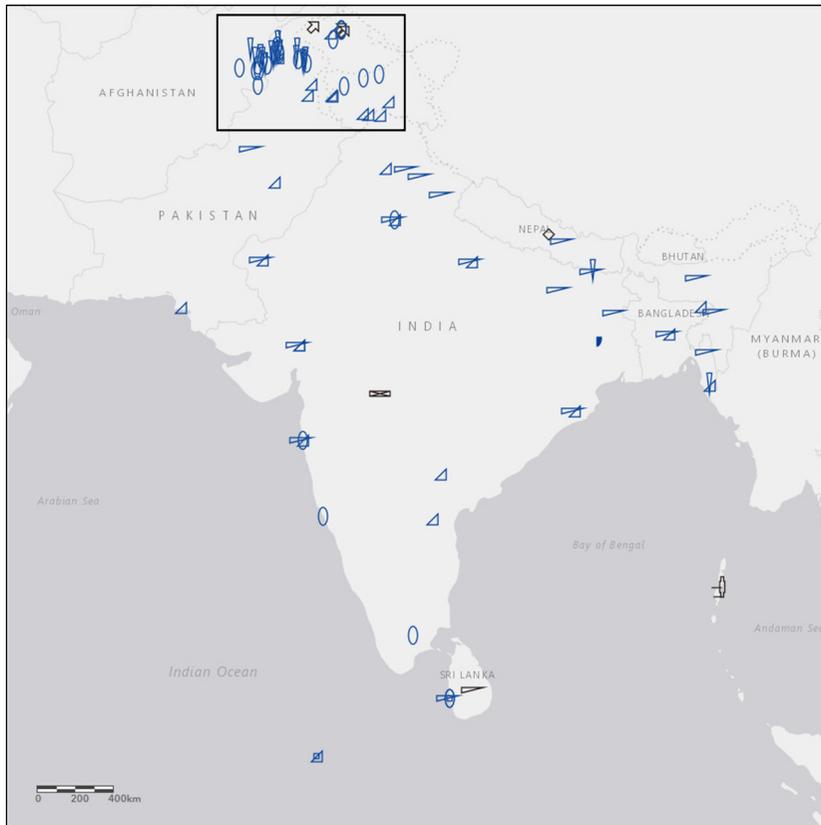
Figure 4.14. 1: 'Dog' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).
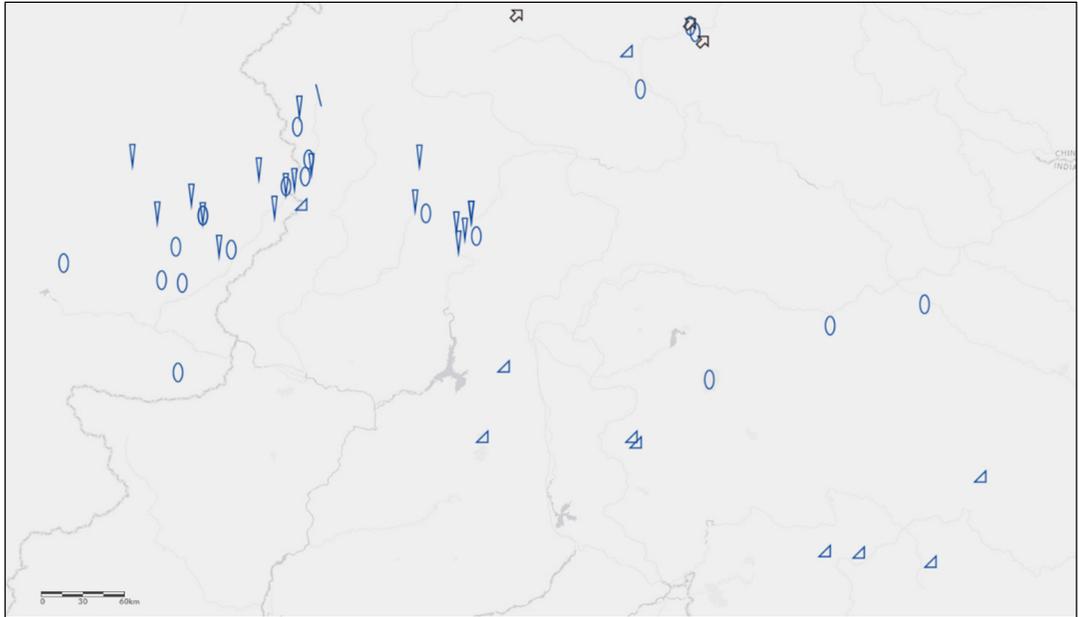
Figure 4.14.2: 'Dog' in northern Pakistan (the area encloed by the rectangle in Figure 4.14.1).
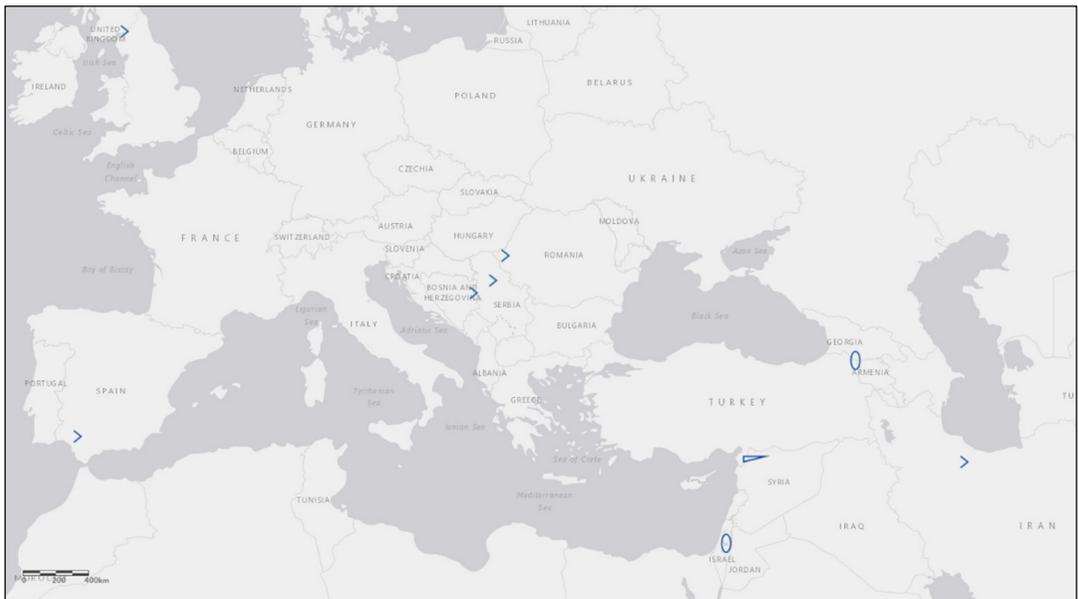
Figure 4.14.3: Types for 'Dog' in Indo-Aryan languages outside South Asia.

'Dog' in Dravidian

The possible Proto-Dravidian etymon NAY (DEDR #3650) usually covers the genera *Canis* (dog, wolf, jackal) and *Cuon* (dhole), although reflexes of another etymon NARI (DEDR #3606) 'a jackal, a fox' are distinguished from 'a dog' in most South, South Central and Central Dravidian languages.

Reflexes of NAY without an affix is retained in all the three subgroups mentioned above. Suffixed forms (N)A-TE and NE-KUDI are respectively distributed in Central and South Central Dravidian.

Kurukh alla: and Malto ale are reflexes of an isolated etymon. DEDR (#1796) relate Classical Tamil verb kurai-'to bark' to Telugu *kukka* 'a dog', which is identical with Pkt *kukka* 'dog', (< Skt *kurkura* of non-IE origin). Brahui employs a borrowing from Iranian.

(KODAMA Nozomi)

⌄ nāy, nāyi, nay

↙ noˑy

⌣ ney

✳ nēte, nette, netta

+ aˑte, (n)āte

∨ neh'ūri, nehuḍí, nih'uṛi, nehˀuṛi, nakuṛi, nekuṛ

∧ nukuṛi

▱ ale, allā

△ kukka

◒ kućāk < Iranian

Figure 4.15.1: 'Dog' in Dravidian.

'Dog' in Iranian

Type A is distributed in the vast area, mainly in Western Iranian. Etymologically, it derived from PIr. *tsuān-/ tsun-*. Type B came from PIr. *kuta-*, is concentrated on the edge of the Iranian world.

This type of the word can be observed in other western Iranian languages meaning 'puppy'.

Type C is now limited to Munji-Yidgha, which derived from PIr. *gadu̯a-*, also attested in Old Iranian, Avestan gaðwa- '(a kind of) dog'.

Type D is found only in Gorani/ Hawrami. It might have a link with Type B, assuming that the onset assimilated to the second consonant.

(IWASAKI Takamasa)

A: *span* type
B: *kut-* type
C: *ɣalv* type
D: *tuta* type
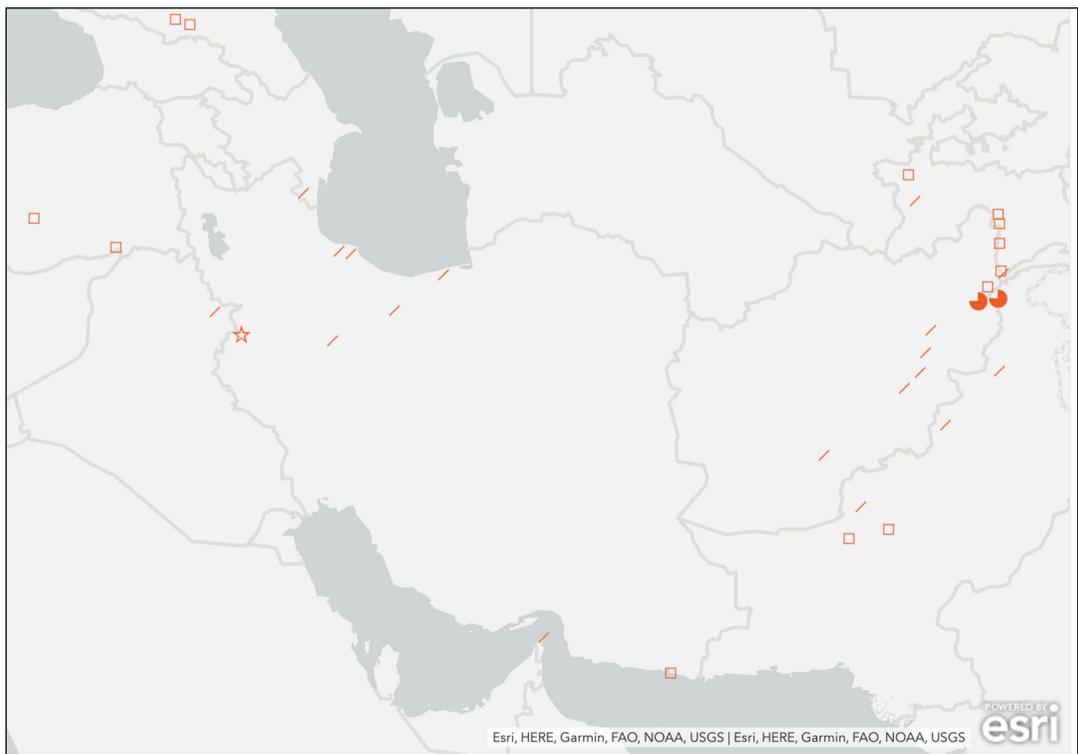
Figure 4.16.1: 'Dog' in Iranian.

'Dog' in Caucasian languages

Type A appears in Kartvelian and Nakh languages, of which a Proto-Kartvelian form can be reconstructed as *dz$_1$aɣl-. Type B is found in Abkhaz and Abaza, and Type C is found in the other Abkhazo-Adyghean languages. Type D is widely attested among Dagestanian languages. Types E and F are loans from Iranian and Azerbaidjani, respectively.

Type D is classified into two subtypes according to the voicing of the initial consonant. Type D1 appears widely in the Dagestan area, whereas Type D2 is distributed on a limited basis at the contact zone with Kartvelian languages.

Type F is presumed to be a loan from an Iranian language and is potentially a form borrowed from Ossetian, spoken near Lak and Archi.

(SUZUKI Hiroyuki)

● A: dz-type; *dzaɣli*, *žeɣ*, *žˤäla*, etc.
○ B: la-type; *a-la*, *la*.
▻ C: h-type; *hə*, *wˤa*.
D: χw-type

△ D1: χw-type; *χʷoj*, *χʷe*, *hʷe*, *pħu*, etc.
▽ D2: ʁw-type; *ʁʷaj*, *wo*, *ʁʷe*, etc.
◖ E: kw-type; *kkʷačči*, *gʷači*, *kicʼ*.
▭ F: t-type; tula, tɨla.

Figure 4.17.1: 'Dog' in Caucasian languages.

'Dog' in Semitic

Almost all Semitic including Akkadian (*kalbu-m*) has *kalb* form. The following types are phonetic variants.

A. *kalb* type (●) is found in Arabic. Variants are *kälb* (Hassaniya Ar.), *kelb* (Maltese Ar.), *kilp* (Cypriot Ar.).

kəlb type (○) with *ə* (<*a*) is found in Arabic of Tunisia, Algeria, Morocco.

kalib type (▲) with *i* between *l* and *b* is found in Chadian Ar., Sudanese Ar. on the Sudan belt and *kélib* (Juba Ar.) In Ethiopic, both *kəlbi* (●Tigrinya ከልቢ) and *kalib* (▲Tigre) are found. But *kel* (◇ Nubi Ar.)

tʃalb (✚) with palatalized *tʃ* (<*k*) is found in Najdi Ar., Gulf Ar. and *tʃalib* (Iraq).

kalba (■) is Aramaic form with the ending *-a*: *kalba* (Koy Sanjaq), *kelba* (Jilu), *kalbo* (Turoyo). In Hertevin *b* > *w kalwa* (□).

kelev (✖) is Hebrew form with *b* > *v*.

kawb (▼ Mehri) and *koób* (Hobyot) in South Arabia changed *l* to *w*. And *kaʕb* (△ Jibbali) may be formed by changing *l* > *ʕ*.

B. *wiʃʃa* (★Amharic ውሻ) is the only form that is not *kalb*.

(NAGATO Youichi)

A *kalb* type
  ● *kalb, kälb, kelb, kilp, kəlbi*
  ○ *kəlb*
  ▲ *kalib, kélib*
  ◇ *kel*
  ✚ *tʃalb, tʃalib*

  ■ *kalba, kelba, kalbo*
  □ *kalwa*
  ✖ *kelev*
  ▼ *kawb, koób*
B type
  ★ *wiʃʃa*

Figure 1.18.1: 'Dog' in Semitic.

'Dog' in Nilo-Saharan

What follow are heuristically reconstructed (marked with a hash #) Nilo-Saharan roots for 'dog' (121 languages surveyed). We simplify diacritics and notations for the [±ATR] feature in the original data.

There are four roots attested across major Nilo-Saharan branches in addition to a few branch-unique roots. Types A, B and C may be related to Afroasi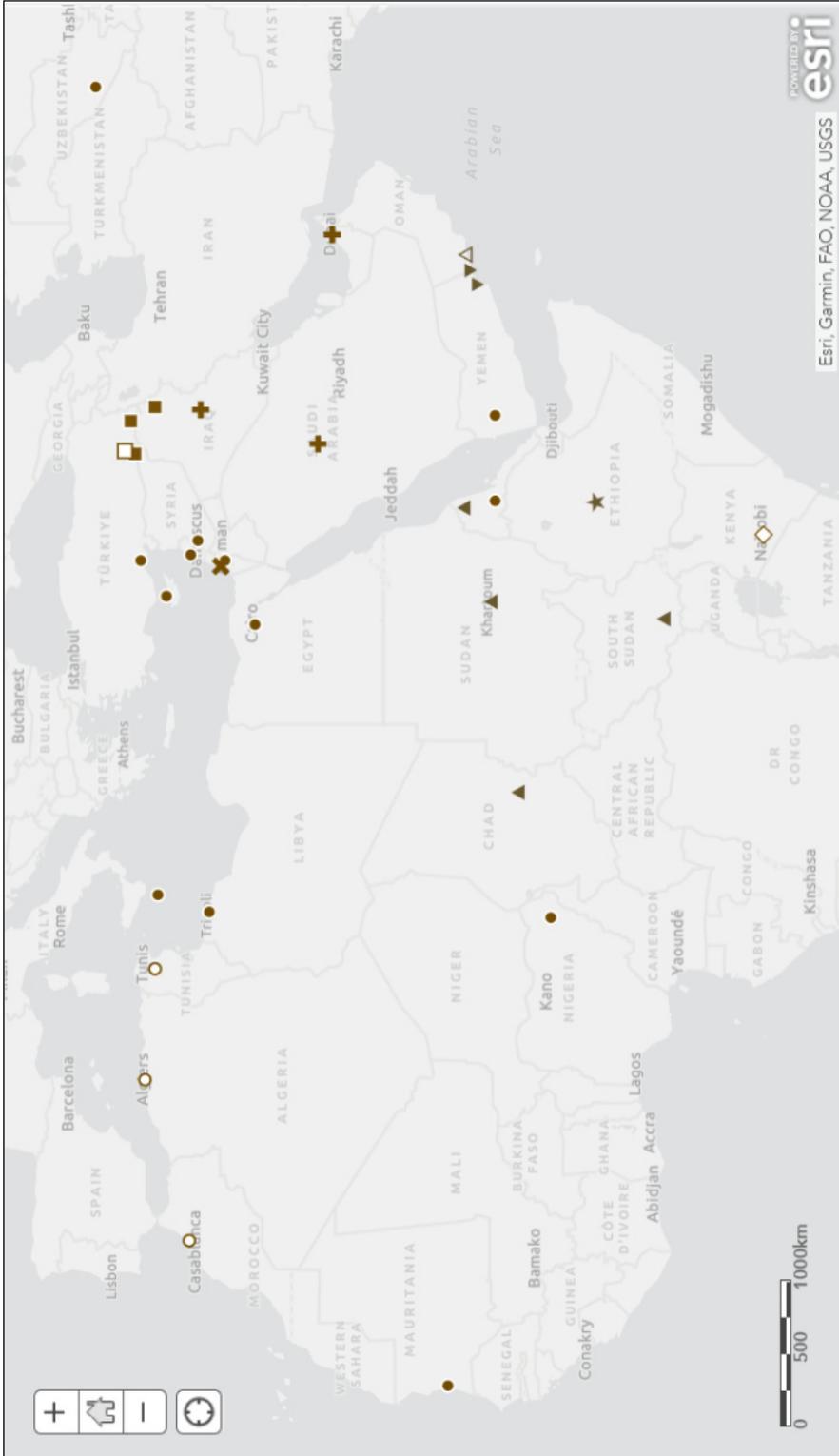atic and Niger-Congo roots. The existence of so many wandering words could be related to the fact that the dog is not native to Africa (Blench 2008b).

Type A1 *#bis(i)* and its possible variants, A2 *#bi* and A3 *#is(i)* are the most widely attested roots, represented in Taman (Abu Sharib *wis*), Daju (Dar Daju *iise*), Nara (*wos*), Nilotic (Nandi *seset*), Gaam (*aza*), Fur (*assa*), Central Sudanic (Lugbara *oce*, Bongo *bihi*, Ngambay *bisi*) and Songhay (Zarma *hansi*). This root was first noticed by Greenberg (1963), who coined the term 'Nilo-Saharan', although he compares only Songhay and Fur. Bender (1981) gives more extensive data set, including Ari *aksi* (Omotic, Afroasiatic) and Orig *wusu* (Kordofanian, Niger-Congo). Ehret (2001) reconstructs *'wɛns* for proto-Nilo-Saharan based on a similar set of cognates but explicitly excludes Fur. Blažek (2008) and Blench (2008b, #-*si*) compare them with some possible cognates from other African phyla, including Adamawa (Dza *iicwa*), Atlantic (Manjaku *u-bus*), Bantoid (Ndoro *siɛ*) and Benue-Congo (Nupe *efi*) in Niger-Congo, Semitic (Amharic *wəšša*, cf. Arabic *ʾaws* 'jackal'), Cushitic (proto-Highland East Cushitic *waša*), Omotic (Seze *wišši*) and Chadic (Bidiya *ʾusu*) in Afroasiatic. To these we could add Semitic Harari *buči*, Berber Tuareg *uššən* 'jackal' and Egyptian *wnš* 'wolf' (Leslau 1979, Lipiński 2001).

Type B1 *#gok*, B2 *#ŋok* and B3 *#goŋ* are attested in Nilotic (Turkana *ingok*, Kipsigis *ng'okto*, Päri *gwok*, Nuer *jiok*, Dinka *jioŋ*), Kuliak (Ik *ŋoka*) and less possibly Mabang (Maba *ñuk*) and Nubian (Nobiin *mug*). Bender (1981) was first to notice this isogloss. He compares with Zande (Niger-Congo, Ubangian) *ango*.

Type C1 *#kal*, C2 *#kan* and C3 *#kud* are attested in Eastern Jebel (Aka *kele*), Berta (*gali*), Nyimang (*gil*), Nilotic (Datooga-Buradiiga *gureera*), Saharan (Kanuri *kəri*, Tudaga *kudi*), Surmic (Koegu *kiani*), Koman (Gwama *kana*), Shabo (*kani*) and Central Sudanic (Kresh *kono*), Amdang (*kut*) and Kuliak (So *kudo'*). These seem comparable to the Afroasiatic isogloss for 'dog' as attested in Semitic (Arabic *kalb* with the alleged 'wild animal' suffix *-b*), Chadic (Hausa *karee*, Migama *kanya*, Mokilko *gede*), Berber (Guanche *cuna*), Cushitic (Saho *kare*, Konso *kuta*) and proto-Omotic *#kan(a)* 'dog' (Ehret 2001, Blažek 2008, Blench 2008b).

Type D *#bel/ber* is attested in Nubian (Kadaru *bol*, Midob *pəəl*, Birked *mel*, Kenuzi *wel*), Saharan (Beria *biri*) and less possibly Surmic (Majang *war*). Afitti *wil*, which we mechanically categorize as type D, might better be compared with Ama *gil* (Type C) of the same branch (Nyimang).

There are three branch-unique types, namely E *#k'aw/k'wa* (Gumuz and Koman, which might form a branch/family), F *#iira/eera* (Kadu) and G *#di* (Eastern Nilotic; Maasai *oldia*, Bari *'dioŋ*).

(NAKAO Shuichiro)
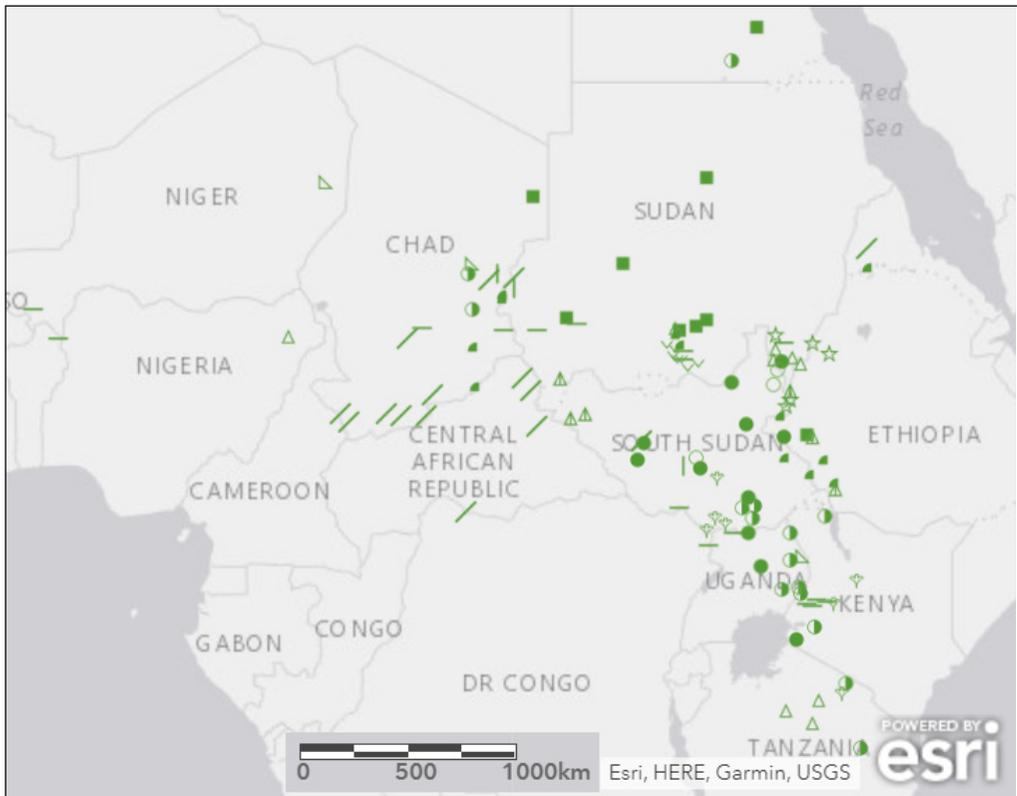
Figure 4.19.1: 'Dog' in Nilo-Saharan.

A. #(b)is(i)

    A1. #bis(i)

    A2. #bi

    A3. #is(i)

B. #gok/ŋok

    B1. #gok

    B2. #ŋok

    B3. #goŋ

C. #kal/kan/kud

    C1. #kal

    C2. #kan

    C3. #kud

D. #bel/ber

E. #k'aw/k'wa

F. #iira/eera

G. #di

H. The other types
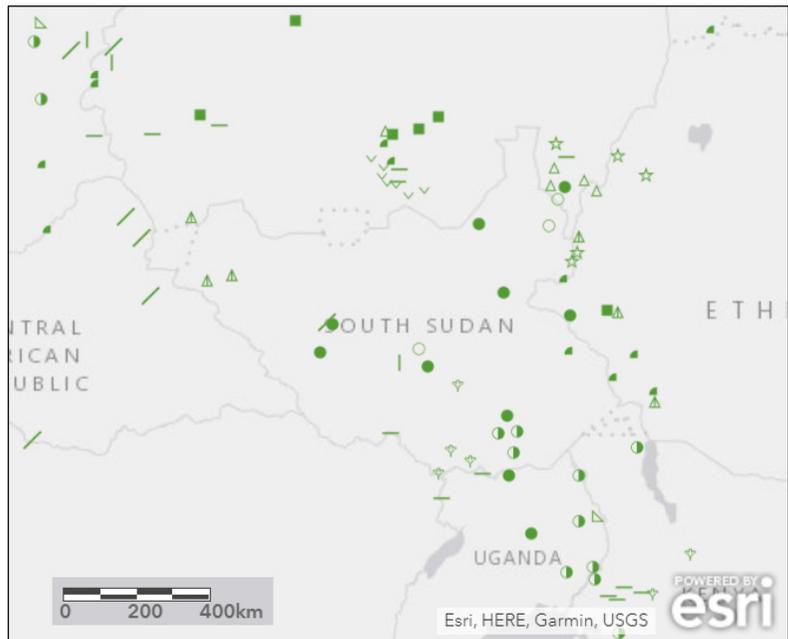
Figure 4.19.2: 'Dog' in Nilo-Saharan around South Sudan.

203

'Dog' in Bantu

The lexical distribution of 'dog' in Bantu is quite uniform. However, unlike 'horse', the common form distributed across the board is not a loanword introduced from outside but descendant forms of *-bʊ́à [BLR-MAIN-282], which is reconstructed by Meeussen (1969) as a main entry of his lexical reconstructions and its reflexes are attested in all 16 Bantu zones; e.g., in Duala [A24] mbɔ, Wanzi [B501] mvwéa, Lingala [C30B] límbwa, Mbole [D11] ŋma, Rundi [JD62] imbwa, Luhyia [E32] isimbwa, Sukuma [F21] mva, Kutu [G37] dibwa, Kongo-Central [H16b] mbwa, Lwena-Lubale [K14] katuwa, Lunda [L52] katuwa, Ndali [M21] kabwa, Manda [N11] libwa, Rufiji [P12] mbwaa, Herero [R31] ombwa, Venda [S21] mbɣa. Moreover, according to Blench (2007: 553), the PB form can further be traced back to the Proto-Niger-Congo and tentatively reconstructable as #-bu, which might have referred to 'jackal' at the stage.

Typical locally distributed forms include <kuri> in North Eastern Bantu languages including zones E, F, G, except for Kilimanjaro Bantu [E60], where <kite> are exclusively distributed, suggesting that they might be cognate.

(SHINAGAWA Daisuke and
KOMORI Junko)
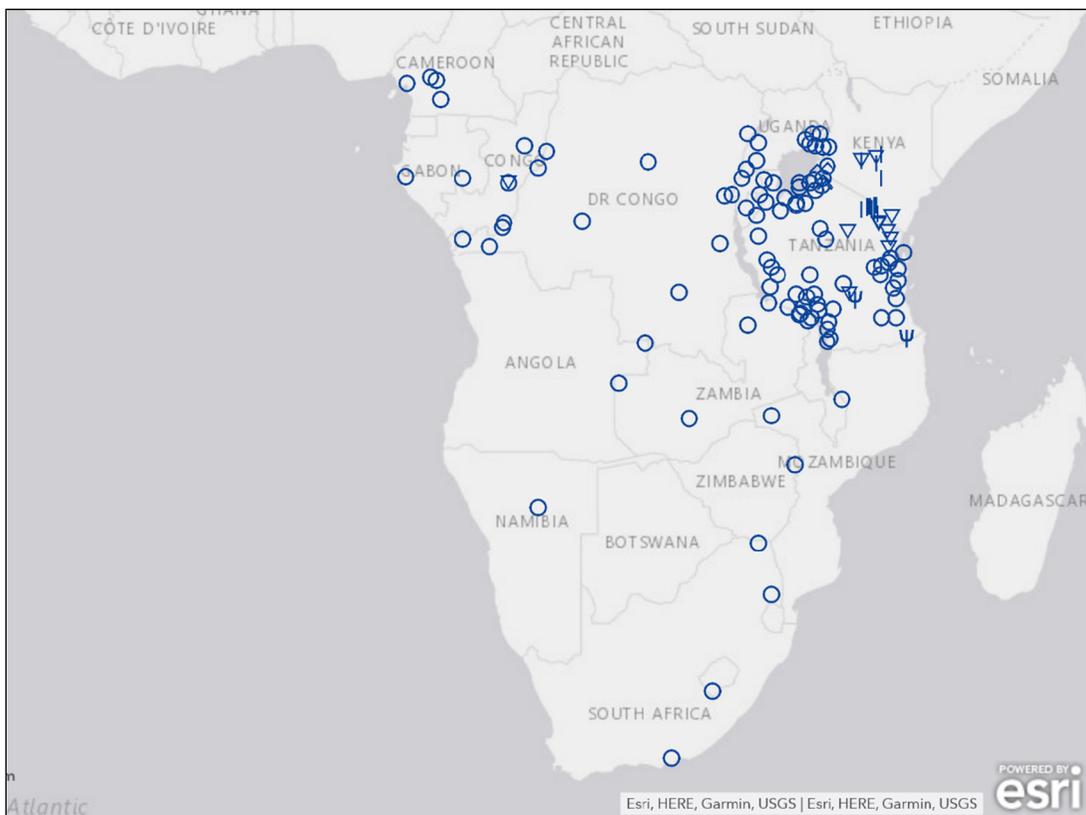
Figure 4.20.1: 'Dog' in Bantu.
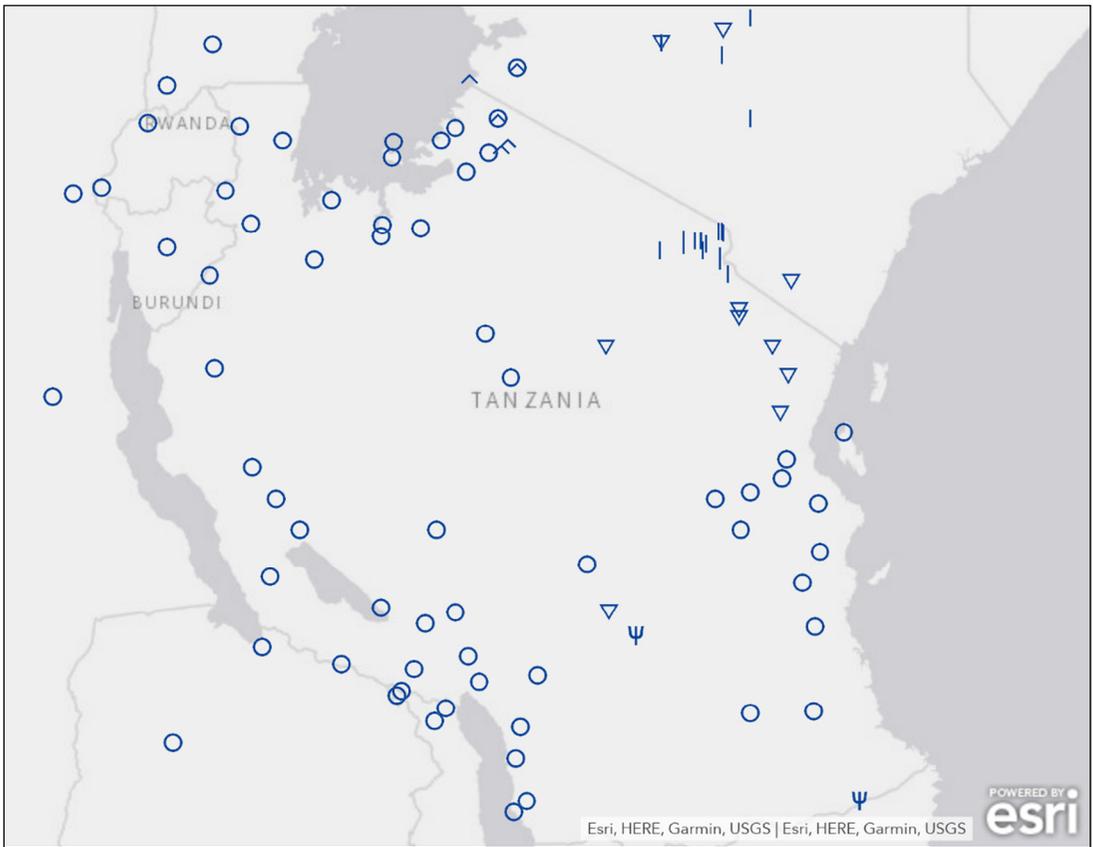
Figure 4.20.2: 'Dog' in northeastern Bantu zones.

'Dog' in the Kalahari Basin area

A total of seven types of the word "dog" are found in 15 sample languages from the Tuu, Kx'a, and Khoe-Kwadi families in the Kalahari Basin area (KBA). The classifications of these word types are summarized in Table 1. Word forms that share the same etymological origin are classified into one of the attested Types A-G, that is, ‡qʰài (B1) and ǂʰìì (B2) in Table 1 are cognates integrated into Type B, for instance.

The distribution of each word type is limited within one of the three language families; that is, word forms in Types A and B are observed only within the Tuu family, Types C and D in the Kx'a family, and Types E, F, and G in the Khoe-Kwadi family.

Table 1: Geographical variation of 'dog'.

	Tuu	Kx'a	Khoe-Kwadi
A	A: ǂʰun		
B	B1: ‡qʰài B2: ǂʰìì		
C		C: cām̀mà	
D		D1: gǂʰóé D2: gǂʰúí̋	
E			E1: ʔaba E2: ʔābá
F			F1: aˤagu F2: haˤ-gu
G			G: ʔä̋ri

(KIMURA Kimihiko, NAKAGAWA Hirosi)

Figure 4.21.1: Geographical variations of 'dog' in KBA.

Chapter V

Wolf

'Wolf' in Asian and African languages

Wolves (*Canis lupus*) are found across Eurasia and North America. They have more than thirty subspecies. Some languages have different forms for different members of a subtribe, such as 'wolf', 'jackal' and 'coyote'; however, we do not focus on this discrepancy in the current project.

No data in Sinitic, Hmong-Mien, Austroasiatic, Austronesian, Dravidian, Nilo-Saharan, Niger-Congo or the Kalahari Basin area languages were gathered for this word. Because the item for 'wolf' is supplementary for this project, the data were not collected. However, most of those languages are spoken outside the native areas of wolves. Because the wolf is considered the ancestor of the domestic dog (*Canis lupus familiaris*), languages that do not have an independent word form for 'wolf' may use the counterpart for 'dog' in its place. See also Chapter IV DOG.

Table 1: Main word forms for 'wolf'.

Languages	Word forms
Chukotko-Kamtchatkan	*ʕiɣəlŋən* *χajne*
Ainu	*hórkew* *ónrupus kamuy*
Japonic	*ookami* 狼 *yamainu* 山犬 ('mountain' + 'dog')
Korean	*iri*
Sinitic	*lang* 狼 (added personally)
Tibeto-Burman (TB)	PTB *s-*pjang* (WrT *spyang*) PTB *s-k-ywal* PTB *na* PTB *kla*
Kra-Dai	*ma naj* ('dog'+'wild') *la:ŋ²* (< Sinitic)
Tungusic	*irɡiči* ŇON-
Uralic	*susi* *vergiz*

	sarmik
Mongolic	*čono*
Turkic	*böri* *kurt* *čanavar*
Indo-Aryan (II)	*vŕka* *śŕgālá* *káḍāra* *bhēḍriya* *grastṛ* *huḍahāra*
Burushaski	*urk* (< IA *vŕka*)
Iranian	PIr *uṛka-/varka-* *lewə* *ṣapt* *dâmī* *dib*
Caucasian	t-type k-type b-type g-type
Semitic	*ði:b* *təkwila*

Borrowings beyond the language family are found in languages such as Kra-Dai (from Sinitic), Burushaski (from Indo-Aryan) and Nakho-Daghestanian (from Turkic).

Some compounds are found, such as 'mountain+dog' (Japonic) and 'dog+wild' (Kra-Dai). Some roots used in TB for 'wolf' are identical to those for 'dog'. Hence, 'dog' and 'wolf' are considered essentially identical, and the distinction is usually made by adding a modifier to the form for 'dog'. Biologically, the emergence of wolves was prior to that of dogs, and dogs were more closely related to the lives of the speakers of these languages, so they were given priority.

Language communities that have either detested or deified wolves appeared due to the long history of interactions between wolves and human beings. Pastoral com-

munities, in particular, tend to view them as antagonistic because they attack livestock, although they are respected in some hunter-gatherer societies.

This attitude to wolves is reflected in language in the use of the word for 'wolf' as a taboo word, which is replaced it with other metaphoric words. For example, pastoralists speaking Amdo Tibetan use words denoting 'those whose mouth should be closed' for 'wolf'; conversely, Ainu speakers use words denoting 'spirit-deities of hunting'.

(SUZUKI Hiroyuki)

'Wolf' in Chukotko-Kamchatkan

In Chukchi, the wolf is called *iʔnə*, in Alutor *ʕiɣəlŋən* and in Koryak *ʕeɣəlŋən* (Kurebito et al. 2001). The word for 'wolf' in Alutor and Koryak has a suffix *-lŋən* for the singular that Chukchi lacks.

Itelmen has different word for 'wolf'. The northern dialect (Tigil Village of the Tigil district) has *χajine*, while in the southern dialect (Kovran Village) it is *xiwne* (Kurebito et al. 2001).

(ONO Chikako)

■ A. *iʔnə* type
B. *ʕiɣəlŋən~ʕeɣəlŋən* type
◧ B-1 *ʕiɣəlŋən*
☐   B-2 *ʕeɣəlŋən*
C. *χajine~xiwne* type
○ C-1 *χajine*
● C-2 *xiwne*
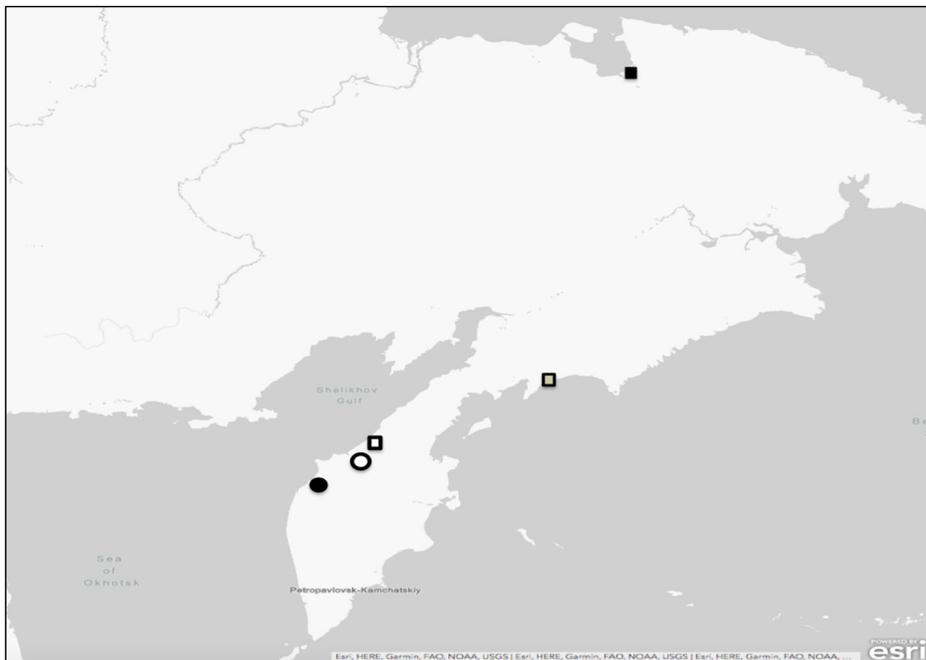
Figure 5.1.1: 'Wolf' in Chukotko-Kamchatkan.

'Wolf' in Ainu

The term for 'wolf' (*Canis lupus*) is divided into two types. Type A, the *hórkew* type, includes three sub-types, which are based on dialectal characteristics: the coda of /-r/ in Hokkaido corresponds to Sakhalin /rA/, and the word-initial /h/ in the eastern Hokkaido dialects have elided in contrast to the western Hokkaido dialects.

The Ainu worship animals, plants, fire, wind, mountains, and rivers, as the *kamuy* or 'spirit-deities.' Type B, *ónrupus kamuy*, means 'the spirit-deities of hunting.' In particular, the Ainu have respect for wolves as *nupuripa kor kamuy* or 'the spirit-deities who govern the upper part of the mountain.' Therefore, the terms for Type A are usually addressed with *kamuy*, such as *hórkew kamuy* (Chiri 1976 [1962]: 141).

(FUKAZAWA Mika)

A. *hórkew* type
- ○ A-1. *hórkew ~ horkew*
- ● A-2. *horokew*
- ◐ A-3. *órkew*

B. *ónrupus kamuy* type
- ╱ *ónrupus kamuy ~ onrupus kamuy*

Figure 5.2.1: 'Wolf' in Ainu.

'Wolf' in Japonic

Wolves have forms such as OOKAMI (*ookame[sama]*, *okami*, *ookabe*), OOINU (*ooinu*, *ooin*, *oinu[sama]*, *oino*, *oin*), and YAMAINU (*yamainu*, *yama no inu*, *yamaeno*, *yameeno*), but most Ryukyuan languages have no forms for wolf.

OOKAMI, OOINU, and YAMAINU are forms that can be analyzed: OOINU is "big dog" or "great dog", and YAMAINU is "mountain dog" or "wild dog"; both are probably newly created words. In contrast to the dogs that are familiar to humans, it is thought that wolves were called "wild dogs" or "big dogs", or even OOKAMI 'great god'. However, for OOKAMI, it may be necessary to consider the possibility that there was a word KAMI (*kame*) that referred to dogs, derived from dog's bark KAA and MEE (cf. *byoobyoo* in Middle Japanese and *bow-wow* in English).

As there are no wolves in the Ryukyu Islands, the word for wolf is basically not found, and *ookami* was borrowed from Japanese. If OOKAMI were a native word for Ryukyuan languages, the forms would have to be *uukami* or *upukam*, but actual form is *ookami*. This clearly shows that this word is borrowed from Japanese.

(NAKAZAWA Kohei and YOKOYAMA Akiko)

	OOKAMI
—	OOINU
⩘	NR
∧	YAMAINU
○	KAMI
⌄	YAMA
	YAMANOANI
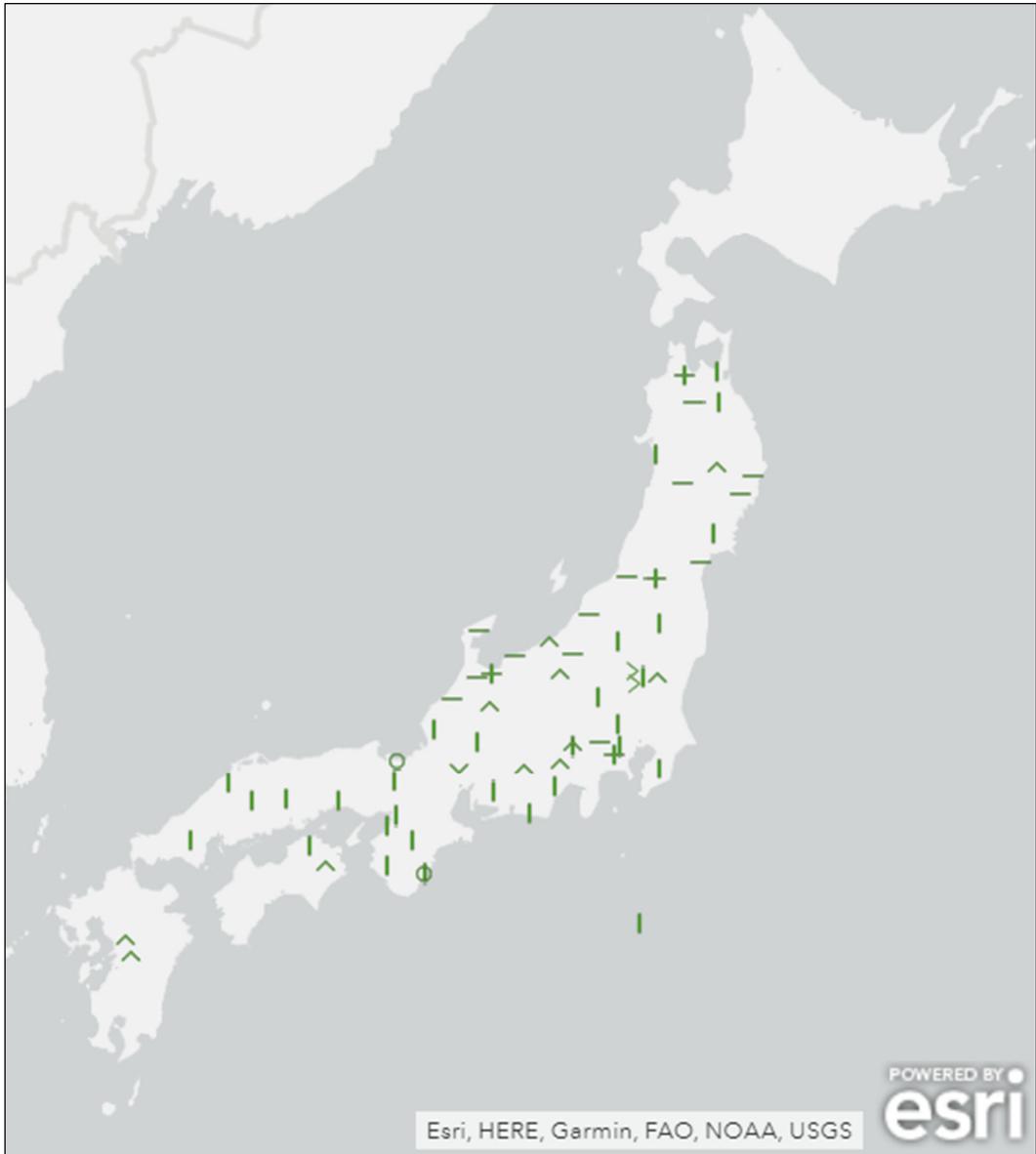
Figure 5.3.1: 'Wolf' in mainland Japan.

Figure 5.3.2: 'Wolf' in Northern Ryukyu Islands.

Figure 5.3.3: 'Wolf' in Southern Ryukyu Islands.

'Wolf' in Korean

We have two etymologically separate words for 'wolf' in modern standard Korean. One is 'iri' and the other 'nɨktɛ'. Only the former form appears as 'irhi' in Middle Korean which is the language spoken from the middle of the 15th century to the end of the 16th century. To my knowledge, the latter form 'nɨktɛ' first appears as late as in the end of the 19th century in written records so that it is a relatively new word. The source of this new word is unknown.

Dialect variation is not so great. There are a few minor phonetic varieties but basically we have two separate words in many dialects.

(FUKUI Rei)

Figure 5.4.1: 'Wolf' in Korean.

'Wolf' in Kra-Dai

Type A has a construction with 'dog' + 'wild', which means 'wild dog'.

Type B is a Sinitic loan word meaning "wolf": B1 and B2 include lang 狼, B3 is chailang 豺狼, and B4 is chaigou 豺狗. This type is distributed along the northern and eastern borders adjacent to the Sinitic speaking area.

Types C to H appear primarily in the Kra branch. The meaning and etymology are unknown.

(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)

- A. 'dog' + 'wild'
- • A1: ma^1 năj^2, ma^{1}lai^2, ma^{1}nai^2, ma^{1}nai^5, ma^{1}nwai^2, ma^{31}na:i^{24}, mā^{6}ma^{1}nai^6
- ○ A2: ma^{1}pe^1, หมาป่า
- ⊠ A3: m̩^{2}nɔ^{23}
- ▣ A4: mpau^{33}zau^{21}
- ◪ A5: ma^{1}ŋa:ŋ^1
- B. Sinitic loan words
- △ B1: la:ŋ^2, laŋ^1, laŋ^4, lɔŋ^4, loŋ^{21}, naŋ^{24}ʑi^{31}, tə^{0}la:ŋ^2
- ▲ B2: qha^{1}laŋ^1
- ▲ B3: tshai^{31}laŋ^{31}
- ◸ B4: tshe^{35}kəu^{42}
- ◿ C. lɔ^{33}ji^{33}kɤ^{33}
- ◣ D. ŋəɯ^{31}ʑi^{13}
- ◪ E. pjuŋ^1
- ■ F. a^{44}tje^{35}
- ▣ G. haŋ^{44}nɗaŋ^{44}
- ▣ H. kɯ:ŋ^{24}

Figure 5.7.1: 'Wolf' in Kra-Dai.

Figure 5.7.2: 'Wolf' in Kra-Dai (enlarged).

'Wolf' in Tibeto-Burman

There are ten major stems (word roots) for 'wolf' in Tibeto-Burman (TB). Most of these forms are etyma of proto-level forms of Proto-Tibeto-Burman (PTB; see STEDT) and Proto-Kuki-Chin (PKC; see VanBik 2009). They include a single stem, a stem plus affix(es), or compound stems. We thus first classify the word forms for 'wolf' into stem type and then into compound type.

The etymology of Type A is *s-pjaŋ (WILD DOG) in the PTB etymon.

The etymology of Type B is *s-k-ywal (WILD DOG/JACKAL/DHOLE/WOLF) in the PTB etymon. This root was borrowed into Indo-Aryan; cf., Sanskrit 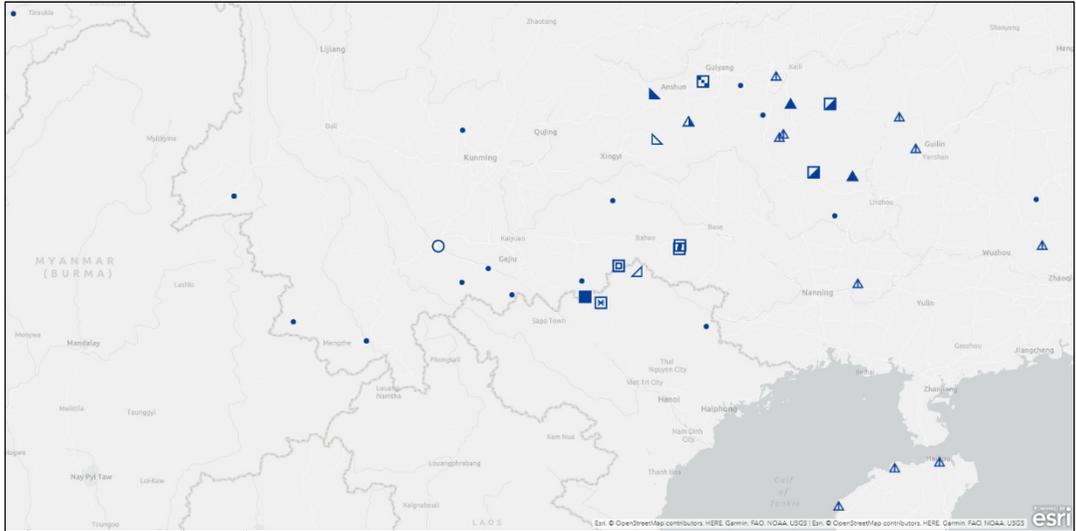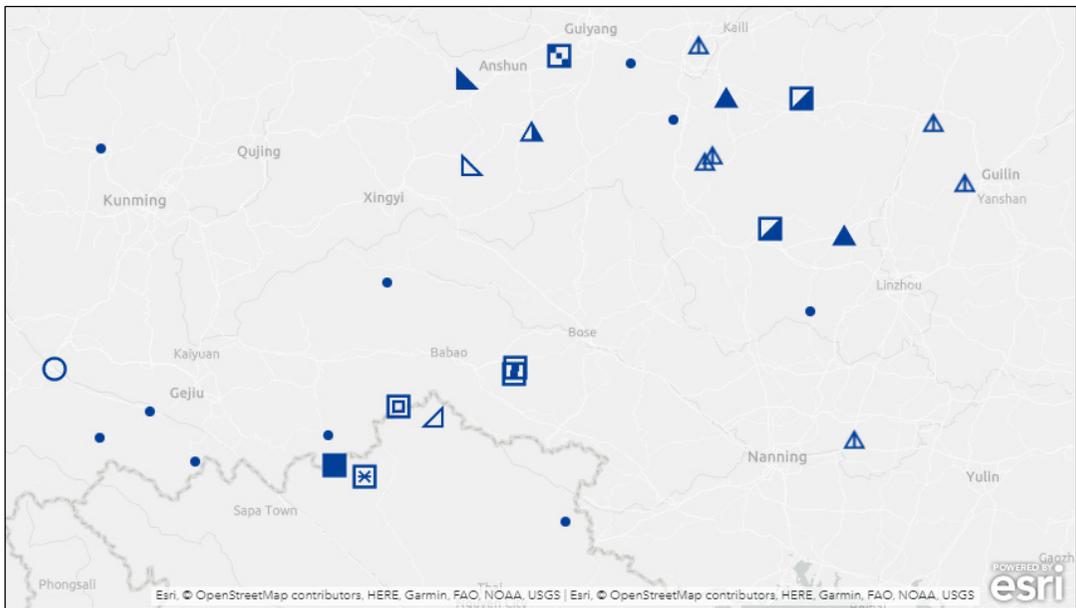śṛgāla. Later on, such an Indic form was borrowed into many TB languages, especially in Nepal. This is an example of a 'backloan' where TB > IA > TB. See Matisoff (2010).

Type C is derived from *na (DOG), and Type D is from *k-la (TIGER) in the PTB etymon. The etymology of Type E hp might be related to that of Type I. Type F tsw is etymologically unknown. Type G is derived from *d-kʷəy-n (DOG) in the PTB etymon. Type H is derived from *tsiŋ-hŋia (FOX/DHOLE/WOLF/WILD CANINE) in the PKC etymon. Type I is derived from *m-par ✕ pra (WILD DOG/WOLF) in the PTB etymon. The etymology of Type J is *d-wam (BEAR) in the PTB etymon. The PTB etyma mentioned above contain several meanings in addition to 'wolf'.

Apart from the ten major types, there are some marginal roots labelled as Type X. Some of them are Tibetan words (kha 'ching, kha dam 'to tie the mouth', and gzig 'snow leopard'), and loans from Chinese (豺狗 chaigou and 豺狼 chailang), whereas the other forms are etymologically unknown.

We also found five types of compound forms. All of them are compounds with Type G: G+B, I+G, X(ram, se)+G, G+X (m-l(e/ə)y), and J+G.

Type A is the most widespread across the branches of TB. This type is found in the northern and central-eastern parts of the TB area (Tibetic, rGyalrongic, and Qiangic), as well as in the south-eastern part (Loloish). Type B is found in the southern (Lolo-Burmese) and central-western part (Sal: Jinghpaw, Northern Naga, and Deng) of the TB area. Type C is found in Loloish; Type D in rGyalrongic, Qiangic, and Loloish; Type E only in Lahu and Kucong; and Type F in Bai, Anong, and Trung. Type G is found as a component of compounds rather than as a single stem. Type H is found in eastern, central, and southern regions; Type I in central and eastern regions; and Type J in Burmish and Karenic only.

The word forms for 'wolf' have much more variety than the ones for 'dog'. This might be related to the fact that 'wolf' is a taboo word and sometimes refers to other words, including slang terms, that do not directly mean 'wolf', because wolves are the most dangerous predators of livestock.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

✓ A. *s-pjaŋ type

tɕã⁵⁵ki⁵⁵, htɕaŋkhu, ɸtɕaŋkhu, spjaŋku, ʃaŋku, ɕiŋ⁵⁵ku⁵⁵, ʰɕãkʰɯ, ʰsã, etc.

❘ B. *s-k-ywal type

kɑl⁵⁵, sho, ʃɔ⁵⁵, son, shan, vi³³, ve¹, viə²¹, etc; vi⁵⁵pa²¹, ve³³tho³⁵ (suffixed); jəkhyon, tʃɑ̃³³khjon³³ (prefixed).

○ C. *na type

na;ŋɕ⁵⁵phɛ³³ (suffixed); mɯ¹ na⁶, ma⁴ nɛ³ (prefixed).

⬚ D. *kla type

lɑ, lou, etc; ləmi, lamɛ, lɔpi, latsi, etc. (suffixed); ãŋõlɔ, hi⁵⁵tɐ³³lɐ³³, paqhala, tɔ³³lɔ³³, etc. (prefixed).

⊕ E. hp type

hɛ phɯ, etc.

⊟ F. tsw type

tsɯ, dzɯ; iɯ dzɯŋ, pipu tsɯ, pi po ɕɯ̃ (prefixed); pi pɯ tɕhiɯ̃ tiɯ (affixed)

— G. *d-kʷəy-n type

▽ G+ type

□ H. *tsiŋ-hɲia type

△ I. *m-par ✂ pra type

⋈ J. d-wam type

∨ X. others

chaigou, chailang, de, kha dam, kha 'ching, gzig, od, u, baŋjEn, ram, m-l(e/ə)y, se, apɵ, etc.
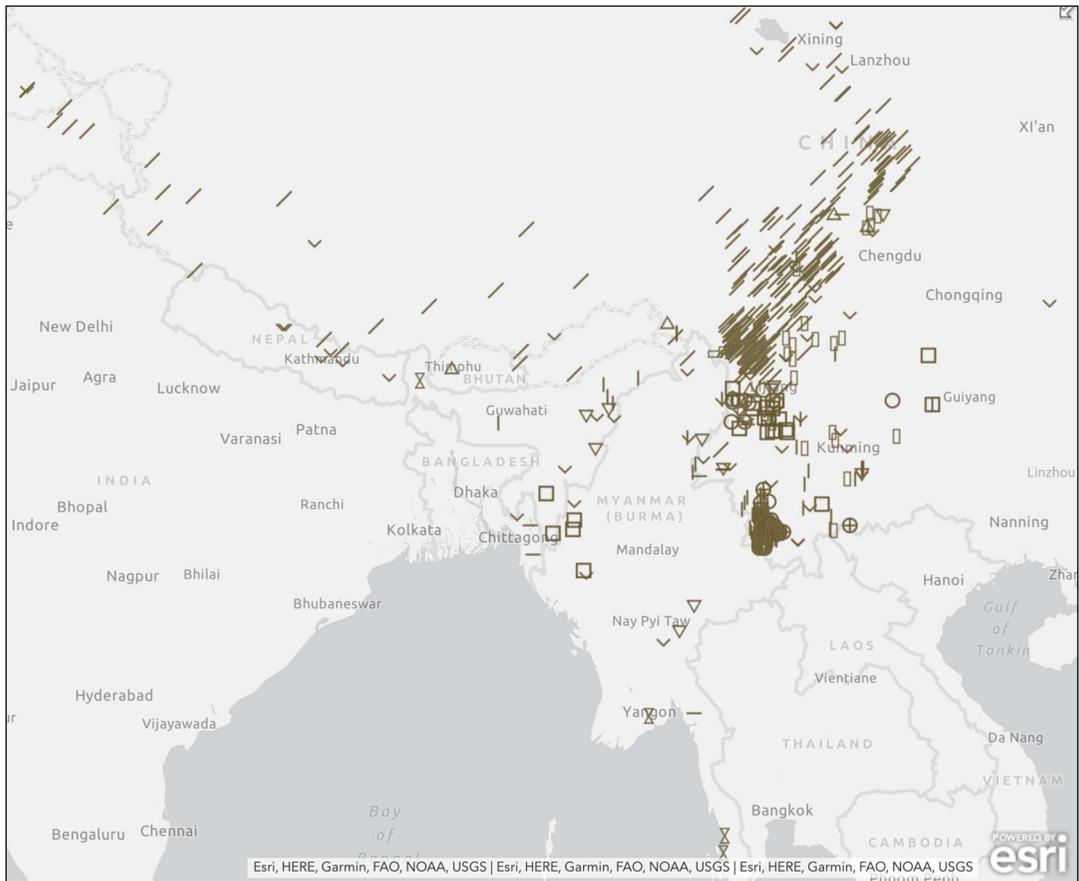
Figure 5.8.1: 'Wolf' in Tibeto-Burman.
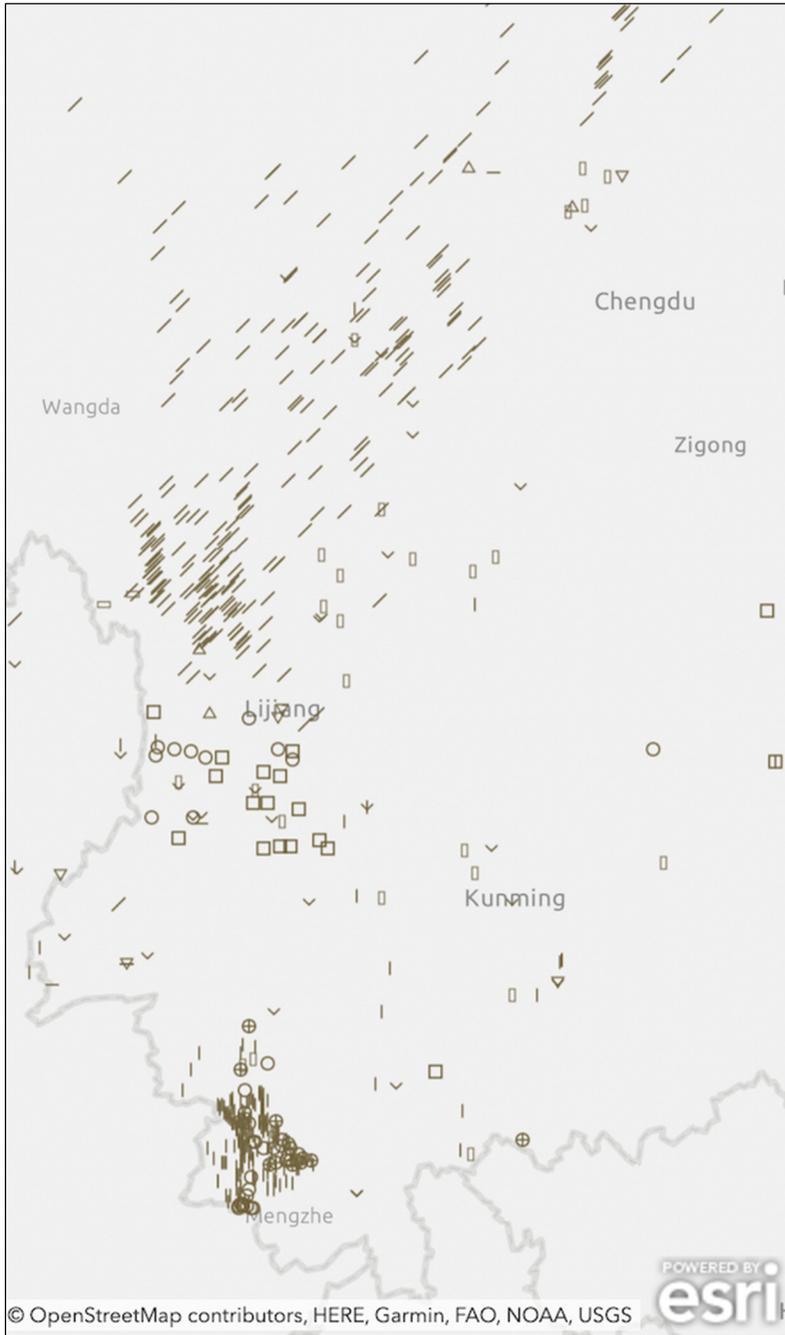
Figure 5.8.2: 'Wolf' in Tibeto-Burman (detailed).

'Wolf' in Tungusic

In Tungusic languages the word forms for 'wolf' would be classified in some types as below:

A Evenki *irgiči*

B ŇON-: Ewen *ňončak,* Udehe *ňeŋu,* Nanay *jeŋgur*

C Oroch *čagu*

D Orochon *gujkə*

E Ewenke *tuuggu*

F Hezhe *ləluki*

G Sibe *yxw*

It is not easy to presume the origin of these words, what they originally mean, where they are from if borrowed. For example, Evenki *irgiči* could be morphologically analyzed into *irgi-* 'tail' and the suffix *-či* 'having something', so *irgiči* means 'animal having a tail'. This is considered that Evenki replaced the word for wolf with *irgiči* as a taboo word preventing wolves from coming to harm their livestock.

(MATSUMOTO Ryo)

○ A *irgiči*

● B ŇON-

◿ C *čagu*

❙ D *gujkə*

♀ E *tuuggu*

☆ F *ləluki*

✳ G *yxw*

Figure 5.11.1: 'Wolf' in Tungusic.

'Wolf' in Uralic

In Uralic there are various forms for 'wolf'. Here I just classified according to the sound forms, it could not be referred to the reason why such many forms they have.

A Finnish *susi*, Estonian *susi*, Votic *susi*, Ingrian *suzi*, Livonian *suž*

B Sami *gumpe*

C Karelian *hukku*

D Veps *händikaz*

E Mordvin *vergiz/vergaz*

F Mari *pire*

G Hungarian *farkas*

H Mansi *legyŋ*

I Khanty *por woj*

J Udmurt *kion*, Komi *köin*

K Tundra Nenets *sarmik*, Enets *sami*

L Forest Nenets *tyh kanunta*

M Selkup *čympyna*

Basically they have each different word forms. For example, even in Balto-Finnic, which languages are relatively closer and have many cognate lexicon, A, C, D types are included. As one of the reasons of this distribution, they avoided calling wolf's name because wolf was regarded as the vermin for their livestock

(MATSUMOTO Ryo)

╲	A *susi, suzi, suž*	N	H *legyŋ*
●	B *gumpe*	♱	I *por woj*
▮	C *hukku*	■	J *kion, köin*
☆	D *händikaz*	☐	K *sarmik, sami*
◎	E *vergiz, vergaz*	⊠	L *tyh kanunta*
✕	F *pire*	◊	M *čympyna*
✵	G *farkas*		
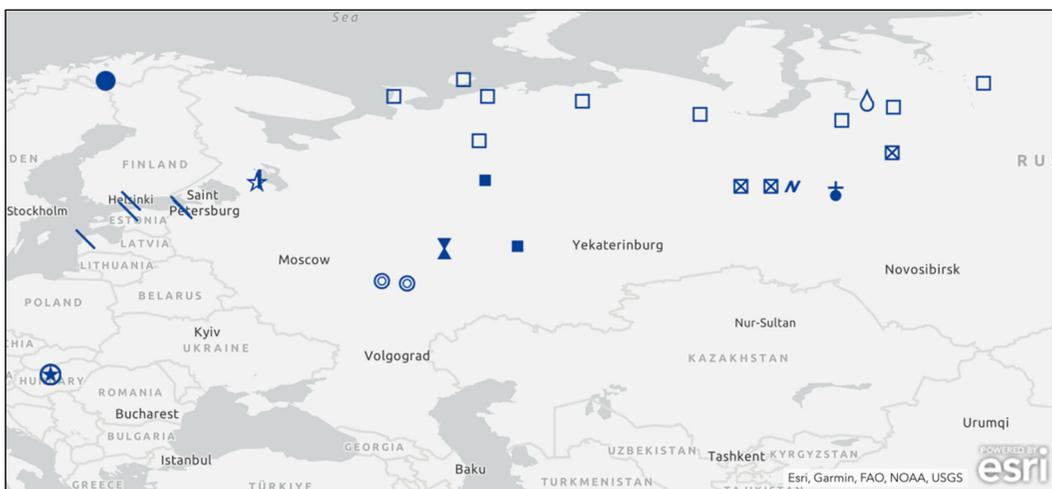
Figure 5.12.1: 'Wolf' in Uralic.

'Wolf' in Mongolic and Turkic

1. Mongolic

In all Mongolic languages except some spoken in peripheral areas, *čono*-type words are used.

Dagur in the northeastern periphery and its branch in Xinjiang use *guskō*-type words (*guskā̄*, *guskō*), which is similar in form to the Tungusic *gujkə*.

We find some other forms in the southern periphery (Gansu province): Monguor has *kadam* and *kadan*, which may be from Tibetic *kha dam* (cf. Ebihara et al. 2022), Dongxiang has a Tibetic loanword, *dʐaŋɢəi*, and Shira Yughur uses the word *ūlīn noxgui*, which literally means 'mountain dog.'

2. Turkic

In Turkic, *böri*-type words are spread over a vast area from northeastern Siberia to Eastern Europe.

In the southwestern region, mainly in Turkey, Iran, and Afghanistan, *kurt*-type words are found.

Azeri has the form *čanavar* in addition to *gurd*. Turkmen has the form *möǰek* as well as *gurt* and *böri*.

Some languages in the central part of the Turkic-language distribution area (Kazakh, Kyrghyz, etc.) use *kaškïr*-type words.

Sarïg Yughur in Gansu province has the form *derdeŋ*. It also has the words *tala əʂt* 'steppe dog' and *tala kük* 'steppe wild animal.'

(SAITÔ Yoshio)

A. *čono* type
- ⌣ čono, čɔn, čʊɑn, šono, šɔn, sʲɔnɔn, čina, čunā

B. ūlīn noxgui
- ⬚ ūlīn noxgui

C. *guskō* type
- ‖ guskō, guskā̄

D. loanword
- ∠ kadam, kadan (< Tibetic)
- ╱ dʐuʐaŋɢəi (< Tibetic)

E. *böri* type
- = böri, bīeri, büre, böre, börö, börü, bioriu, borju, burə, pörü, pȫrü, pǖr

F. *kurt* type
- ▭ kurt, gurd, gurt, kort

G. *kaškïr* type
- | kaškïr, kaskïr, karïškïr, káškǎr

H. čanavar
- ╲ čanavar

I. möǰek
- 0 möǰek

J. derdeŋ
- ▽ derdeŋ
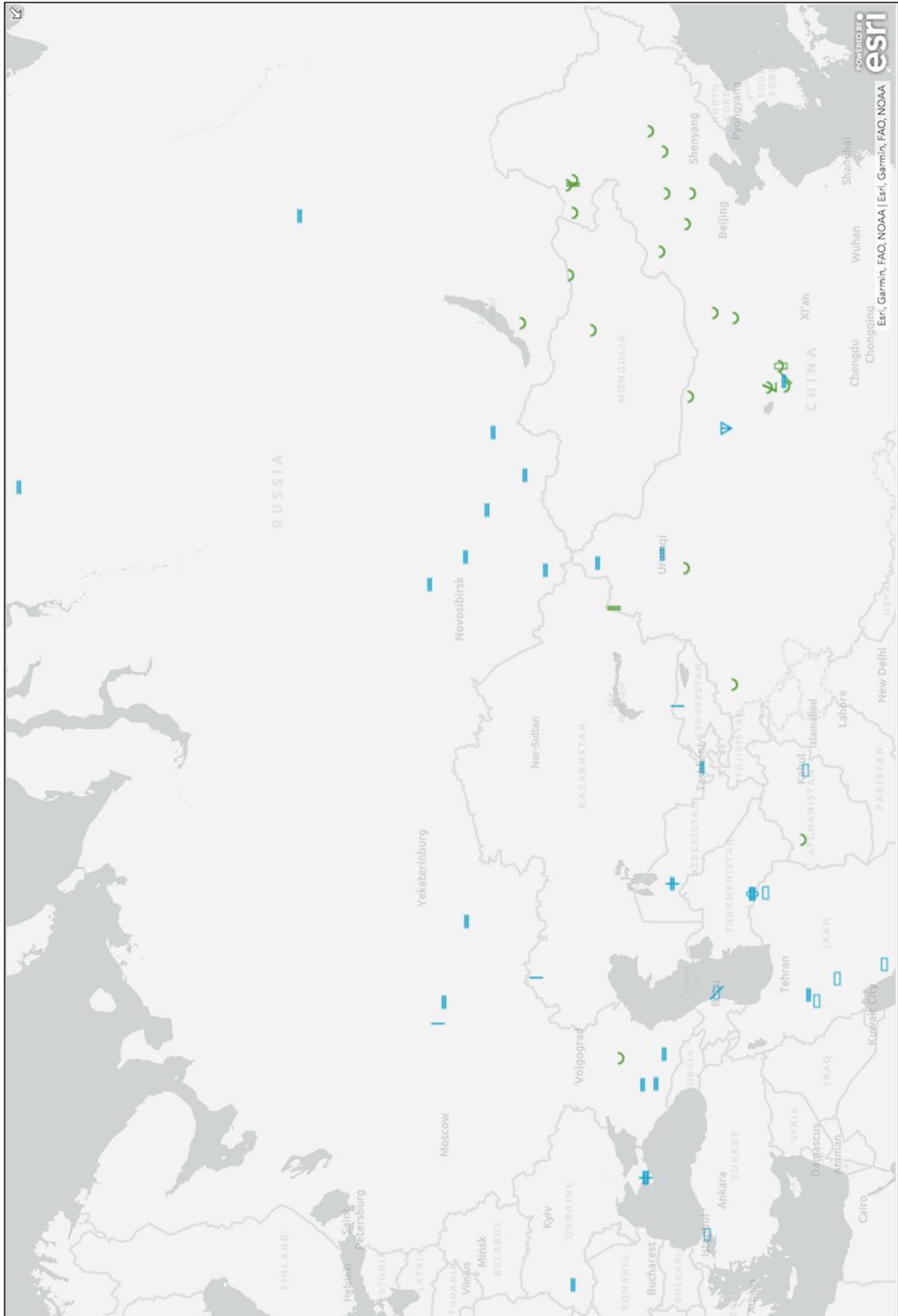
Figure 5.13.1: 'Wolf' in Mongolic and Turkic.
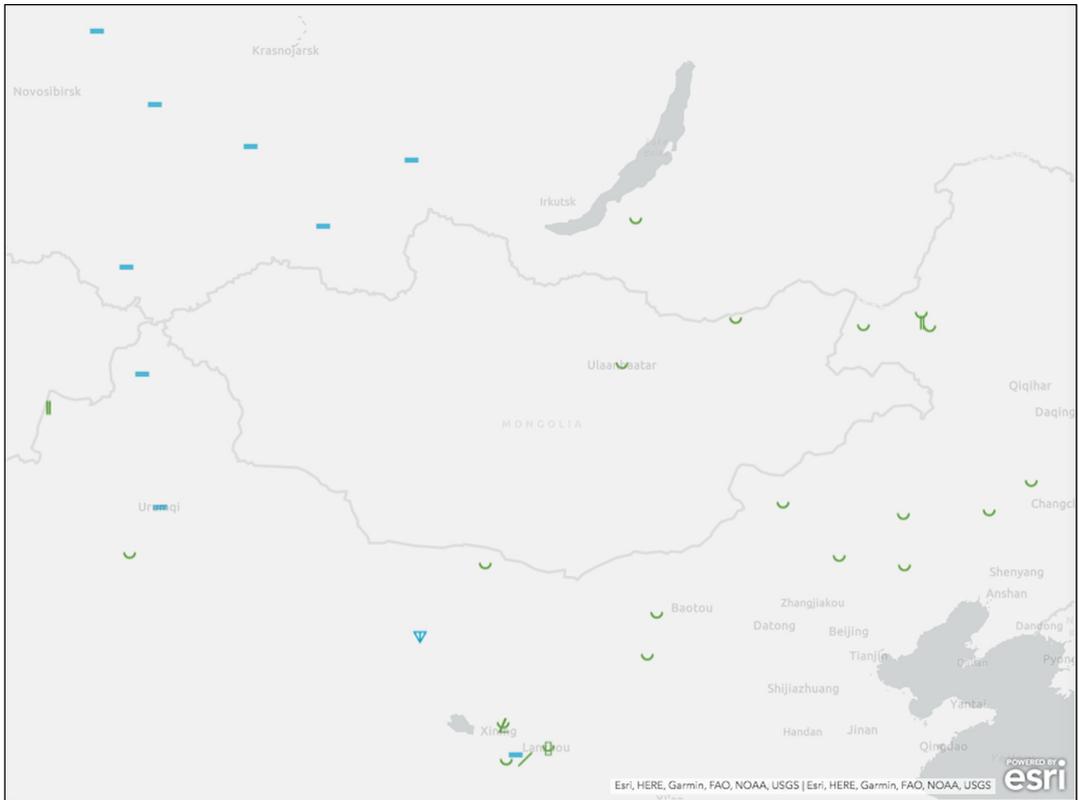
Figure 5.13.2: 'Wolf' in Mongolic and Turkic (The Mongolian Plateau and its vicinity magnified).
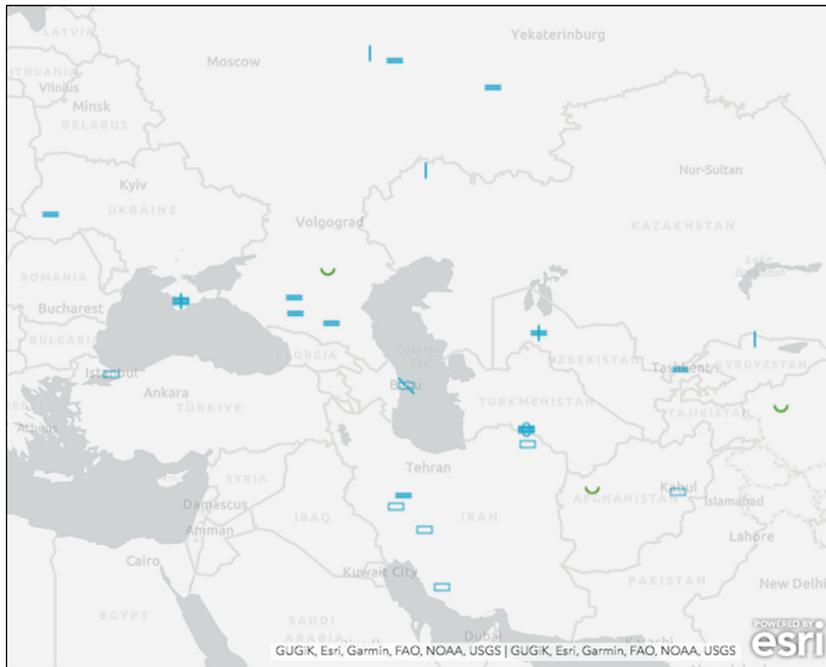
Figure 5.13.3: 'Wolf' in Mongolic and Turkic (Central Asia to East Europe magnified).

'Wolf' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. I did not find many words for 'wolf' in the languages which I treat, as far as I could find. When a language has several words for wolves, I targeted 'male adult wolf'.

The distribution of 'wolf' words is hard to draw the shape. The type A is plotted mainly in the periphery including Europe. In Karakoram-Hindukush, on the one hand, Nuristani languages employ Types B. On the other hand, IA languages employ C and E. Types D and F are for IA the rest region.

The type A is the most major one and is derived from Sanskrit *vŕka* वृक 'wolf'. This can be traced back in a direct line and reconstructed as PIA *wŕkas* and Proto-Iranian *wŕkah*, which are derived from PII *wŕkas*, and untimately PIE *wĺkwos* 'wolf' (originally means 'dangerous'). So this type is cognate with Latin *lupus*, Ancient Greek *lúkos* λύκος, and of course English *wolf*. Forms of this type are used by IA and Burushaski languages. Historically the Burushaski form *urk* was borrowed from Ishkashimi, a Iranian language in Afghanistan and Tajikistan. Thus the route is not same between the IA and Burushaski (and then Domaaki *urk* and Drasi Shina *urúk*) forms. Nepali *bwãso* ब्वाँसो and Sinhala *vŕkayā* වෘකයා are derived from Sanskrit *vŕkadãśa* वृकदँश 'wolf as biter' (cf. *dãśa* दँश 'biting').

The *śŕgālá* type appears in IA and Nuristani, which are concentrated in the area of Himalaya, Karakoram, and Hindukush mountain ranges. The original Sanskrit form *śŕgālá* शृगाल refers to 'jackal', which is inherited from PIA *śŕgālás*. Its Proto-Iranian counterpart is *čárguš* 'lion', derived from PII *ćárguš* 'predatory animal', and the PIE form is reconstructed as *ḱérgús*. Proto-Nuristani may be considered as *şiyól*, from Kam *şiól*, Kati *şyōl*, and Prasun *şil*.

The third major type *káḍāra* is used in Indo-European languages around the border between northwestern Pakistan and eastern Afghanistan. The Sanskrit word *káḍāra* कडार means 'having projecting teeth' (Turner 1966: 132) or 'tawny' (Monier-Williams 1899: 245). Waigali *ḍẽkar* suffered metathesis.

Next, the *bhēḍriya* type can be seen only in IA languages. The forms originate in Sanskrit *bhēḍriya* भेड्रिय 'sheep-killer' (see also the type F), related to *bhēḍra* भेड्र 'sheep'.

The type E of *grastṛ* is found in IA languages at the most northwestern part of Pakistan. The word *grastṛ* ग्रस्तृ in Sanskrit originally meant 'eclipser' and then 'swallower'.

The *huḍahāra* type F is detected in northern India and Nepal. Sanskrit *huḍahāra* हुडहार refers to 'ram-taker', made of *huḍa* हुड 'ram' and *hāra* हार 'take away'.

In some languages, words derived from Proto-Iranian *daywáh* 'devil' (< PIE *daywós* 'heavenly, god'), or *l/nekṛa* 'rag, tattered cloth' (< Sanskrit *lēkka* *लेक्क 'defective') for the concept 'wolf'. In Khowar, they employ *šapír*, which is originated in Sanskrit *śápyati* शप्यति 'curses'. Xaladikta Romani *ryč* is a descendant of Sanskrit *ŕkṣa* ऋक्ष 'bear', see the paper on 'Bear' in this volume. Kashmiri *rāmĭ hūn* رامؠ ہون means 'dog of the god Rāma'.

(YOSHIOKA Noboru)

A. *vŕka* type (12) ☆

    urk, urúk, varu, ruv, ruyi

 [+*dãśa*]

    vṛkayā, bwãso

B. *śṛgālá* type (6) ⟨

    šā̃l, šiōl, ṣiól, ṣyōl, ṣil, šyāŋku

C. *káḍāra* type (6) ⟍

    kaṛál, kaṛák, karā́ṛ, karā́ro, ḍēkar

D. *bhḗḍriya* type (5) ◎

    bheṛiyā, bheḍīyo, bheḍyā, pə̀gyaṛ

E. *grastṛ* type (4) ◆

    grast, grāsta, grhas, grac

F. *huḍahāra* type (4) /

    huṛār, hū̃ṛār, hunā̃r

G. others

    [*lēkka* type (2)] lakṛa, nekṛe; dēu, ryč, baghiāṛ, lāṇḍgā, rāmi̇ hūn, šapír
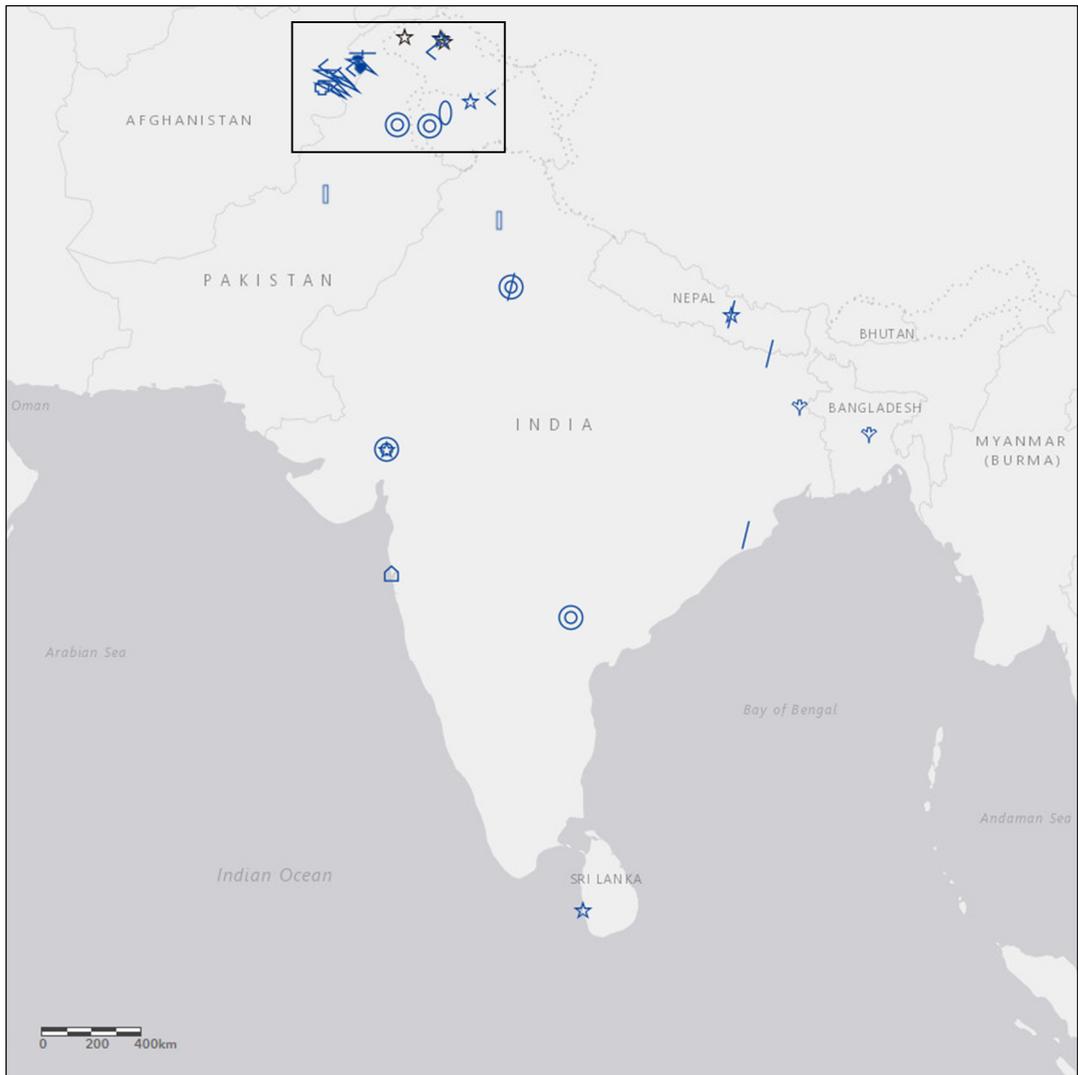
Figure 5.14.1: 'Wolf' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).

Figure 5.14.2: 'Wolf' in northern Pakistan (the area encloed by the rectangle in Figure 5.14.1).

Figure 5.14.3: Types for 'Wolf' in Indo-Aryan languages outside South Asia.

'Wolf' in Iranian

Type A is the most widespread type, which traces back to the PIr word *u̯r̥ka-/varka-* (< PIE *u̯l̥kʷos-*). Type B is confined only to Pashto. It originated from OIr. *daivya-* 'daevic (i. e. devilish) animal', which replaced an original word for 'wolf', probably because the word became a taboo word. Type C, observed in Wakhi and Sarykoli, may also be a euphemism, a loanword from Dardic (cf. Khowar *šapīr* 'wolf' < Skt. *śapita-* 'cursed'). Ormuri has Type D *dâmī*, which literally means 'hunter' from Persian (ultimately from Arabic). Type E is a loanword from Arabic equivalent ذنب (ð'b) 'wolf'.

Type B through D are similar to each other in that they replaced inherited words and had bad meanings in order to avoid taboo words.

Ossetic word *bīræg* 'wolf', which I classified it Type A here, is controversial. A form inherited from Old Ossetic is fossilized in person names *Wærxæg*. Some scholars assume that it is a loanword from Turkic (cf. Uzbek *büri* 'wolf'), whereas Abaev (1958) argues that it is likely to come from an Eastern Iranian word (ex. Khotanese *birgga- [birɣa]* 'wolf'.) in order to explain the last sound [g].

(IWASAKI Takamasa)

- — A: *vruk* type
- | B: *lewə* type
- ▽ C: *ṣapt* type
- ⚘ D: *dâmī* type
- ◗ E: *dib* type
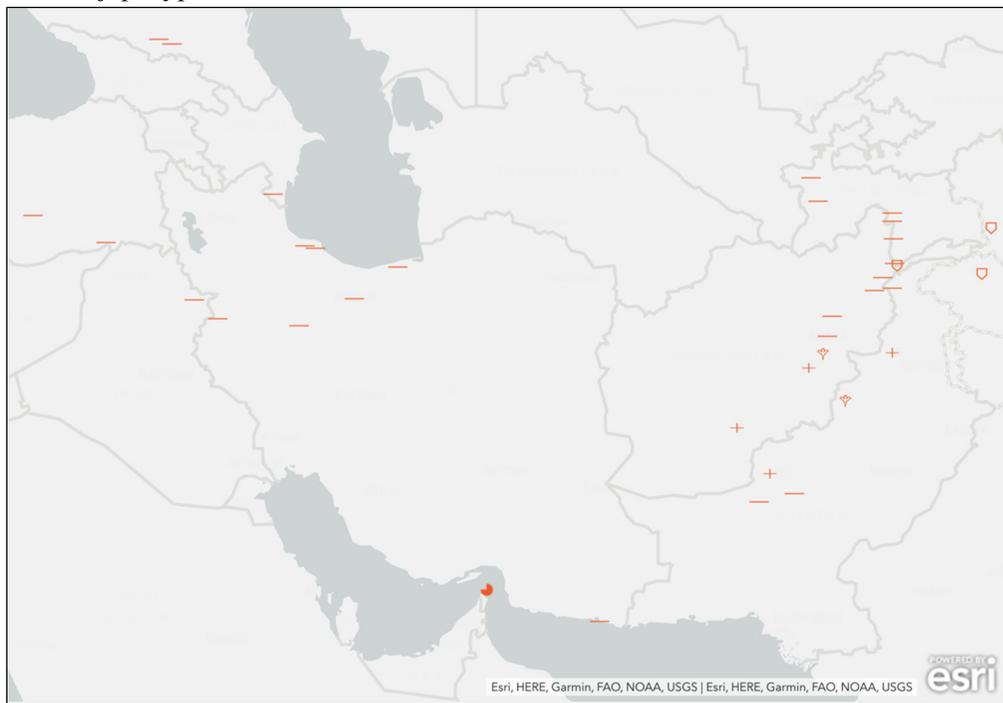
Figure 5.16.1: 'Wolf' in Iranian.

'Wolf' in Caucasian languages

Type A appears in Svan, Adyghe, and Kabardian. Type B is found in the other Abkhazo-Adyghean languages. Type C is widely distributed in Nakho-Dagestanian languages. Type F appears in the Dagestanian languages spoken in the southernmost area.

Type D is a loan from Armenian *gajl*. It appears in Kartvelian languages, but they are not always in close contact with Armenian. It is assumed that the borrowing process occurred in the Proto-Kartvelian period. Type E is borrowed from Azerbaidjani *džanavar*, which originally came from Persian. Type E is found in the periphery of the Azerbaidjani-speaking area rather than in languages in close contact with Azerbaidjani; the borrowing possibly occurred in an early time.

(SUZUKI Hiroyuki)

A: t-type; *txere*, *tiʁʷʒʲ*, etc.
B: k-type; *a-kwɨdžʲma*, *kʲabɨ*, etc.
C: b-type
  C1: *boc'o*, *bac'a*, *boc'e*, *bec'*, etc.
  C2: *borz*, *bʕorc'*, *barc'*.

D: g-type; *mgel-i*, *ger-i*, etc.
E: dž-type; *džanavar*, *žanavar*, etc.
F: ub-type; *ubul*, *eb*, *ul*.
G: others; *jam*, *gra*.
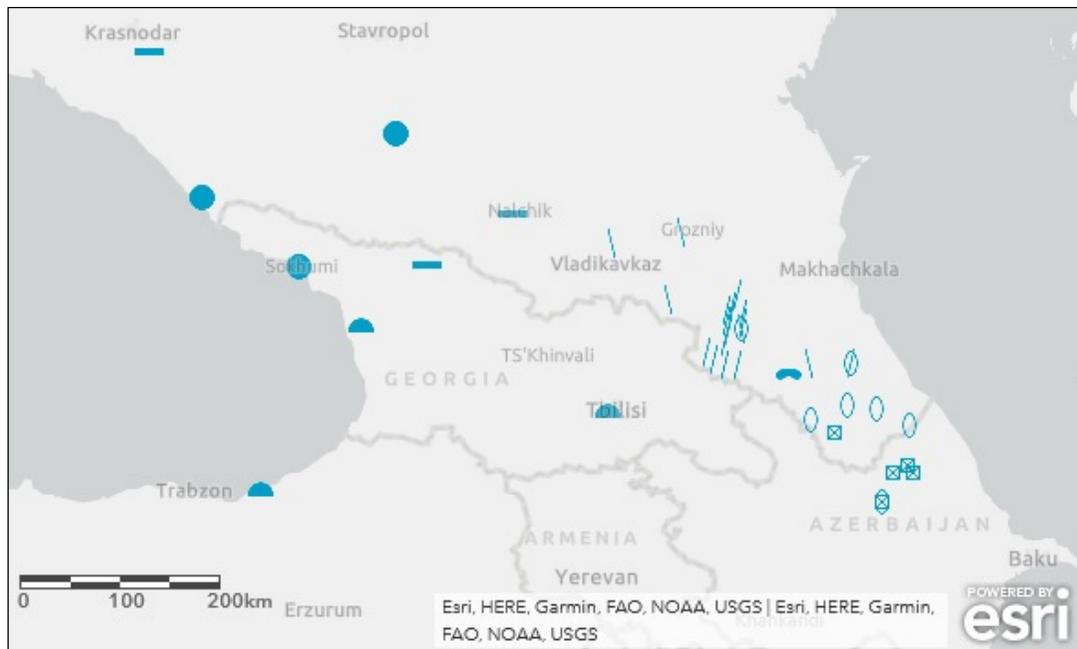
Figure 5.17.1: 'Wolf' in Caucasian languages.

'Wolf' in Semitic

Semitic except Ethiopic have cognates of *ði:b*.

A. *ði:b* type (•) with a fricative *ð* is the form of nomadic Arabic (Najdi, Gulf, Iraqi, Yemen, Libya).

di:b type (⬤) with a plosive *d* is the form of sedentary Arabic dialects (Syria, Lebanon, Egypt) and *díb* (Malta, Morocco).

zi:b (✖) is the Bukhari Ar. form derived from Iraqi *ði:b*. *ze'ev* (✚) in Hebrew.

diwa type (▲) with *w~v* (<*b*) is Aramaic form: *diwa* (Jilu), *dewa* (Hertevin), *divo* (Mlaḥsô).

B. *təkwila* (▽ Amharic ተኵላ), *təkla* (▽ Tigre) are Ethiopic.

kawb, *koób* (✻) is used in South Arabia: *kawb* (Mehri) and *koób* (Hobyot).

(NAGATO Youichi)

A type
- • *ði:b*
- ○ *di:b*
- ✖ *zi:b*
- ✚ *ze'ev*
- ▲ *diwa, dewa, divo*

B type
- ▽ *təkwila, təkla*
- ✻ *kawb, koób*

Figure 5.18.1: 'Wolf' in Semitic.

Chapter VI

Bear

'Bear' in Asian and African languages

There are several species of bears in Asia and Africa: 'brown bear,' 'polar bear,' 'Asiatic black bear,' 'sun bear,' and 'sloth bear.' Some languages distinguish the terms for 'bear' according to these species. For example, Chukotko-Kamchatkan languages distinguish between 'polar bear' and 'brown bear.' Indo-Aryan (IA) and Tibeto-Burman (TB) languages use the terms 'brown bear' and 'Asiatic black bear' differently. Conversely, in areas, such as the northern region of Hmong-Mien, there are no data for areas where bears do not inhabit, and in areas, such as the small Austronesian islands, the vocabulary gap is filled with loan words.

Additionally, the Ainu language uses different bear names for different ages and sexes, which are broadly classified as bear cubs, female bears, and male bears. The terms 'bear cub' can also be referred to by age, such as *hepér* (1-year-old cub), *riyáp* (2-year-old cub) 'lit. winterized thing,' and *cisúrap* (3-year-old cub) 'lit. separated thing (from their parents).' This change in bear terms according to age and sex may be closely related to the custom of the bear-sending ceremony in the Ainu culture.

Uralic has rich words for bears, which are difficult to find cognate as each other. This may be due to the taboo against direct references to bears since they are ferocious animals, are feared by hunters, and are ritualized as something to be honored (See the detailed discussions in Hallowell 1926, Emeneau 1948, and Petrov 1989).

The Ainu respect bears as "mountain spirit-deities," and a general term for 'bear' is *kamuy* 'spirit-deities.' The languages of Tungusic, Turkish, and TB use the term 'grandfather' or 'old man' for 'bear.' Other animals, such as 'dog' in Sinitic, 'monkey' in Sanskrit, and 'rat' in Dravidian, can also be related to 'bear.' This is a way to eliminate fear by replacing a bear term with a kinship term and a more familiar animal term.

Other euphemisms, such as the words originating from 'brown,' 'black-skinned,' or 'destroying' indicated by a real/imaginal property, can be found in Proto-Indo-European (PIE) and Sanskrit languages, from which IA and Dravidian languages were derived.

In many cases, as in Japonic, a word for 'bear' originates from a loan word. Because the bear is thought to understand human speech, humans often use a special style or lexicon when speaking of or to a bear (Janhunen 2003). This may be another reason for the more active use of loan words.

Table 1: Main word forms for 'bear'.

Languages	Word forms
Chukotko-Kamchatkan	'Brown bear'
	kejŋen ~ kajŋən type
	weqanł
	met'sk'aj
	'Polar bear'
	umqə ~ umqa type
Ainu	'Male bear'
	síyuk type 'big game animals'
	'Female bear'
	kucán type
	'Bear cub'
	péwrep type
	hepér type
	'Bear (synecdoche)'

	kamúy type < 'spirit-deities' *iso* type < 'game animals'	Turkic (Tk)	*adïg ~ ajï type* (< OTk *aðïg*) *aba* type (< 'old man, grandfather') *ire* type *(*< possibly Samoyedic 'grandfather and old man'*)*
Japonic (J)	*kuma* type (< OJ < pre-Middle Korean) *sisi* type < 'beast and meat'		
Korean (K)	*ko:m* (< MK)	South Asia	*ŕkṣa* type (< Sanskrit and Proto-Nuristani < PIE 'destroying') *bhalluka* type (< Sanskrit 'auspicious and favorable') *babhru ~ bhrāru* type (< Sanskrit 'reddish brown' or 'brown animal' < PIE 'brown, bright') *nya* type *śvāpáda* type (< Sanskrit 'beast of prey, wild beast') *kaḷōttī* 'black bear' (< Sanskrit 'black-skinned') *ruvdič* (cf. *ruv* 'wolf') *metčkoi* (< possibly Church Slavonic: *mečĭka* 'lit. she-honey-eater')
Sinitic (Sn); Chinese (C)	熊(-) type: *ɕyŋ*(-) (< MC < OC) 狗 type: *kou-* < 'dog (-like)' or a kind of prefix 黑/瞎 type: *xei-/ɕia-*		
Hmong-Mien	*kləp* type *ɕoŋ* type *zuŋ* type		
Kra-Dai (KD)	*mi* type (< **hmi A*) *jong* type (< Sn *xiong* 熊)		
Tibeto-Burman (TB)	Proto-Loloish etc.: **(k-)d-wam* type WrT **dred* type *yiel* type (< **(k-)d-wam* type)		
Austroasiatic	Proto-Mon-Khmer (PMK) etc.: **ɟkaw* type PMK etc.: **[k]r[e]s* type PMK etc.: **[k]mum* type Proto-Pramic: *bi:ʔ* type Proto-Pray-Pram: **bɛ:k* type Proto-Khmuic: **suəl* type Munda: **bana* *rəwaj su:t* *kawap*	Dravidian	*elugu, erj, ili, oḍi, eju* (cf. 'rat') *karaṭi, ka·ṟ* (cf. *kar* 'black') *guṟṟi* 'black'
		Iranian (Ir)	*xers* type (< PIr < PIE) *melu* type (< possibly Hindi) *mamm* type (< probably Dravidian) *noghondom* type
Austronesian	*tsmoi* type *ōso* *bauaŋ* type *ber* type (< English)		
Tungusic	*amaka* type (cf. Evenki 'grandfather') *nakat type* *mapa type*	Caucasian	*d-* type (< **datv-*) *mš-* type *ča* *s/š/z-* type
Uralic	*karhu, karu, kondi, ofta, ovto, okš, wark, app'ji*		
Mongolic	*bābgai type* *xar görōs* type < 'black wild beast' (< C 黑熊 'lit. black bear') *noxoē xar görōs* type 'dog black wild beast' (< C 狗熊 'lit. dog(-like) bear')	Semitic	*dubb* type *debba* type *dəbi* type *dov* (< **dob*)

The symbol # marks a (heuristically/ author's) reconstruction.

(FUKAZAWA Mika)

'Bear' in Chukotko-Kamchatkan

The brown bear (Ursus arctos) is distributed in the Chukotkan- and Kamchatkan peninsulas.

In Chukchi, the brown bear is called *kejŋən*, in Koryak *kajŋən* (Kurebito et al. 2001); here the vowel /e/-/a/ change has occurred. Alutor calls it *keŋən*, so the Alutor name lacks /j/ in the middle of the word.

Itelmen has a different name for the brown bear. The northern dialect (village Tigil of the Tigil district) calls it *weqanɬ*, while in the southern dialect (village Kovran) it is called *met'sk'aj* (Kurebito et al. 2001).

Polar bear (Ursus maritimus) in Chukchi is called *umqə*, which Koryak and Alutor call *umqa* (Kurebito et al. 2001). Here we see the last vowel /ə/ changes into /a/ in the south.

Itelmens have no name for the polar bear, because it is not distributed in Kamchatka and people have had almost no experience of seeing it before. If they see this animal, they would call it "white/light" (*atxlaχ*) bear (*weqanɬ* or *met'sk'aj*).

(ONO Chikako)
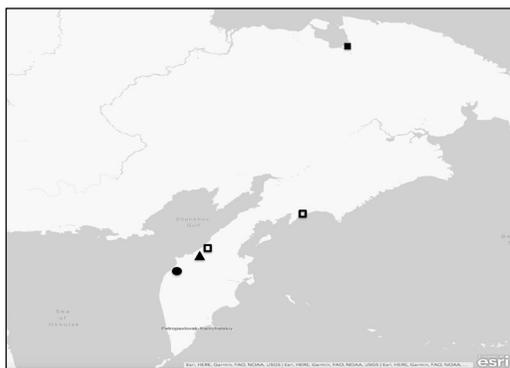
Figure 6.1.1: 'Brown bear' in Chukotko-Kamchatkan.
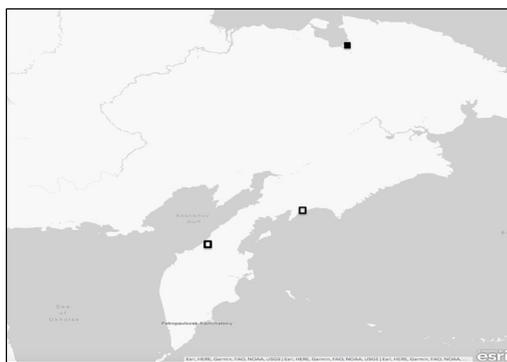
Figure 6.1.2: 'Polar bear' in Chukotko-Kamchatkan.

A. *kejŋen~kajŋən* type
- ■ A-1 *kejŋən*
- □ A-2 *kajŋən*

- ▲ B. *weqanɬ* type
- ● C. *met'sk'aj* type

A. *umqə~umqa* type
- ■ A-1 *umqə*
- □ A-2 *umqa*

'Bear' in Ainu

The 83 vocabulary items for 'bear' are listed in Chiri's (1976 [1962]) *Classified Ainu dictionary: Animals*. The Ainu worship animals, plants, fire, wind, mountains, and rivers as their *kamuy* or "spirit-deities." Among them, bears are an important *kamuy* in the Ainu culture. They are worshipped as the spirit-deities of mountains, known as the *kimún kamuy*, *nupúrikor kamuy,* and *metótus kamuy*. *Kamuy* or 'spirit-deities' can express the synecdoche meanings, as follows: 'bear' (*Ursus arctos yesoensis*) in Hokkaido and Kuril, 'seal' in the eastern Sakhalin, and 'Steller sea lion' in the western Sakhalin. In Sakhalin, bears (*Ursus arctos collaris*) are called *iso*, which means 'game animals' in Hokkaido.

Here, word forms are divided into three categories according to age and sex, as shown in Figures 6.2.1 to 6.2.3. Figure 6.2.4 shows the distribution of the words *kamúy* and *iso* for 'bears' that are used as a synecdoche.

(FUKAZAWA Mika)

A. *síyuk* type (lit. 'big game animals')

  ○ A-1. *síyuk ~ siyuk ~ síuk*

B. male X type

  ▭ B-1. *piine iso* (lit. 'male bear')

  | B-2. *pínne kamuy* (lit. 'male bear')

  ‖ B-3. *piineh* (lit. 'male one')

C. others

  ∨ *iso* (lit. 'bear'< 'game animals')

  • *áska kucán* (lit. 'clean female bear')

A. *kucán* type

  ● A-1. *kucán*

  ⊖ A-2. *kucán yuk* (lit. '(female) bear game animals')

B. female type

  | *mátne kamúy* (lit. 'female bear')

Figure 6.2.1: 'Bear (male)' in Ainu.

Figure 6.2.2: 'Bear (female)' in Ainu.

A. *péwrep* type (lit. 'young one')   A & B type

    ○   *péwrep ~ pewreh*   ◉   *péwrep* and *hepér*

B. *hepér* type   C. small X

    ●   *hepér ~ epér ~eper*   ✚   *pon iso* (lit. 'small bear')

Figure 6.2.3: 'Bear (cub)' in Ainu.

A. *kamúy* type ('spirit-deities')   B. *iso* type ('game animals')

    ▮   *kamúy ~ kamuy ~ kamui*   ▭   *iso*
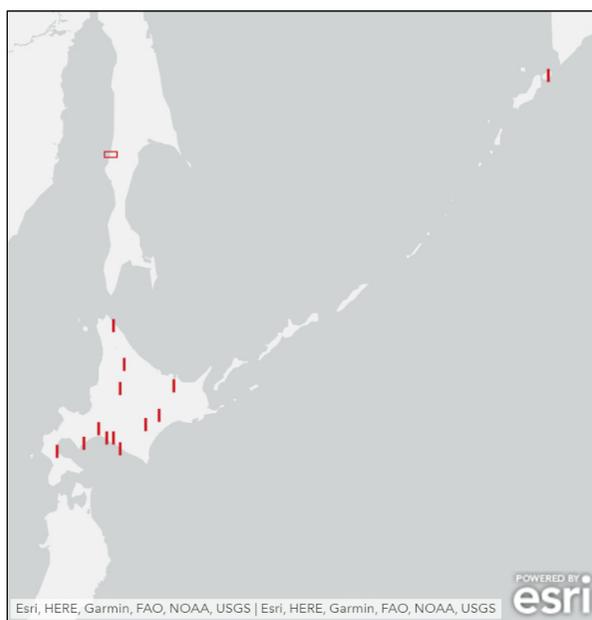
Figure 6.2.4: 'Bear' (synecdoche) in Ainu.

'Bear' in Japonic

The major form for the bear is KUMA (*kuma*, *kumame*). SISI (*sisi*, *notarizisi*, *suzisi*) and KUMANOSISI are found in the Tohoku region. Most Ryukyuan languages have no corresponding forms for the bear.

As there are no bears on the Ryukyu Islands, the word for bear is basically not found, and it is thought that *kuma* was borrowed from Japanese. Brown bears are distributed in Hokkaido, and black bears south of Honshu, thus the types of bears differ depending on the region; therefore, KUMA originally referred to black bear, which is attested in Old Japanese *kuma*. Brown bears are called *siguma* in Old Japanese, later *higuma*, which contains the morpheme *kuma*. The *si* of *siguma* is probably the same morpheme as that of *sika* 'deer' (cf. *ka* 'deer' and *meka* 'doe'), which means "male" or "big". *sisi* is a word that means "beast" or "meat" and is thought to refer to wild animals that are edible. The area where bears are called SISI or KUMANOSISI is probably named because bears are typical wild-food animals, like *inosisi* 'wild boar' and *kanosisi* 'deer.' The HONGUMA type is used to distinguish the bear from other animals called *kuma*, such as *anaguma* 'badger'.

There are no other forms of bear in Japonic languages other than KUMA. Loanwords usually have few variants. Therefore, KUMA must be a loanword and is probably derived from *kuma ~ koma* on the Korean Peninsula found in Gongju's ancient name "고마ᄂᆞᄅ" *komanʌnʌ* (Ungjin, literally "bear port").

(NAKAZAWA Kohei and YOKOYAMA Akiko)

	KUMA
	SISI
^	KUMANOSISI
ᴗ	HONGUMA
ᴎ	NR
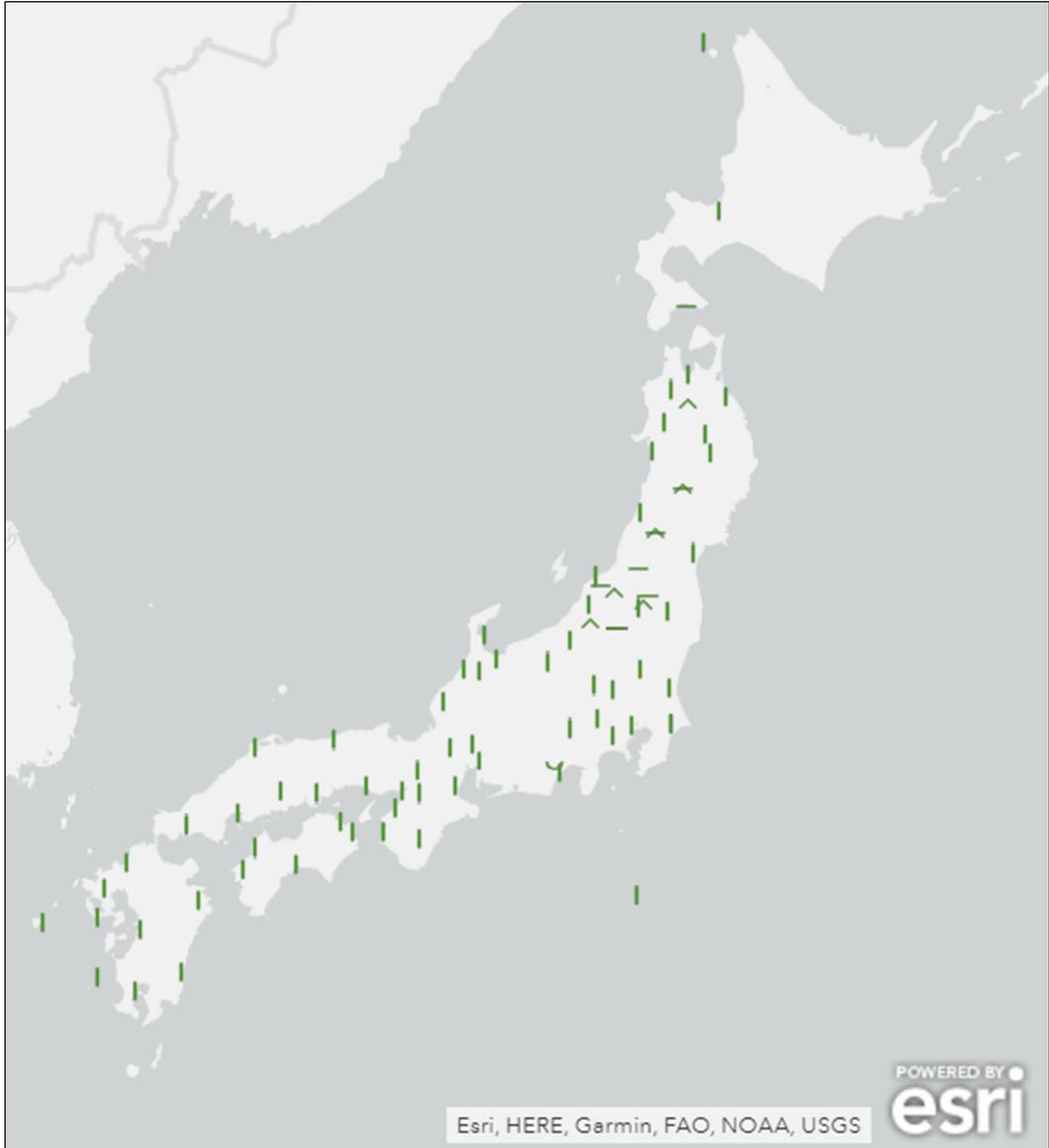
Figure 6.3.1: 'Bear' in mainland Japan.

Figure 6.3.2: 'Bear' in Northern Ryukyu Islands.
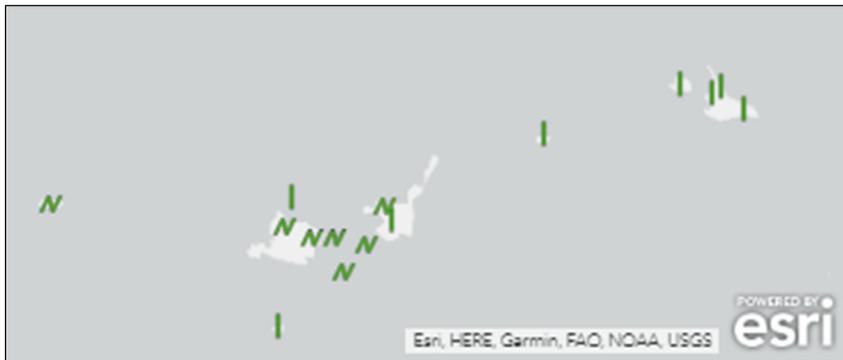
Figure 6.3.3: 'Bear' in Southern Ryukyu Islands.

'Bear' in Korean

Modern standard form for 'bear' is 'ko:m' and Middle Korean form 'kom (R)'. Middle Korean is the language spoken from the middle of the 15th century to the end of the 16th century. These two forms are almost the same except that the Middle Korean form is marked with a rising tone (abbreviated as R above), which is usually interpreted as having rising pitch contour and vowel length.

As for the modern form, older speakers of Seoul pronounce this word with a long vowel which is a reflex of the Middle Korean rising tone. However, this long vowel has been lost among younger speakers.

Historically, it is usually the case that words with the Middle Korean rising tone go back to a disyllabic word. In this case it was recorded as 'koma' in the place name '熊津' in a 15th century document. Also, this disyllabic form has often been compared with the corresponding Japanese word 'kuma'.

Dialect variation is not so great. In Ogura (1944) we have forms like 'nɔ:-pʰje', 'nɔ-pʰe', 'nɔ-pʰɛŋ-i' and 'kom-pʰje, other than the form mentioned above. The form 'nɔ-pʰe' is believed to be a borrowing from Manchu 'lefu' (Ogura (1944: 2nd vol., 580-581). However, all these forms are special terms used by wild Ginseng hunters often referred to as 'Simmani. Therefore these forms are not included in the Map.

(FUKUI Rei)
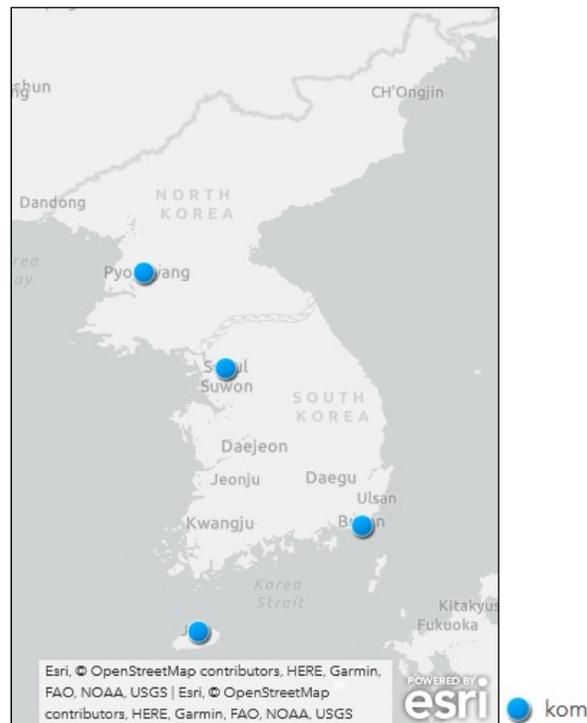
Figure 6.4.1: 'Bear' in Korean.

'Bear' in Sinitic

We classify forms based on the first syllable, then subclassify by next morpheme types.

A1: 熊ɕyŋ(集宁)　　am21 / hoŋ21 (广宁)
A2: 熊瞎子ɕyŋ ɕia tsɻ(承德)
A3: 熊人hoŋ iɐn (广州)
A4: 熊家iuŋ ka (梅县)
A5: 熊嘎婆ɕioŋ ka pʰo (重庆)
A6: 熊婆iuŋ pʰɔ (北流塘岸)
A7: ʃan mɛ (贵港南汇)
B1: 狗熊kou ɕyŋ (北京), 狗熊儿
B2: 狗黑, 狗黑子kou xe tsɻ (徐州)
B3: 狗驼子kəu tʰuɤ tsɻ (白城)
C1: 黑熊xei ɕyŋ (哈尔滨)
C2: 黑狗熊xe kou ɕyŋ (徐州)
C3: 黑瞎子xei ɕia tsɻ (济宁)
C4: 黑傻子xei ʂa tsɻ (大连)
D1: 人熊iən xioŋ (柳州)
E1: 老熊lao ɕioŋ (蒙自)
E2: 大老黑ta lɔ xei (诸城)
F: kui (龙胜红瑶)
G: Others：山里棲 (庆元)

Forms of bear in Sinitic dialects show relatively complicated distribution, however it may be interpretated most types are descendant of 熊 type.

Monosyllabic type of 熊(A1) are distributed all over China. Many dialect forms of 熊 have glottal or velar fricative onsets like [x-][h-][ɕ] (including palatalized one) and velar nasal ending [ŋ]. These forms correspond to Middle Chinese form. Some Min 閩 dialects have forms with labial nasal ending like [im],[am].These forms may preserve Old Chinese form. Reconstructed forms of 熊 of Middle Chinese and Old Chinese are shown below.

熊	1	2	3	4
MC	ɣĭuŋ	jiuŋ	hjuwŋ	-
OC	ɣĭəu	gium	*C.[ɢ]w(r)əm	*wəm

1:郭錫良(2010), 2: Karlgren (1957[1997]), 3: Baxter & Sagart (2014), 4: Schuessler (2007)

In the north east area forms containing 黑(black) or 瞎(blind) are distributed (Map 3). These forms have fricative initials corresponding to 熊, and 黑 type is distributed peripheral area of 瞎. 黑 may be the form which 熊 lost the ending [-ŋ].Types like B2 狗黑 or E1 大老黑 may support this scenario. 瞎 may be the palatalized form of 黑, however further research is needed.

Types containing 狗 form are distributed mainly in northern dialects (including south west area). 狗 means dog so if it uses as modifier, 狗熊 means 'dog-like bear'. However, 狗 is also often used as prefix, e.g. 狗蚁(ant), 狗夹(earwig), 狗毛虫(caterpillar) and so on., so 狗 of bear is also possible to be a kind of prefix.

(YAGI Kenji)

A.
  A-1　熊ɕyŋ, am, hoŋ
  A-2　熊瞎子ɕyŋ ɕia tsʅ
  A-3　熊人hoŋ iɛn
  A-4　熊家 iuŋ ka
  A-5　熊嘎婆 ɕioŋ ka pʰo
  A-6　熊婆 iuŋ pʰɔ
  A-7　ʃan mɛ

B.
  B-1　狗熊kou ɕyŋ
  B-2　狗黑，狗黑子kou xe tsʅ
  B-3　狗驼子kəu tʰuɤ tsʅ

C.
  C-1　黑熊xei ɕyŋ

C-2　黑狗熊xe kou ɕyŋ
C-3　黑瞎子xei ɕia tsʅ
C-4　黑傻子xei ʂa tsʅ

D.
  D-1人熊iən xioŋ

E.
  E-1老熊lao ɕioŋ
  E-2大老黑ta lɔ xei

F.
  F-1 kui

G.
  G-1 Others：山里棲

Figure 6.5.1: 'Bear' in Sinitic.

‘Bear’ in Hmong-Mien

There are six types in BEAR: A: *kləp*; B: - (no word for the entry); C: *ɕoŋ*; D: *zuŋ*; E: *klai*; F: *mi loŋ*.

Type A exhibits the widest distribution both geographically and phylogenetically, thus strongly suggesting that this type is the most archaic. What is interesting in BEAR is that there are many lects that do not have a word for the entry. The lects that exhibit the data gap are distributed in the northern edge of the entire region of Hmong-Mien. This probably corresponds to the gap in the habitat of Asian black bear in Southern China.

(TAGUCHI Yoshihisa)

Figure 6.6.1: ‘Bear’ in Hmong-Mien.

'Bear' in Kra-Dai

Type A1 is dominant in the whole area, which should be the original form of Kra–Dai. Li (1977) reconstructed *hmi A based on sound correspondence and Siamese orthography. A2 can be interpreted as a denasalized form plus the prefix "lao" 老 borrowed from Sinitic. The stem of A3 is treated as a phonological variety of type A.

Type B is distributed to the places next to the Sinitic speaking area. B1 and B2 are varieties of Sinitic xiong 熊 meaning 'bear', and B3 originated from the Sinitic gouxiong 狗熊 meaning 'bear' as well.

Types C to G are scattered in many locations, mainly belonging to the Kra branch. They differ from one another; it is difficult to infer their formation process.
(ENDO Mitsuaki, TOMITA Aika, and HIRANO Ayaka)

A. *mi* type
- A1: $mɑi^1$, $məi^1$, me^1, mei^1, $məi^{24}$, mey^1, $mɤɤy^2$, $mɤy^1$, $mi:^{A1}$, mi^1, mi^{322}, $mī^6$, mi^{A1}, $miai^1$, mii^1, mii^2, mii^5, mje^2, $mɯay^1$, $mɯay^2$, $mɯəi^1$, $mɯi^1$, $mo:i^1$, mo^1hui^2, moi^1, mou^4, $mu:i^1$, $mu:əi^1$, $mu:y^1$, $muəi^1$, mui^1, mui^4, $pwa^2mɛ^1$, $qɑ^2mjɔ^{23}$, $ta^0mɛ^{312}$, $tə^0mie^{33}$, $ʔmi^1$, $lɔ^{33}mi^{55}$
- A2: $lɔ^{33}pi^{45}$
- A3: $a^{44}mua^{55}$

B. *jong* type
- B1: $hioŋ^4$, $huŋ^4$
- B2: $jɔŋ^2$, $jiuŋ^{21}$, $joŋ^2$, $juŋ^2$
- B3: $kau^3 juŋ^2$

C. *lauli* type
$lau^{55}li^{33}$

D. *khui* type
$khũ:i^1$

E. *khuangtci* type
$khuaŋ^{42}tɕi^{44}$

F. *nanglung* type
$naŋ^{24}luŋ^{24}$
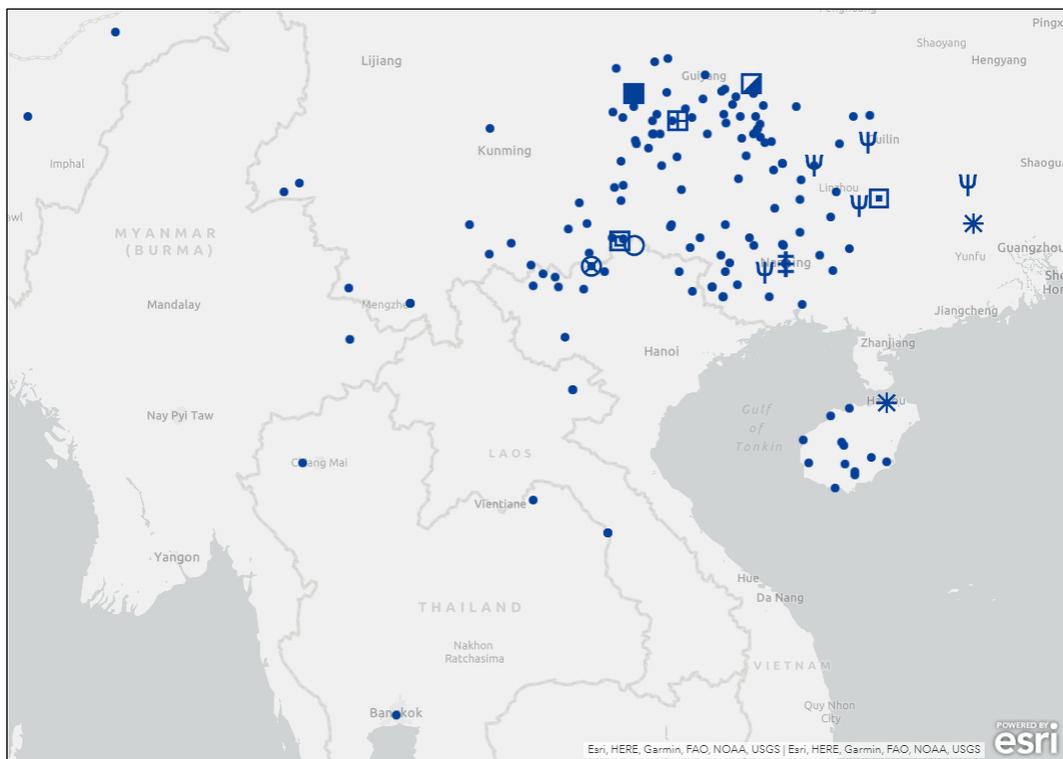
G. *dza* type
dza^{33}

Figure 6.7.1: 'Bear' in Kra-Dai.

'Bear' in Tibeto-Burman

There are three major stems (word roots) for 'bear' in TB. Two of them are etyma of the proto-level forms of Proto-Tibeto-Burman (PTB), Proto-Loloish (PL) and Proto-Southern-Qiang/Rma (PSQ). The word forms contain word formations that consist of a single stem, a stem plus an affix, or two compound stems. We thus first classify the TB word forms for 'bear' according to stem types and compound types. A list of three stems is as follows:

A type *(k-)d-wam* is derived from *d-wam* (BEAR (animal)) in PTB, and *(k-)d-wam* (BEAR (animal)) in PL etyma. These roots are recognized as cognate with Chinese 熊 *xiong* (OC *d-wam*) (STEDT). In PL etymon *k-d-wam¹, k-* is an animal prefix (see Matisoff 2003: 138-139). Word forms of Loloish and some other languages correspond to *(k-)d-wam*. The etymology of Type B *dred* is derived from Written Tibetan (*dred* 'brown bear'). Tournadre & Suzuki (2022) mention that Type B is widespread in the northern part of the Tibeto sphere. A morpheme *pra* in Type X might be related to this type. The etymology of Type C *yiel* is unknown. The form of Type C is derived from the word form $z\varepsilon^{31}mi^{35}t\mathfrak{o}^{31}phu^{53}$ 'bear' which originally has a Type A word root $t\mathfrak{o}^{31}$ as a head, and it lost the lexical root $t\mathfrak{o}^{11}$. This might be related to a custom to avoid speaking out taboo words, such as 'bear.'

In addition to the three major types, there are several marginal roots, labelled as Type X. *tɕi* (Bai) might be related to Type A (*(k-)d-wam* > *tiam* 'bear' in PSQ etymon), and *rədaʔ* (in Basum) is related to *ri dwags* 'wild animal' in WT. The other word forms are etymologically unknown.

Most of the above-mentioned forms consist of a single stem, but we also found two types of compound forms. *myi* means 'human' in WT. *myi*+B might be a loan translation from Chinese 人熊 *renxiong*.

In some languages and dialects, two or three word-forms coexist: two word-forms (A and B, B and *ri dwags),* and three word-forms (A+*pra*, *dza ŋi rʌ,* and *dza trui).* In some dialects of Amdo Tibetan which have both A and B, A is a general term for 'bear,' and refers to 'black bear,' whereas B refers only to 'brown bear'.

Type A is the most widespread across the branches of TB. This type is found in the northern and eastern parts of the TB area (Tibetic, rGyalrongic, and Qiangic groups), and in the south-eastern part (Lolo-Burmese). Moreover, a compound with Type A (A+*pra*) is found in Qiangic. Type B is distributed widely in Tibetic, and sTau. Type C is only found in the southwestern part of Yunnan province (Lahu).

In terms of the distribution, Type B, which refers to 'brown bear,' is found only in Tibetic located in northern part of the TB area. It is noteworthy that this type is not found in the southern part. This fact might be relevant to the habitat of 'brown bear'. Furthermore, the area where both Type A and B are used (Amdo, and the northern part of Khams Tibetan) overlaps the area where both black and brown bear live.

(EBIHARA Shiho, IWASA Kazue, KURABE Keita, SHIRAI Satoko, SUZUKI Hiroyuki)

✏ A. *(k-)d-wam* type

to:m^{13}, tɑwɑm, tom, ton, tho:m^{13}, thɛ^{55}, thɔ^{55}, tʰán, tvóeeŋ, doŋ, kom, cap^3, gom, goŋ, gĩ^{35}, gɣ^{21}, ɣɔ, ŋuɛ̃^{35}, ŋu^{33}mu^{53}, ze^{21}, xom^{35}, vɔɔ̃, wom^{35}, woms, wɛr^{13}, wɛ̃^{55}, hom^{35}, ɣuɛ̃^{53}, lúwaiŋ, sap, tsáp, ɔm^{35}, uɛ̃^{5}, ^{1}e^{2}we; ɛngwi, ã^{31}gui^{55}, bhə.li.yom; tɐ^{13}nɑ^{53}, khu^{21}tɕhi^{21}, wǒ-pa, ɣɒ^{21}mɔ^{55}, je^{11}mɒ^{55}, ɣu^{44}mA^{21}, zi^{21}bɑ^{21}, vɛ^{31}ti^{55}, ɔ^{31}je^{55}, vompi, etc. (suffixed); ɕathom, ahoŋ, kəʃtʃək, kəsʰàp, re^{33}we^{55}, zi^{21}ba^{33}, chaba, xa^{31}ɔ^{55}, xɔ^{31}ɣ^{55}, a^{33}ø^{44}, zɛ^{31}mi^{35}tɔ^{11}, iɹəm, ʔə-wám, si-tĩ, su-tum,

sevan, sitom, ʃap^{33}daw^{55}, ʃəwi^{53}, thok-wam, chawom, wɛʔwùn, ovu, (prefixed*)*; zɛ^{31}mi^{35}tɔ^{31}phɯ^{53} (affixed).

⭘ B. *dred* type

tʂel, pʈet, ʁre, etc.; drenmo, ʈe^{11}mɔ:ŋ^{55}, ʈe^{11}po:ŋ^{55}, ʈe^{22}wo:ŋ^5, etc. (suffixed).

△ C. *yiel* type

zɛ^2, zɛ^{2}pa^2, ɣɛ^{2}pa^2, ɣɯ^2

⌄ X. others

tɕi, ɕou, bhālu, pra, mu-pur, thega, rədɑʔ, dzinami, dZinAmi, tɕhi^{21}ku^{55}, wi̥wũ̀
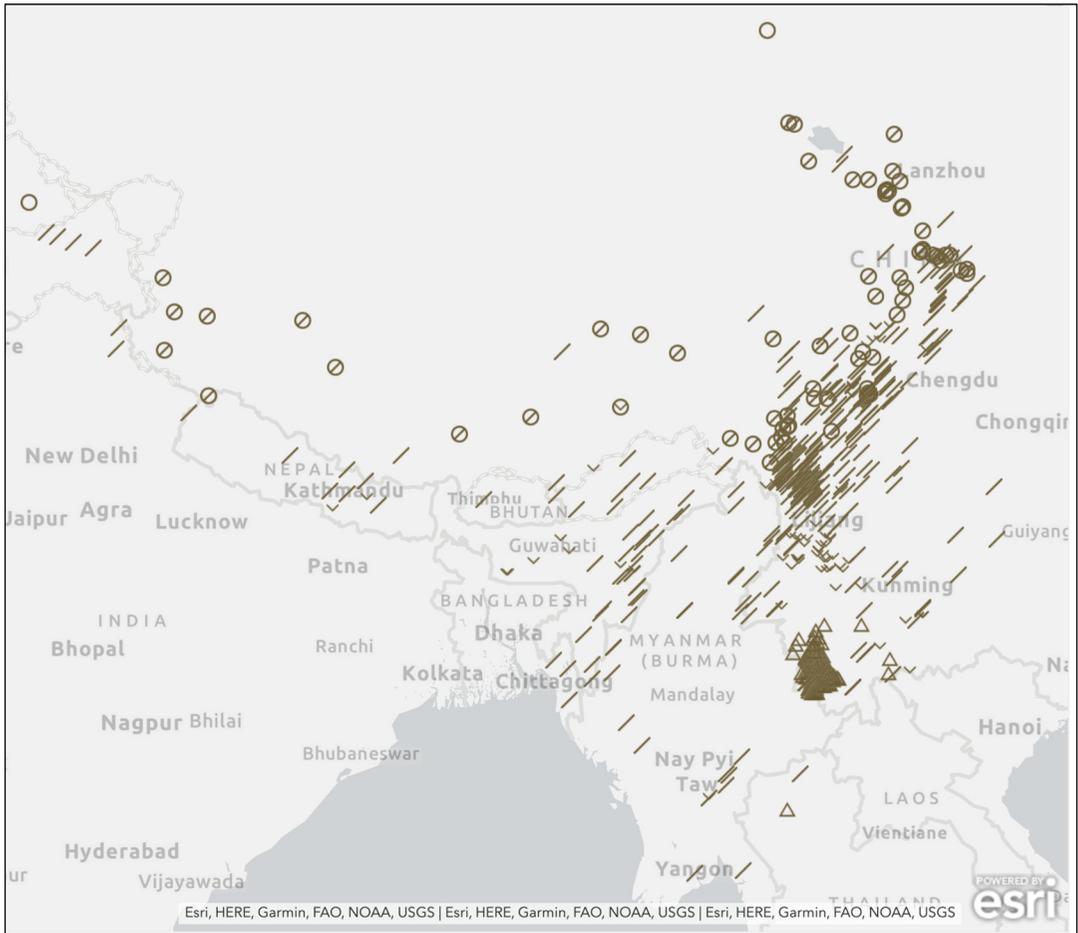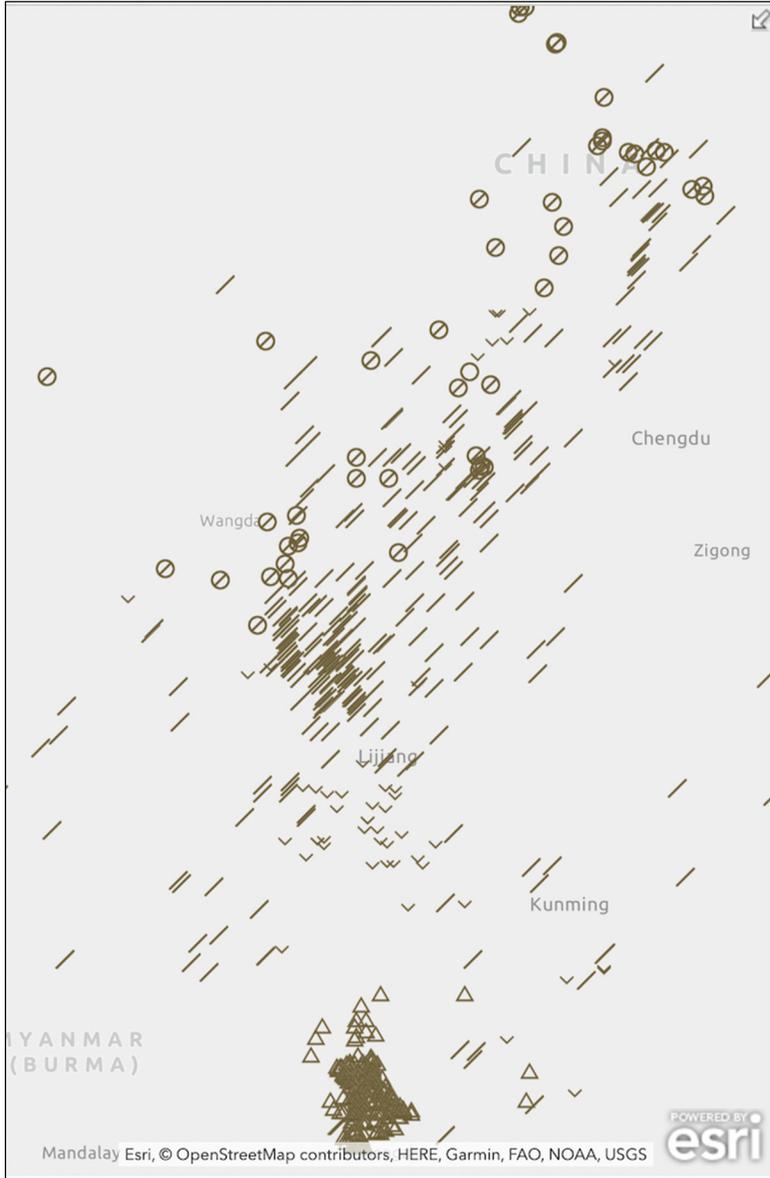

Figure 6.8.1: 'Bear' in Tibeto-Burman.

Figure 6.8.2: 'Bear' in Tibeto-Burman (detailed).

'Bear' in Austroasiatic

The word forms meaning "bear" in Austroasiatic are classified into nine types, as follows.

A. ɹkaw type
Proto MK: *ɹkaw (Shorto 2006)
Proto Bahnaric: *ckaw~*gaw (Sidwell 2011); cakaw (Alak), cəgəw (Bahnar [Pleiku]), cəkaw (Sapuan)
Proto Katuic: *hŋkaw (Sidwell 2005); hankʌːw (Ngeq), haŋkaw (Ta'Oi [of Sekong])
Proto Vietic: *c-guːʔ/c-kuːʔ (Ferlus 2007); cakù: (Chứt [Rục]), təkow⁴ (Malieng)

B. [k]r[e]s type
Proto MK: *[k]r[e]s (Shorto 2006)
Proto Palaungic: *krees (Sidwell 2010); krer (Palaung), kres (Lawa [Umphai])

C. [k]mum type
Proto MK: *[k]mum (Shorto 2006)
Proto Monic: *kmṳm (Diffloth 1984); hmúm (Nyah Kur [Central]), khəmúm (Nyah Kur [Klang])
Khmeric: klaa kmum (Khmer)

D. biːʔ type
Proto Pramic (Khmuic): *biːʔ (Sidwell 2013); biːʔ (Phong), biː (Tai Hat)

E. bɛːk type
Proto Pray-Pram (Khmuic): *bɛːk (Sidwell 2013); biɯk (Mlabri)

F. suəl type
Proto Khmuic: *suəl (Sidwell 2013); suːl (Khsing-Mul)

G. bana type
Munda: *bana (Proto Kherwarian: Munda 1968)

H. rəwaj suːt type
Pearic: rəwaj suːt (Chong)

I. kawap type
Aslian: kawap (Kensiu)

The word forms for "bear" in Austroasiatic are quite diverse. Among them, three forms are reconstructed as the proto Mon-Khmer forms: A type ɹkaw, B type [k]r[e]s and C type [k]mum.

The forms succeeding A type ɹkaw are spread among Bahnaric, Katuic and Vietic, along the east coast of mainland Southeast Asia. The Bahnaric and Vietic languages preserve the proto sesqui-syllabic forms, while Katuic has developed a new presyllabic form, *hŋ-, from the proto *ɹ-.

The B type [k]r[e]s is distributed across the Palaungic (Palaung and Lawa) area.

The C type [k]mum is distributed throughout the Monic (Nyah Kur [Central] and Nyah Kur [Klang]) and Khmeric (Khmer) areas.

Khmuic languages distributed around Northern Laos have three different forms: D type biːʔ (Phong, Tai Hat), E type bɛːk (Mlabri) and F type suəl (Khsing-Mul).

As for the other types, each language group has its own forms: G type bana for Munda, H type rəwaj suːt for Pearic and I type kawap for Aslian.

(SHIMIZU Masaaki,
MINEGISHI Makoto)

Figure 6.9.1: 'Bear' in Austroasiatic.
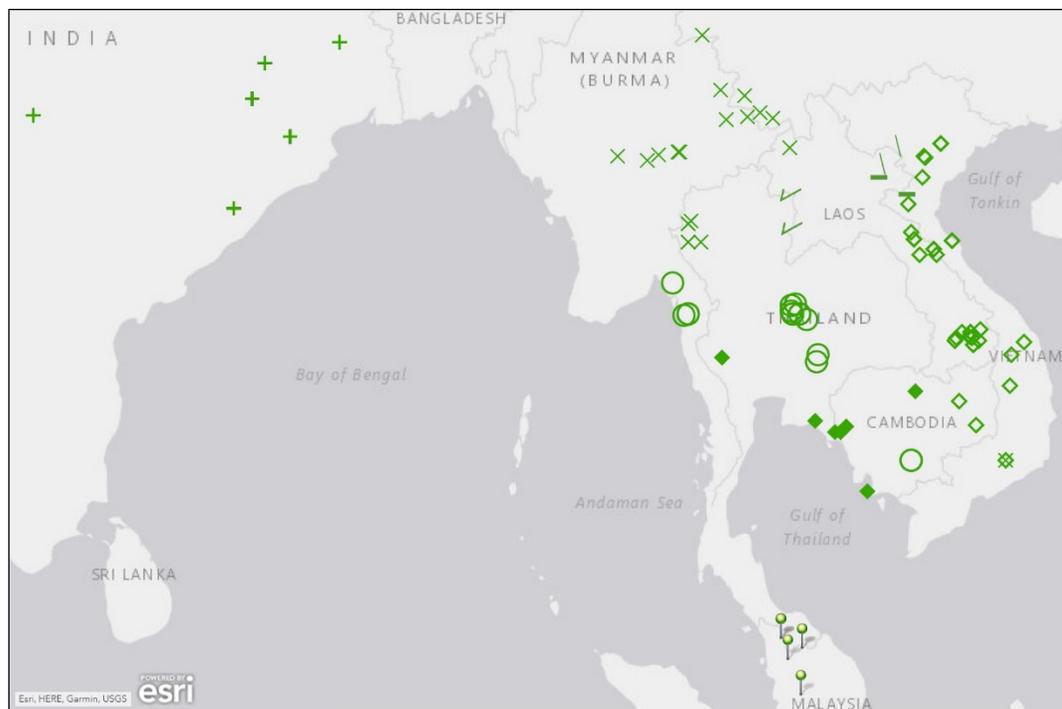
◇	A	*ɹkaw type*
×	B	*[k]r[e]s type*
○	C	*[k]mum type*
⁻	D	*biːʔ type*
↙	E	*bɛːk type*
\	F	*suəl type*
+	G	*bana type*
◆	H	*rəwaj suːt type*
⸙	I	*kawap type*

'Bear' in Austronesian

Among the regions where Austronesian languages are spoken, bears are typically prevalent in Taiwan, Malaysia, and large islands of Indonesia. In contrast, fewer bears are found in the smaller islands of Indonesia and the Pacific; consequently, there are no specific words for "bear" in these regions. Furthermore, there is much variation in the words denoting "bear." Among these variations, words containing the alveolar affricate /ts/ and /m/ as well as those with /b/ and /ŋ/ are found in multiple languages. Loanwords derived from the English term *bear* are also prevalent among languages spoken in the Pacific, where the animal has hardly existed in recent history.

We can divide word forms for 'bear' into four major types (Types A to D) and other variable types which are altogether included in type E. Types A and B are found in Formosan languages (Taiwan), whereas Type C is predominant in Indonesian languages, particularly Java and Sumatra.

Type A includes word forms containing /ts/ and /m/ such as *tsmoi* (Tsou), *tsomay* (Rukai), and *tsumay* (Paiwan). Type B represents a form *ōso* which is found in Tagalog and Aklanon. Type C consists of word forms containing the word-initial /b/ and word-final /ŋ/ and often entailing the word-internal /r/: *bauaŋ* (Murut), *biruaŋ* (Minangkabau and Sundanese), *bəruaŋ* (Indonesian), *bruaŋ* (Javanese), *baruaŋ* (Bugis), etc. Type D consists of loan forms from English 'bear', such as *ber* (Yabem), *bea* (Roviana and Eastern Fijian), *pea* (Rotuman, Tongan, and Tahitian), and *beya* (Kilivia). Other types are included in type E, for example, *ŋarux* (Atayal), *kokoman ṣo tao* (Yami), *mɔnturun* (Palawan), *ursa* (Malagasy Merina), *cagɛə* (Aceh), *gɔppul* (Batak Toba), and *urosa* (Samoan).

(UTSUMI Atsuko)

⚡ A: word forms containing /ts/ and /m/ *tsmoi*, *tsomay*, and *tsumay*

⌒ B: *ōso*

▽ C: *bauaŋ*, *biruaŋ*, *bəruaŋ*, *bruaŋ*, *baruaŋ*, etc

✺ D: *ber*, *bea*, *pea*, *beya*

✚ E: Various forms

Figure 1.1.1: 'Bear' in Taiwan and the Phillipines

Figure 1.1.2: 'Bear' in Indonesia
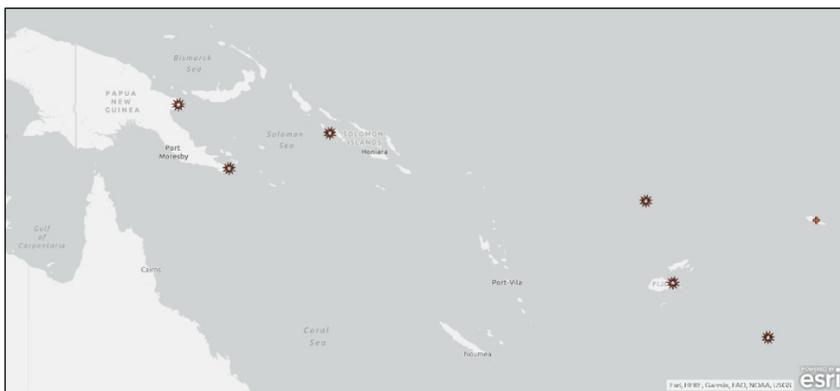
Figure 1.1.3: 'Bear' in Papua and the Pacific

'Bear' in Tungusic

In Tungusic languages the word forms for 'bear' would be classified in some types as below:

A AMAKA: Evenki *amaka*, Negidal *amaxa*, Orochon *ətirkən*, Ewenke *ətəggəŋ*

B NAKAT: Ewen *nakat*

C MAPA: Nanay *mapa,* Oroch *mafa*, Udehe *mafa,* Hezhe *mafka*

D Sibe *ləf*

Tungusic words for 'bear' are phonetically very similar each other. Type A and Type C differ only in the initial vowel – it is not clear whether the initial vowel was added or deleted. Type C shows the alternation of the initial nasal *m/n*. Only Type D has another from the others.

There is also the semantic similarity. For example, *amaka* in Evenki and Negidal, *ətirkən* in Orochon (Chinese Evenki) etc. means on only 'bear' but also 'grandfather', because they use it for the taboo word to avoid inviting the fiercest creature in Siberia by calling its real name. Such a rhetorical way about bear is found in other languages in Siberia.

(MATSUMOTO Ryo)

○	A	AMAKA
	B	NAKAT
	C	MAPA
☆	D	*ləf*

Figure 6.11.1: 'Bear' in Tungusic.

'Bear' in Uralic

In Uralic there are various forms for 'bear'. Here I just classified according to the sound forms, it could not be referred to the reason why such many forms they have.

A Finnish *karhu*, Ingrian *karhu*, Estonian *karu*, Votic *karu*

B Veps *kondi*, Karelian *kondii*

C Moksha (Mordvin) *ofta*, Erzya (Mordvin) *ovto*, Livonian *okš*

D Sami *bierdna*

E Mari *maska*

F Mansi *mojpyr*

G Khanty *pupi woj*

H Hungarian *medve*

I Komi *rudoš*

J Selkup *ķorķy*

K Tundra Nenets *wark*

L Forest Nenets *app'ji*

M Enets *boglja*

There are too different forms for 'bear' in Uralic. Even in the branches of Uralic family no common words are observed except between some dialects of Balto-Finnic branch. It must be because they avoided using the original word not to call the harmful beast to the human life, which is a same situation as the word for 'wolf.'

(MATSUMOTO Ryo)

- ╱ A *karhu, karu*
- ⌒ B *kondi*
- ● C *ofta, ovto, okš*
- □ D *bierdna*
- ✛ E *maska*
- ✖ F *mojpyr*
- Ψ G *pupi woj*
- ☆ H *medve*
- ◊ I *rudoš*
- ✖ J *ķorķy*
- ⊥ K *wark*
- ⊠ L *app'ji*
- ⬤ M *boglja*

Figure 6.12.2: 'Bear' in Uralic.

'Bear' in Mongolic and Turkic

1. Mongolic

Mongolic languages in the Mongolian plateau and its vicinity have *bābgai*-type words for 'bear.'

In languages in southern Mongolia, the word *xar görō̄s* and its variants, which literally mean 'black wild beast,' and the word *noxǽ xar görō̄s* and its variants, which mean 'dog black wild beast,' are used. They are loan translations of Chinese words for 'bear.' (Cf. 黑熊, lit. 'black bear'; 狗熊, lit. 'dog bear')

Dongxiang in Gansu province has the forms *gəušin* and *xašin* (< Chinese 狗熊).

Oirad people and their modern branch in the lower Volga region, the Kalmyks, use a Turkic word *ajū*.

In Qinghai and Gansu provinces, Baoan has *dɛrmoŋ*, which may be related to the Tibetic forms *drenmo* and *ʈemɔ̄ŋ* (cf. Ebihara et al. 2022), and Kangjia has the form *mɔsɔkɔ*.

Dagur in Heilongjiang province has *ətərkᾱn* as well as *bᾱbəg*.

Moghol in Afghanistan uses a Persian loanword, *xirš*.

2. Turkic

The words of *adïg* type and *ajï* type are cognate and are descendants of Old Turkic *aðïg* 'bear.' Most languages distributed from Xinjiang to Europe share these types of words.

The Turkic languages in Iran and Afghanistan use a Persian loanword, *xïrs*.

In Siberia, due to the local folklore about bears, words meaning 'old man, grandfather' are used to denote bears (cf. *apā, ebē, epē* 'ancestor, grandfather/ grandmother, father/mother'): *up*á (Chuvash), *aba* (Khakas); *apšaq* (Shor); *ehe* (Sakha); *ebeke* (Dolgan). In addition to *ehe* ('grandfather'), Sakha uses the word *tïatāɣï* ('forest dweller') for 'bear.'

Tofalar in southern Siberia has the forms *ire*, *irezaŋ*, etc., which may be related to the Samoyedic word for 'grandfather, old man.'

Sarïg Yughur in Gansu province uses a Chinese loanword, *gəušuŋ* (< 狗熊).

(SAITÔ Yoshio)

A. *bābgai* type
- ⌣ bābgai, bambu, bə̄bug, bə̄bəg

B. loan translation type
- ╱ **B-1.** *xar görȫs* type
  xar görȫs, xara görösən, xar gyrēs, xar gɯrə̄həi
- ╲ **B-2.** *noxǣ xar görȫs* type
  noxǣ xar görȫs, noxō xar görȫs, noxoi xar gɯrēs

C. mɔsɔkɔ
- △ mɔsɔkɔ

D. ətərkə̄n
- ∠ ətərkə̄n

E. loanword
- ‖ **E-1.** dɛrmoŋ (< Tibetic)
- ⚡ **E-2.** gəušin, xašin (< Chinese)
- ✚ **E-3.** xirš (< Persian)
- ▯ **E-4.** ajū (< Turkic)

F-1. *adïg* type
- — adïg, adəɣ, atəχ

F-2. *ajï* type
- ▯ ajï, aju, ajïv, ajju, ajuv, ajū, ajïq, ejiq

G. *aba* type
- Ƴ **G-1.** aba, upá
- ⌠ **G-2.** apšaq
- Ψ **G-3.** ehe
- Ƴ **G-4.** ebeke

H. loanword
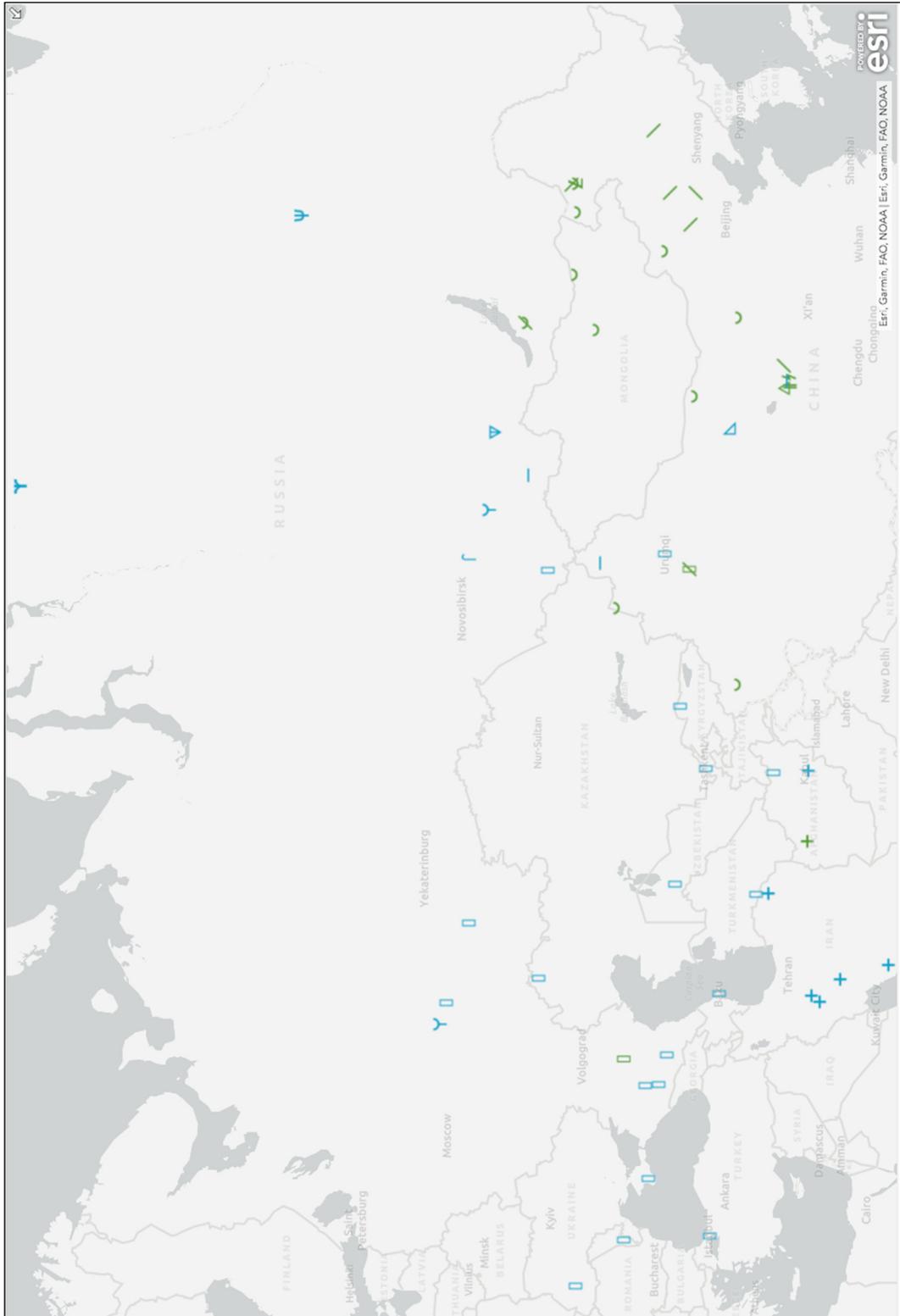- ▽ ire, irezaŋ, ireaŋ, irej (< Samoyedic)
- ◣ gəušuŋ (< Chinese)
- ✚ xïrs (< Persian)

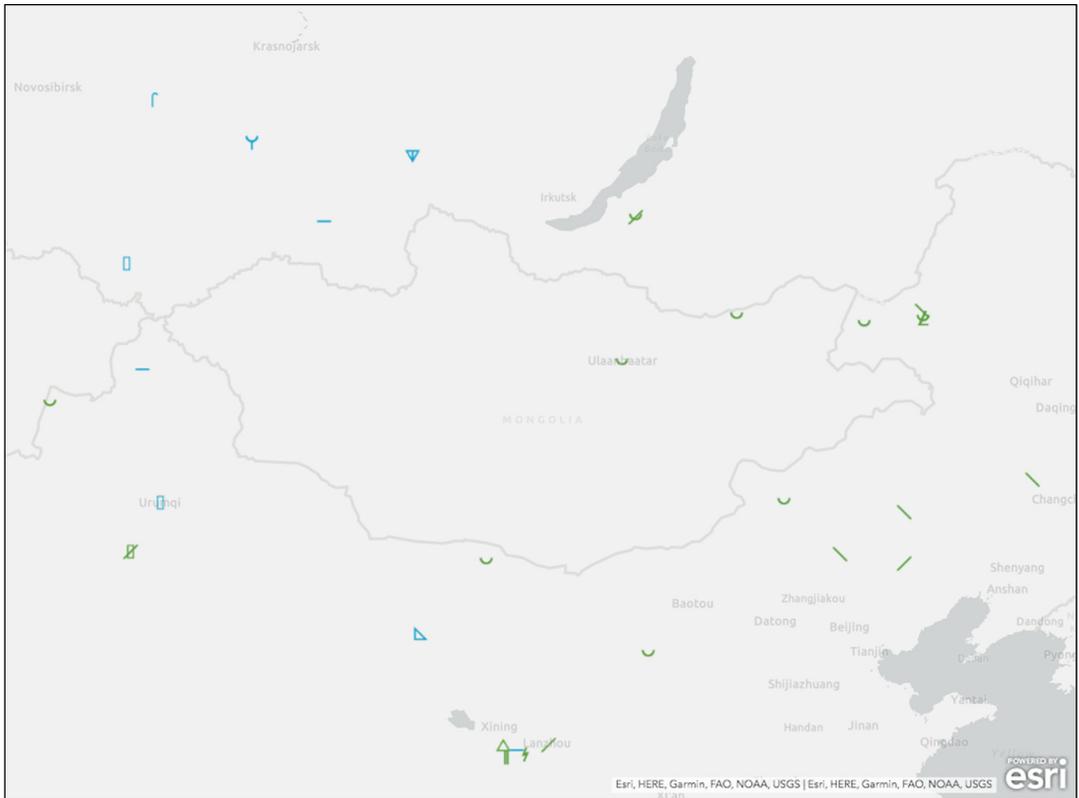Figure 6.13.1: 'Bear' in Mongolic and Turkic.

Figure 6.13.2: 'Bear' in Mongolic and Turkic (The Mongolian Plateau and its vicinity magnified).
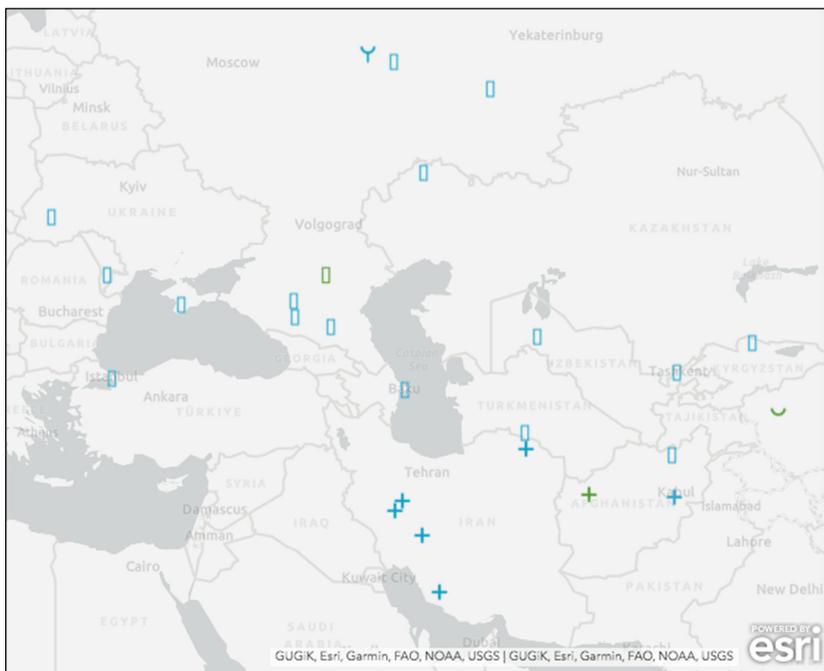

Figure 6.13.3: 'Bear' in Mongolic and Turkic (Central Asia to East Europe magnified).

'Bear' in South Asia

I describe the languages of Indo-Aryan (IA), some small language families/branches, and language isolates in South Asia. When a language has several words for bears, I targeted 'male adult bear'. Both black and brown bears are treated here.

The distribution of 'bear' words is partially multilayered. Indo-European, widely employ the type A except in the east and in the islands. The type B is employed by IA and Nihali in Sri Lanka and inland from central Pakistan to Nepal, Bangladesh, and eastern India. In and around the Kashmir region, Type C is distributed, and the centre of the area, that is, in the east of Indian Kashmir they use froms of the type E.

The most major type is A. This type is derived from Sanskrit *ŕkṣa* ऋक्ष 'bear' (< Proto-IA **Hŕ̥ṭsas*) or Proto-Nuristani **írca*, which are inherited from Proto-Indo-Iranian **Hŕćsas*, ultimately Proto-Indo-European **h₂ŕ̥tḱos* 'destroying'. So this type is cognate with Latin *ursus* and Ancient Greek *árktos* ἄρκτος. Forms of this type are used by many IA languages, including languages outside South Asia. Some languages use forms derived from the combination of the two Sanskrit words *ŕkṣa* and *bhalla* भल्ल, both of which refer to 'bear'.

The type B appears in IA languages and Nihali, which are spread over the South Asia. The original Sanskrit form is *bhalluka* भल्लुक 'bear', and also may be *bhalla* already mentioned above. *bhalluka* (alternative form is *bhallaka* भल्लक) is made of *bhalla*, which originally refers to 'auspicious, favourable'. According to Monier-Williams (1899: 748), the word *bhalluka* also refers to 'monkey' and

the similar word *bhallūka* भल्लूक can mean 'dog' as well as be a component of a word referring to 'bear cub'.

The type C can be seen in IA languages in and around Kashmir and Waigali, a Nuristani language. The forms originate in Sanskrit *babhru* बभ्रु 'reddish brown' or **bhrāru* *भ्रारु 'brown animal', both derived from the PIE root **bherH-* 'brown, bright'. Some Western Pahari languages employ the type-C words for 'brown-bear', as distinguishing it from 'black-bear' by the type-A terms. In Waigali, however, they employ the type-C word *brō̃* just for 'black-bear' and use the type-A word *ōc* for 'brown-bear' contrary.

Next, the *nya* type is of Burushaski languages, but is also detected in Domaaki. They are all spoken in the Karakoram mountain range. The actual forms differ in three major Burushaski dialects, namely *nya* in Yasin (in west), *ỹa* in Hunza, and *yã* in Nager (both in east). In Domaaki, the form is *ya*, directly borrowed from Hunza Burushaski.

The type E is of *śvāpáda*, a minor type in some Kashmiric lects. The source form is the Sanskrit word *śvāpáda* 'beast of prey, wild beast' made of *śvan* 'dog' and *páda* 'foot'.

Besides the types, there remains some other words for 'bear'. *kaḷōttī* 'black bear' in Chambeali is derived from Sanskrit *kālakṛttika* कालकृत्तिक 'black-skinned'. Gurbet Romani *ruvdič* may be related to *ruv* 'wolf'. Kelderaš Romani *metčkoi* seems derived from Church Slavonic *mečĭka* мечъка 'lit. she-honey-eater'. *veḷkaḍḍu* in Saurashtra is surely borrowed from Dravidian. Urdu *dubb* دب is a loanword of Arabic *dubb* دب 'lit. walker'.

(YOSHIOKA Noboru)

A. *ŕkṣa* type (57) ╲

    rīkh, rikkh, r̃č̌h, r̃č̌h, ričhu, ričhini, rič̌čh, rič, riči, rīču, riš, r̃s̃, orc, ic, ēc, ōc, ich, ācha, içʌ, iç, ĭ̃ç, ẽç, āç, ĭ̃çh, ič̌čh, ẽ̌č, ančeakī, ĭş, ĩş, ẽş, ütrū, hirč, yãič, žãič

 [+*bhalla*]

    asval, ã̃svel

B. *bhalluka* type (17) ☐

    bhāluk, bhāllu, bhālũ, bhal, balluk, bologo, banae, valahā

C. *babhru / bhrāru* type (8) ⊐

    bhrabū, barabbu, bhrabbū, brō, dřenmo, ḍhḷabbū, ḍhḷɛbbū

D. *nya* type (4) ⊣

    nya, ỹa, yã, ya

E. *śvāpáda* type (3) ⬤

    šāput, hāput, ǎput

F. others

    kaḷōttī, mallū, ruvdič, metčkoi, veḷkaḍḍu, dubb, kärimañña
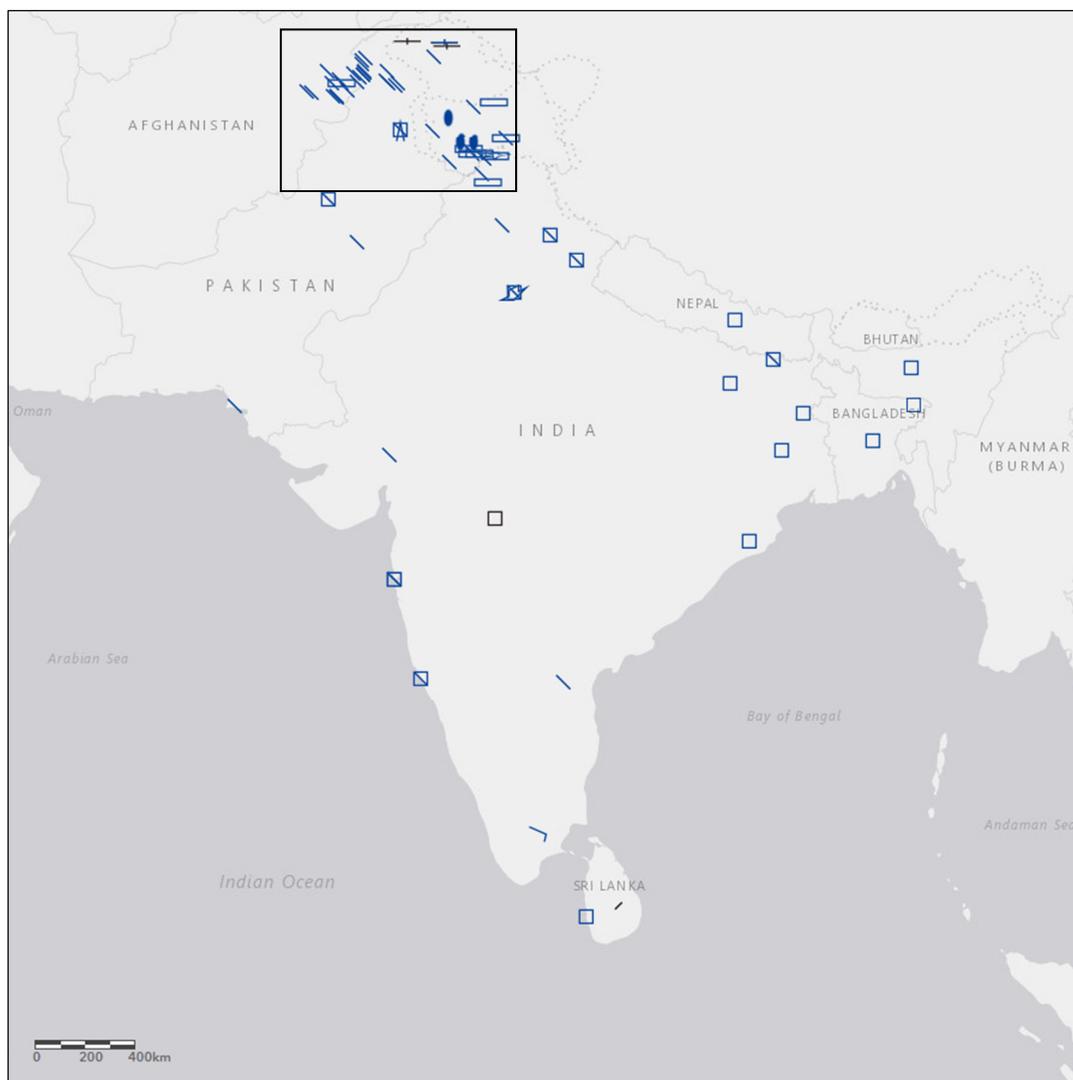

Figure 6.14.1: 'Bear' in SA: Indo-Aryan, Nuristani (both in navy blue), Andamanese, and language isolates (those in black).
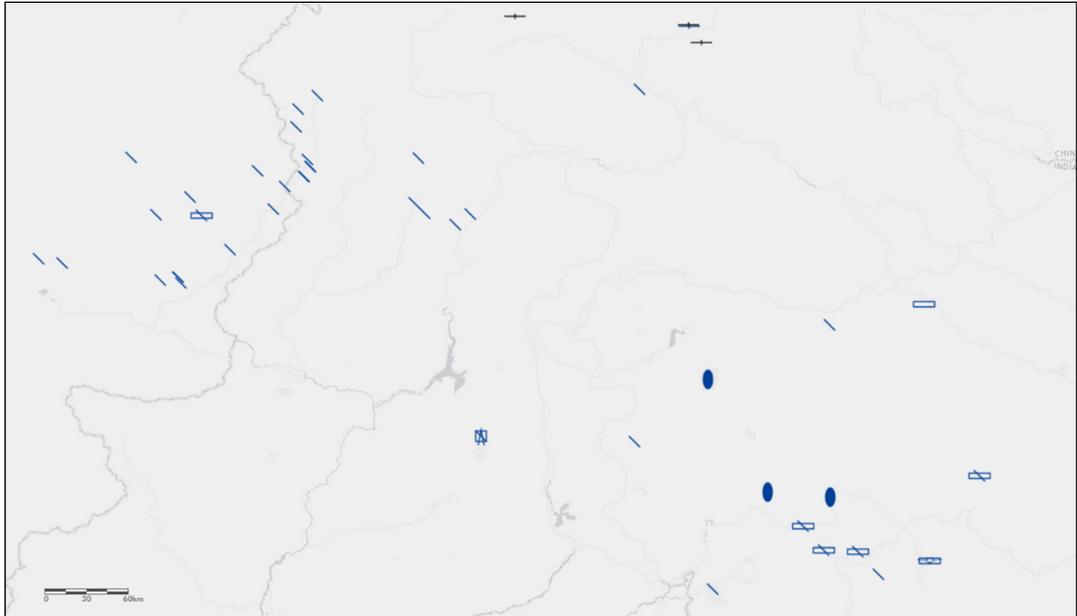
Figure 6.14.2: 'Bear' in northern Pakistan (the area encloed by the rectangle in Figure 6.14.1).
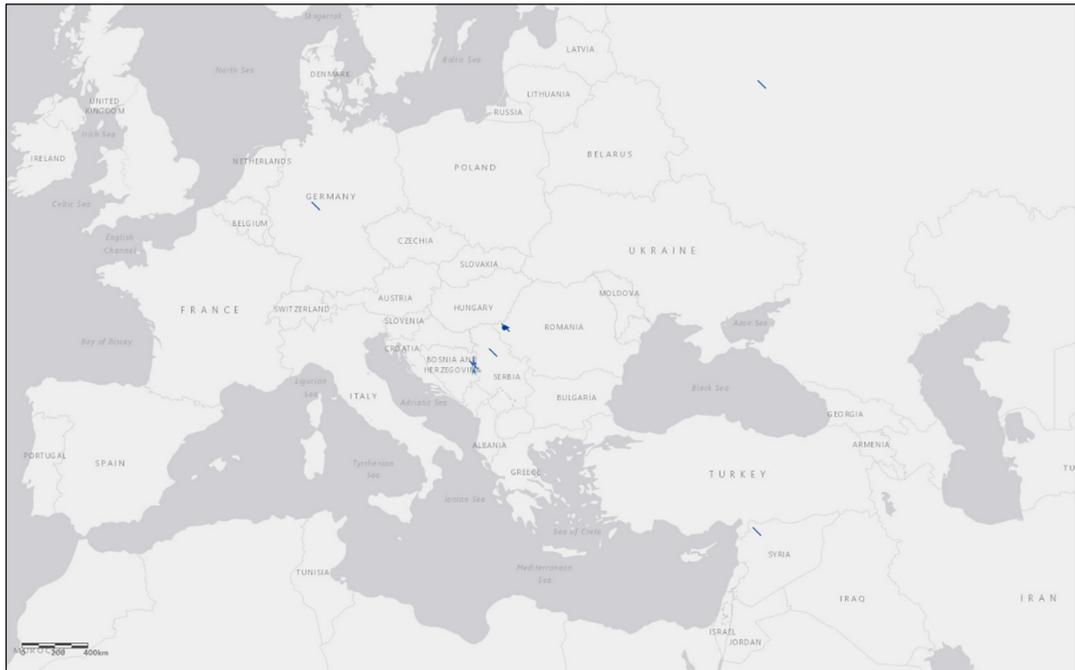
Figure 6.14.3: Types for 'Bear' in Indo-Aryan languages outside South Asia.

'Bear' in Dravidian

Bears of the genera *Melursus* (sloth bear) are endemic to the subcontinent where Dravidian languages are spoken while *Ursus* (Asian black bear) is distributed in the area Brahui is spoken.

The DEDR entry #857 is a rather loosely defined etymon which includes several Classical Tamil words meaning 'a bear' such as *eŋku*, *eḻu*, *uḻiyam* and *iḻai* along with mostly South Central and Central Dravidian forms such as *elugu*, *eṟj*, *oḍi* and *eju*. Interestingly, these words in South Central and Central Dravidian languages often resemble the word meaning 'a rat' in the respective languages, and besides, the Telugu derivative suffix -ēlu apparently meaning 'a creature' as in *kundēlu* 'a hare', *tābēlu* 'a tortoise' and *tōḍēlu* 'a wolf'.

The South Dravidian etymon KARAṬI (DEDR #1263) appears to be related to KAR 'black' (DEDR #1278) as Kolami and Naiki GUḌḌI (DEDR #1679) means 'black' in other Central Dravidian languages, which can be modified forms of taboo as in the case of replacement of Indo-Aryan *ŕkṣa* suggested in CDIAL.

Kurukh and Brahui have replaced the Dravidian form with Indo-Iranian borrowings.

(KODAMA Nozomi)

∨ elugu

∖ eṛj, eṛju, eṛjal, aṛjal

╱ ili, ilij, illij

+ oḍi, oli, ouḍi, o'ṛi, olzu

| eju

☆ intu, imai, iḷai, uḷiyam, eŋku, eḻu

◈ karaṭi, karaḍi, karoḍi

▰ ka·ṛ, karḍy

▱ guṭṭi

⬤ bhalu

◉ xiṛas
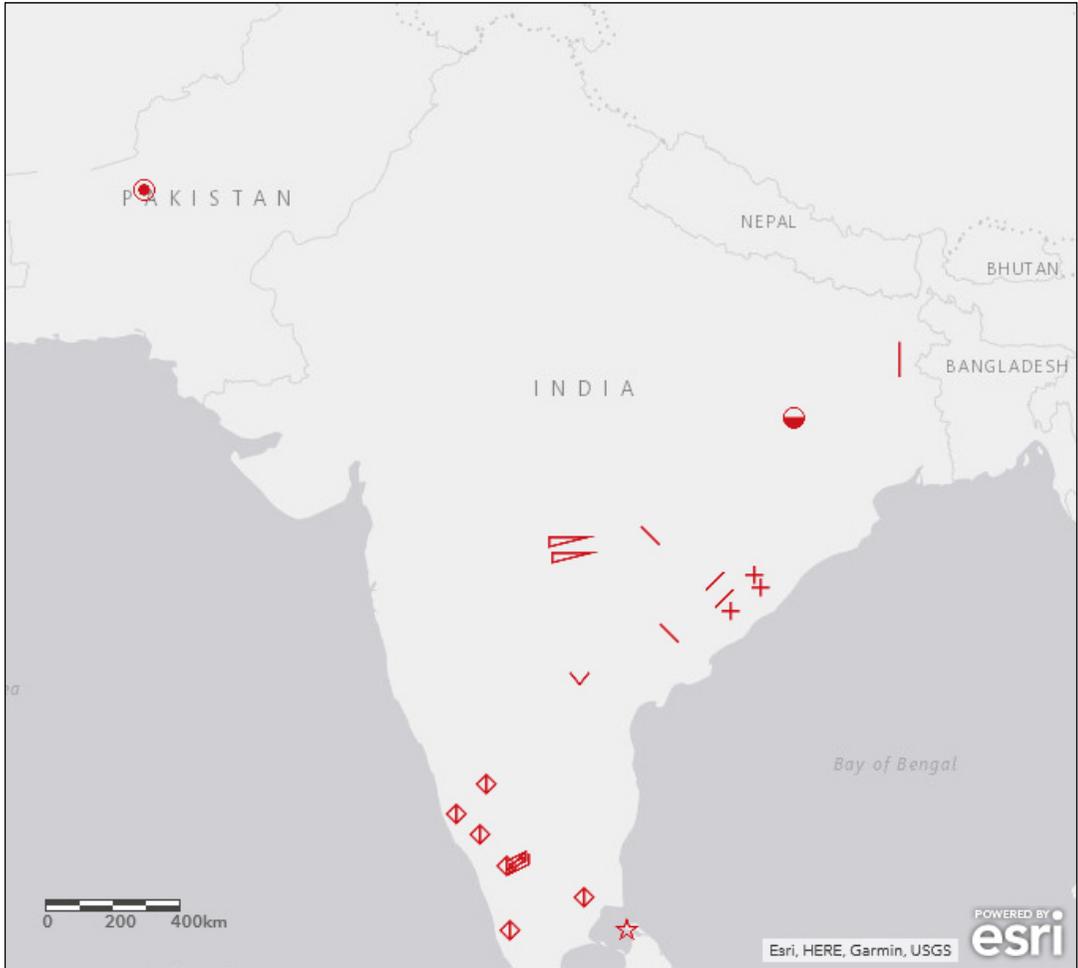
Figure 6.15.1: 'Bear' in Dravidian.

'Bear' in Iranian

With few exceptions almost all Iranian languages have Type A forms, which derived from the PIr word *(H)r̥k̑ca- (< PIE. *h₂r̥tk̑os) Note that many languages have borrowed from the Persian/Tajik/Dari equivalent forms *xers/ xirs* 'bear'. In some cases, both an inherited word and the borrowed one coexist. Rushani, for example, has both the inherited and borrowed words (an inherited word *yurxx* and borrowed one *xers*).

Along with type A, Pashto also has type B word. There seems to be no semantic difference between the two words (cf. *tor melu/ tor yəg* 'himalayan bear'). Its origin is unknown, but it may have relation to an Indo-Aryan word such as Hindi *bhālū* 'bear'. Balochi has Type C word (*mamm*), probably a loanword from a non-Iranian language, Brahui (Dravidian), coexistent with a word *rič* which is also borrowing from Lahnda, an Indo-Aryan language. Wakhi has unique Type D forms *noɣurdum* and *nəɣərdum*, whose origin is still controversial. Due to a heavy influence of Arabic, Kumzari has Type E, a loanword from the Semitic word دبّ *[dub]* 'bear'.

(IWASAKI Takamasa)

─ A: *xers* Type   B: *melu* Type   0 C: *mamm* Type   ▶ D: *noghondom* Type
◕ E: *dubb* Type
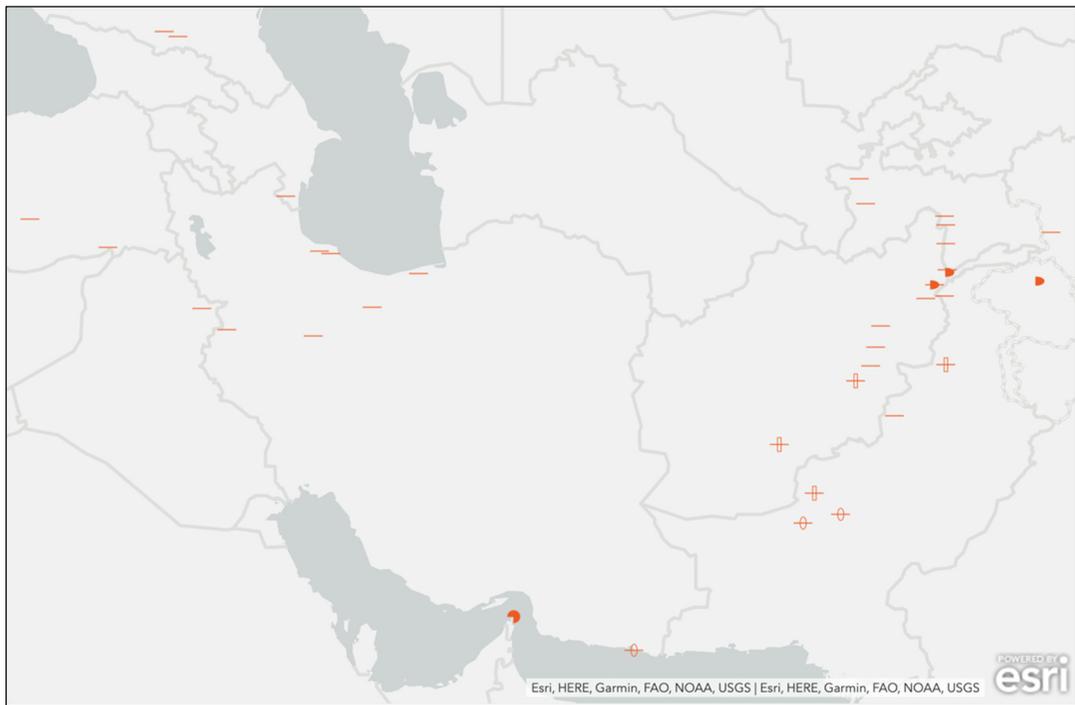
Figure 6.16.1: 'Bear' in Iranian.

'Bear' in Caucasian languages

Type A appears in Kartvelian languages, derived from a common root *datv-. Type B is found in Abkhazo-Adyghean languages. Type C appears in Nakh languages, and Type D is the main form in Dagestanian languages, containing three initial consonants /s/, /š/, and /z/. Types C and D appear closely related since Avar exhibits variants: *ci, si, ši,* and *či*. Avar is classified as Type D based on its vowel feature.

Types E and F are exceptional, and each of them appears only in a single language, namely, Type E in Archi and Type F in Agul. These are Dagestanian languages distributed in other sister languages; the possibility of borrowing is presumably low.

(SUZUKI Hiroyuki)

●A: d-type; *datv-i, tunt-i, mtuti,* etc.
▬B: mš-type: *a-mš^{j/w}, miš^wa, miš^jə,* etc.
■C: *ča*.

□ D: s/š/z-type; *si, sĩ, šĩ, sehĩ, ze, psɨ,* etc.
⌐E: *χχams*.
◗F: *bagniš, bangiš*.

Figure 6.17.1: 'Bear' in Caucasian languages.

'Bear in Semitic

All Semitic languages share the root *d-b*. Akkadian form is also *dabû* (Old Akk. *dabium*). Arabic (*dubb*) and Aramaic (*debba*) forms are results of the adding *b* for making a three consonantal root (*d-b-b*).

A. *dubb* type (•) is the Arabic form in Gulf, Yemen, Sudan, Palestine. Phonetic variations are *dibb* (Egypt, Lebanon, Iraq), *dəbb* (Syria), *ḍəbb* (Morocco).

B. *debba* type (★) is Modern Aramaic form with the ending *-a* of historically the emphatic state: *debba* (Juish Neo-Aramaic) in Koy Sanjaq (Kurdistan, Iraq) and Hertevin in Turkey.

C. *dəbi* type (▲) is Ethiopic form: *dəbi* (Tigrinya ድቢ), *dəbə* (Amharic ድብ).

D. *dov* (✖) is Hebrew form (< **dob* דֹּב.)

E. *ors* is Maltese Ar. form (< *orso* Italian.).

(NAGATO Youichi)

A. ● *dubb, dibb, dəbb, ḍəbb*
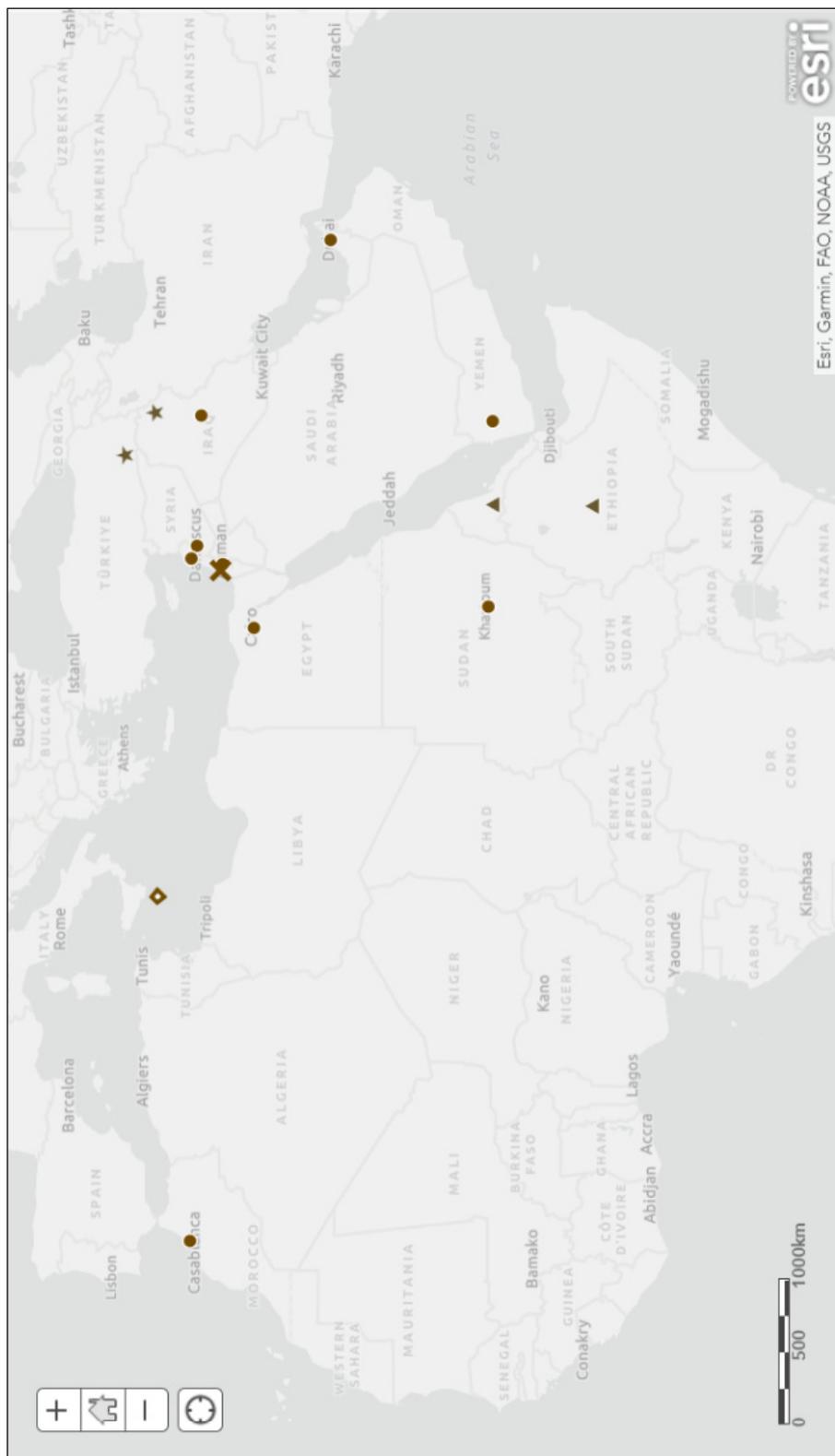B. ★ *debba*
C. ▲ *dəbi, debə*

D. ✖ *dov*
E. ◇ *ors*

Figure 6.18.1: 'Bear' in Semitic.

References

Abaev, V. I. [Абаев, В. И.] (1958)*Историко-Этимологический словарь осетинского языка* т. I a-k'. Москва-Ленинград: Академия наук СССР.

Akitani, Hiroyuki 秋谷裕幸, Weihui Wang 汪维辉, and Masaki Nohara 野原将挥 (2022) Shuo {gou} 说 ｛狗｝. In *Iwata Ray Jiaoshou rongxiu jinian lunwenji* bianjizu 《岩田礼教授荣休纪念论文集》编辑组 (ed) *Iwata Ray Jiaoshou rongxiu jinian lunwenji (shangce)* 《岩田礼教授荣休纪念论文集 上册》, 264–280. Tokyo: Geolinguistic Society of Japan. doi: https://doi.org/10.5281/zenodo.6342364

Alves, Mark J. (2015) Etyma for 'Chicken', 'Duck', & 'Goose' among language phyla in China & Southeast Asia. *Journal of the Southeast Asian Linguistics Society* 8: 39–55. URI: http://hdl.handle.net/1885/16086

Asai, Tōru (1974) Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Cultures, Hokkaido University* 8: 45–13.

Bastin, Yvonne, André Coupez, Evariste Mumba, and Thilo C. Schadeberg (eds.) (2002) Bantu lexical reconstructions 3 / Reconstructions lexicales bantoues 3. URI: http://linguistics.africamuseum.be/BLR3.html

Baxter, William H. and Laurent Sagart (2014) *Old Chinese: A new reconstruction.* Oxford: Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780199945375.001.0001

Behnstedt, Peter and Manfred Woidich (2011) *Wortatlas der arabischen Dialekte. Band I: Mensch, Natur, Fauna und Flora.* Leiden and Boston: Brill. doi: https://doi.org/10.1163/ej.9789004186644.i-665

Bender, M. Lionel (1981) Some Nilo-Saharan Isoglosses. In: Thilo C. Schadeberg and M. Lionel Bender (eds.) *Nilo-Saharan proceedings*, 253–267. Dordrecht: Foris. doi: https://doi.org/10.1515/9783110883466-019

Bennette, Patric R. and Jan P. Sterk (1977) South Central Niger-Congo: a reclassification. *Studies in African Linguistics* 8(3): 241–273. doi: https://doi.org/10.32473/sal.v8i3

Blažek, Václav (2008) Lexicostatistical comparison of Omotic languages. In: John D. Bengtson (ed.) *In Hot Pursuit of Language in Prehistory Essays in the four fields of anthropology*, 57–148. Amsterdam: John Benjamins. doi: https://doi.org/10.1075/z.145.12bla

Blench, Roger (2007) Lexical avoidance taboos and the reconstruction of names for large animals in Niger-Congo, an African language phylum. In Dounias, Edmond, Elisabeth Motte-Florac, and Margaret Dunham (eds.) *Le symbolisme des animaux: L'animal, clef de voûte de la relation entre l'homme et la nature?*, 545–555. Paris: IRD Editions. URI: https://www.documentation.ird.fr/hor/fdi:010041935

Blench, Roger M. (2008a) The problem of pan-African roots. In: John D. Bengtson (ed.) *In Hot pursuit of language in prehistory essays in the four fields of anthropology*, 189–209. Amsterdam: John Benjamins. doi: https://doi.org/10.1075/z.145.16ble

Blench, Roger (2008b) Omotic livestock terminology and its implication for the history of Afroasiatic. In: Gábor Takács (ed.) *Semito-Hamitic Festschrift for A. B. Dolgopolsky and H. Jungraithmayr*, 63–78. Berlin: Dietrich Reimer.

Blust, Robert A. (1980) Austronesian etymologies I. *Oceanic Linguistics* 19: 1–181. URI: https://www.jstor.org/stable/20172307

Blust, Robert A. (1999) Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elisabeth Zeitoun and Paul Jen-Kuei Li (eds) *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Taipei: Academia Sinica.

Bouquiaux, Luc, Jean-Marie Kobozo, Marcel Diki-Kidiri, Jacqueline Vallet et Anne Behaghel (1978) *Dictionnaire sango-français et lexique français-sango*. Paris: Société des Etudes Linguistiques et Anthropologiques de France.

Boyeldieu, Pascal (2000a) *Identité tonale et filiation des langues sara-bongo-baguiriennes (Afrique Centrale)*. Köln: Rüdiger Köppe.

Burrow, T. and M. B. Emeneau (1984) *A Dravidian etymological dictionary*, 2nd ed. Oxford: Clarendon Press.

Cao, Zhiyun 曹志耘 (2008) *Hanyu fangyan dituji* 汉语方言地图集 [*Linguistic atlas of Chinese dialects*]. Beijing: Shangwu yinshuguan 商务印书馆.

Chiri, Mashiho 知里真志保 (1973 [1942]) Ainugohō kenkyū: Karafuto hōgen wo chūshin toshite アイヌ語法研究：樺太方言を中心として [A study of Ainu grammar: With special reference to the Sakhalin dialects]. Reprinted in: *Chiri Mashiho chosakushū* 知里真志保著作集 [*Collected works of Chiri Mashiho*], 3: 455–586. Tokyo: Heibonsha 平凡社.

Chiri, Mashiho 知里真志保 (2001 [1952]) Matagi no yamakotoba to Ainugo マタギの山言葉とアイヌ語 [Japanese mountain language used by Matagi, and Ainu language]. Reprinted in: *Ainugo kō* アイヌ語考 [*The papers on the Ainu language*], 3: 1–3. Tokyo: Yumani Shobō ゆまに書房.

Chiri, Mashiho 知里真志保 (1976 [1962]) *Bunrui Ainugo jiten: Dōbutsu-hen* 分類アイヌ語辞典：動物篇 [*Classified Ainu dictionary*], 3. Reprinted in: *Chiri Mashiho chosakushū* 知里真志保著作集 [*Collected works of Chiri Mashiho*], 1. Tokyo: Heibonsha 平凡社.

Clauson, Gerard (1972) *An etymological dictionary of pre-thirteenth century Turkish*. Oxford: Oxford University Press.

Diffloth, Gérard (1980) The Wa languages. *Linguistics of the Tibeto-Burman Area* 5(2): 1–182. doi: https://doi.org/10.15144/LTBA-5.2.1

Diffloth, Gérard (1984) *The Dvaravati Old Mon languages and Nyah Kur*. (Monic Language Studies). Bangkok: Chulalongkorn University Printing House.

Diffloth, Gérard and N. Zide (1992) Austro-Asiatic languages. In W. Bright (ed) *International Encyclopedia of Linguistics*, Vol. I: 137–142. New York: Oxford University Press.

Dimmendaal, Gerrit J. (2019) Reconstructing Katloid and deconstructing Kordofanian. In: Gertrud Schneider-Blum, Birgit Hellwig and Gerrit J. Dimmendaal (eds.) *Nuba mountain language studies: New insights*, 383–415. Köln: Rüdiger Köppe.

Dimmendaal, Gerrit J., Colleen Ahland, Angelika Jakobi and Constance Kutsch Lojenga (2019) Linguistic features and typologies in languages commonly referred to as 'Nilo-Saharan'. In: H. Ekkehard Wolff (ed.) *The Cambridge handbook of African linguistics*, 326–381. Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/9781108283991.011

Dimmendaal, Gerrit J. and Anne Storch (2016) Niger-Congo: A brief state of the art. In *Oxford handbook topics in linguistics*, online edition. Oxford Academic. doi: https://doi.org/10.1093/oxfordhb/9780199935345.013.3

van Driem, George (2015) Tibeto-Burman. In William S-Y. Wang and Chaofen Sun (eds) *The Oxford handbook of Chinese linguistics*, 135–148. Oxford: Oxford University Press. doi: https://doi.org/10.1093/oxfordhb/9780199856336.013.0007

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.) (2021) *Ethnologue: Languages of the World*. Twenty-fourth edition. http://www.ethnologue.com [accessed on 27 April 2021]

Ebihara, Shiho, Kazue Iwasa, Keita Kurabe, Satoko Shirai, and Hiroyuki Suzuki (2022) 'Horse' in Tibeto-Burman. [In this volume]

Efimov, V. A. [Ефимов, В. А.] (1986) *Язык Ормури: В синхронном и историческом освещении*. Москва: Наука.

Ehret, Christopher (2001) *A historical-comparative reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe.

Emeneau, Murray Barnson (1948) Taboos on animal names. *Language* 24(1): 56–63. doi: https://doi.org/10.2307/410287

Emeneau, Murray Barnson (1997) Brahui etymologies and phonetic developments: New items. *Bulletin of the School of Oriental and African Studies* 60(3): 440–447. doi: https://doi.org/10.1017/S0041977X00032481

Ferlus, Michel (2007) *Lexique de racines Proto Viet-Muong (Proto Vietic Lexicon)*. Unpublished manuscript. (with personal permission of use) http://sealang.net/monkhmer/database/

Fortescue, Michael (2005) *Comparative Chukotko-Kamchatkan dictionary*. Berlin/New York: Mouton de Gruyter. doi: https://doi.org/10.1515/9783110925388

GAJ = Kokuritsu kokugo kenkyūjo 国立国語研究所編 [National Institute for Japanese Language and Linguistics (eds.)] (1989-2006) *Hōgen bumpō zenkoku chizu* 方言文法全国地図 [*Grammar Atlas of Japanese dialects*]: Zaimushō insatsukyoku 財務省印刷局.

Greenberg, Joseph H. (1963) *The languages of Africa*. Bloomington: Indiana University Press.

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti, and Mark Pagel (2015) Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 112(43): 13296–13301. doi: https://doi.org/10.1073/pnas.1503793112

Güldemann, Tom (2014) 'Khoisan' linguistic classification today. In: Tom Güldemann and Anne-Maria Fehn (eds.) *Beyond 'Koisan': Historical relations in the Kalahari Basin*, 1–40. Amsterdam: John Benjamins. doi: https://doi.org/10.1075/cilt.330.01gul

Guo, Xiliang 郭锡良 (2010) *Hanzi guyin shouce (zengdingben)* 汉字古音手册（增订本）[*A handbook of ancient sounds of Chinese characters (enlarged and revised version)*]. Beijing: Shangwu yinshuguan 商务印书馆.

Guthrie, Malcolm (1967–71) *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages*. Vols. 1–4. London: Gregg Press.

Hallowell, A. Irving (1926) Bear ceremonialism in the northern hemisphere. *American Anthropologist New series* 28(1): 1–175. doi: https://doi.org/10.1525/aa.1926.28.1.02a00020

Hammarström, Harald (2019) An inventory of Bantu languages. In Van de Velde, Mark, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.) *The Bantu languages.* Second edition, 17–78. London: Routledge. doi: https://doi.org/10.4324/9781315755946

Hammarström, Harald, Robert Forkel, Martin Haspelmath and Sebastian Bank (eds.) (2020) Glottolog 4.3. http://glottolog.org [accessed on 27 April 2021].

Hattori, Shirō and Mashiho Chiri 服部四郎、知里真志保 (1960) Ainugo shohōgen no kisogoi tōkeigakuteki kenkyū アイヌ語諸方言の基礎語彙統計学的研究 [*A lexicostatistic study of the Ainu dialects*]. *The Japanese Journal of Ethnology* 民族學研究 24(4): 31–66. doi: https://doi.org/10.14890/minkennewseries.24.4_307

Hayashi, Kingo 林欽吾 (1973 [1940]) Shikotan-tō no Ainu-zoku 色丹島のアイヌ族 [The Ainu people on Shikotan Island]. Reprinted in Nihon Jōmin Seikatsu Kenkyūjo (ed.), *Nihon jōmin seikatsu shiryō sōsho* 日本常民生活資料叢書 [*Series of materials on the lives of the people of Japan*] 7: 552–505. Tokyo: Sanichi Shobō 三一書房.

Headley, Robert K. (1985) Proto-Pearic and the classification of Pearic. In Suriya Ratanakult et al. (eds.) *Southeast Asian Linguistic Studies Presented to André G. Haudricourt*, 428–478. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University.

Heine, Bernd (1975) Notes on the Yaaku Language (Kenya). *Afrika und Übersee* 58: 27–61, 119–138. URI: http://hdl.handle.net/2307/5364

Hetzron, Robert (1974) La division des languages sémitiques. In André Caquot et David Cohen (eds.) *Actes du premier congrès international de linguistique sémitique et chamito-sémitique Paris 16–19 juillet 1968*, 181–194. Mouton. doi: https://doi.org/10.1515/9783110879872-019

Hetzron, Robert (1976) Two principles of genetic reconstruction. *Lingua* 38: 89–108. doi: https://doi.org/10.1016/0024-3841(76)90074-7

Hou, Jingyi 侯精一 (2002) *Xiandaihanyu fangyanxue gailun* 现代汉语方言概论. Shanghai 上海: Shanghai jiaoyu chubanshe 上海教育出版社.

Igarashi, Yosuke 五十嵐陽介 (2018) Bunkigakuteki shuhō ni motozuita Nihongo, Ryūkyugo shohōgen no keitōbunrui no kokoromi 分岐学的手法に基づいた日本語・琉球語諸方言の系統分類の試み [An attempt to classify Japanese and Ryukyuan dialects based on cladistic methods]. In Yuka Hayashi 林由華, Tomohide Kinuhata 衣畑智秀, and Nobuko Kibe 木部暢子 (eds.) *Fīrudo to bunken kara miru Nichiryūshogo no keitō to rekishi* フィールドと文献から見る日琉諸語の系統と歴史 [Phylogeny and History of Japanese and Ryukyuan Languages from Fields and Documents], 17–51. Kaitakusha 開拓社.

Ikegami, Jiro 池上二良 (1989) Tungus-shogo ツングース諸語. In *Gengogaku daijiten* 言語学大辞典, Vol. 2. Tokyo: Sanseido 三省堂.

Irving-Pease, Evan K., Hannah Ryan, Alexandra Jamieson, Evangelos A. Dimopoulos, Greger Larson & Laurent A. F. Frantz (2018) Paleogenomics of animal domestication. In C. Lindqvist and O. Rajora (eds.). *Paleogenomics*, 225–272. Cham: Springer. doi: https://doi.org/10.1007/13836_2018_55

Janhunen, Juha (2003) Tracing the bear myth in Northeast Asia. *Acta Slavica Iaponica* 20: 1–24. URI: http://hdl.handle.net/2115/39417

Johanson, Lars and Éva Ágnes Csató (eds.) (1998) *The Turkic languages.* London and New York: Routledge.

Karlgren, Bernhard (1957) *Grammata serica recensa*. Stockholm: The Museum of Far Eastern Antiquities. [Chinese edition: *Hanwendian (Xiudingben)* 《汉文典（修订本）》 translated by Wuyun Pan 潘悟云, Jianqiao Yang 杨剑桥, Zhongye Chen 陈重业, and Hongming Zhang 张洪明, 1997. Shanghai: Shanghai Cishu Chubanshe.]

Kindaichi, Kyōsuke 金田一京助 (1993 [1937]) Sankan no Ainugo 山間のアイヌ語 [The Ainu language in the mountain]. *Sangaku* 山岳 [*Mountains*], 33(1). Reprinted in *Kindaichi Kyōsuke zenshū* 金田一京助全集 [*The complete works of Kyosuke Kindaichi*], 6: 255–260. Tokyo: Sanseido 三省堂.

Klimov, G. A. and M. Sh. Khalilov [Климов, Г. А. и М. Ш. Халилов] (2003) *Словарь кавказских языков: Сопоставление основой лексики*. Москва: Издательская фирма <Восточная литература>.

Kobayashi, Masato and Bablu Tirkey (2017) *The Kurux language: Grammar, texts and lexicon*. Leiden: Brill. doi: https://doi.org/10.1163/9789004347663

Kodama, Nozomi (2021) Animal Vocabulary in Dravidian. The 3rd meeitng of Studies in Asian and African Geolinguistics, ILCAA Joint Research Project, Sat 4 Sep–Sun 5 Sep 2021 (online).

Krishnamurti, Bhadriraju (2003) *The Dravidian languages*. Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511486876

Kurebito, Megumi (ed.), Tokusu Kurebito, Megumi Kurebito, Yukari Nagayama, Chikako Ono, Mitsuhiro Yazu (2001) *Comparative basic vocabulary of the Chukchee-Kamchatkan language family 1*. Osaka: Osaka Gakuin University.

LAJ = Kokuritsu kokugo kenkyūjo 国立国語研究所編 [National Institute for Japanese Language and Linguistics (eds.)] (1966–74) *Nihon gengo chizu* 日本言語地図 [Linguistic Atlas of Japan]: Ōkurashō insatsukyoku 大蔵省印刷局.

Leslau, Wolf (1979) *Etymological dictionary of Gurage (Ethiopic)*. 3 vols. Wiesbaden: Otto Harrassowitz.

Li, Fang-kuei (1977) *A handbook of comparative Tai*. Honolulu: The University Press of Hawai'i.

Li, Paul Jen-kuei (2008) The great diversity of Formosan languages. *Language and Linguistics* 9(3): 523–546.

Liang, Min 梁敏 and Junru Zhang 张均如 (1996) *Dongtai yuzu gailun* 侗台语族概论. Beijing: Zhongguo Shehui Kexue Chubanshe 中国社会科学出版社.

Librado, Pablo, Naveed Khan, Antoine Fages, et al. (2021) The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598(7882): 634–640. doi: https://doi.org/10.1038/s41586-021-04018-9

Lipiński, Edward (2001) *Semitic languages: Outline of a comparative grammar* (2nd edition). Leuven: Peeters.

Luangthongkum, Theraphan. 2019. A View on Proto-Karen phonology and lexicon. *Journal of the Southeast Asian Linguistics Society* 12(1): i–lii. doi: http://hdl.handle.net/10524/52441

MacMichael, H. A. (1918) Nubian elements in Darfur. *Sudan Notes and Records* 1: 30–48. URI: https://www.jstor.org/stable/41715920

Masica, Colin P. (1991) *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Matisoff, James A. (2003) *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley/Los Angeles/London: University of California Press. URI: https://escholarship.org/uc/item/19d79619

Matisoff, James A. (2010) Toward a Eurasian bestiary: (I) OTTER in Tibeto-Burman and Mon-Khmer (II) JACKAL in Sino-Tibetan and Indo-European. *Bulletin of the National Museum of Ethnology* 34(3): 575–591. URI: http://hdl.handle.net/10502/4142

Matisoff, James A. (2015) *The Sino-Tibetan etymological dictionary and thesaurus.* Berkeley: The Regents of the University of California.

Matossian, Mary Kilbourne (1997) *Shaping world history: Breakthroughs in ecology, technology, science, and politics.* Armonk: M.E. Sharpe.

Matumo, Z. I. (1993) *Setswana English Setswana dictionary.* 4th edition. Gaborone: Macmillan Botswana.

Meeussen, Achille Emile (1969) Bantu lexical reconstructions. *Africana Linguistica* 3: 80–122.

Morgenstierne, Georg (1929) *Indo-Iranian frontier languages: vol. I Parachi and Ormuri.* Oslo: Instituttet for Sammenlignende Kulturforskning.

Munda, Ram Dayal (1968) Proto-Kherwarian phonology. M.A. Thesis. University of Chicago.

Nakagawa, Hiroshi 中川裕 (1989) Nihongo to Ainugo tono sōjigoi 日本語とアイヌ語との相似語彙 [Similar vocabulary in Japanese and Ainu]. *Kikan Yamataikoku* 季刊邪馬台国, 38: 190–205.

Norquest, Peter K. (2016) *A phonological reconstruction of Proto-Hlai.* Leiden: Brill. doi: https://doi.org/10.1163/9789004300521

O'Fahey, Rex S. (1980) *State and society in Dār Fūr.* London: C. Hurst & Co.

Ostapirat, Weera (2000) Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23.1: 1–251. doi: https://doi.org/10.15144/LTBA-23.1.1

Ostapirat, Weera (2016) Issues in the reconstruction and affiliation of Proto-Miao-Yao. *Language and Linguistics* 17(1): 133–145. doi: https://doi.org/10.1177/1606822X15614522

Ōtsuki, Fumihiko 大槻文彦 (1932) *Daigenkai* 大言海 [Daigenkai] vol. 1. Fuzanbō 冨山房.

Pellard, Thomas (2015) The linguistic archeology of the Ryukyu Islands. In: Patrick Heinrich, Shinsho Miyara, and Michinori Shimoji (eds.) *Handbook of the Ryukyuan languages: History, structure, and use*, 13–37. Berlin: De Gruyter Mouton. doi: https://doi.org/10.1515/9781614511151.13

Perri, Angela R., Tatiana R. Feuerborn, Laurent A. F. Frantz, Greger Larson, Ripan S. Malhi, David J. Meltzer, Kelsey E. Witt (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences* 118(6): e2010083118. doi: https://doi.org/10.1073/pnas.2010083118

Petrov, A. A. (1989). The bear taboo in Even language and folklore. *Études/Inuit/Studies* 13(1): 131–133. URI: http://www.jstor.org/stable/42869655

Qian, Zengyi 钱曾怡 et al. (2010) *Hanyu guanhua fangyan yanjiu* 汉语官话方言研究. Taiyuan 太原: Qilu shushe 齐鲁书社.

Quint, Nicolas (2013) Integration of borrowed nouns in Koalib, a noun class language. In: Schadeberg, Thilo C. and Roger M. Blench (eds.) *Nuba mountain language studies*, 115–134. Köln: Rüdiger Köppe.

Rybatzki, Volker (2003) Intra-Mongolic taxonomy. In Juha Janhunen (ed.) *The Mongolic languages*, 364–390. London: Routledge. doi: https://doi.org/10.4324/9780203987919

Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill & Johann-Mattis List (2019) Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences* 116(21): 10317–10322. doi: https://doi.org/10.1073/pnas.1817972116

Savaşçı, Özgür (2008) *Visuelles Wörterbuch Türkisch-Deutsch*. München: Dorling Kindersley Verlag GmbH.

Schuessler, Axel (2007) *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawaiʻi Press.

Shorto, Harry L. (2006) *A Mon-Khmer comparative dictionary, Pacific Linguistics*. Canberra: Research School of Pacific and Asian Studies, Australian National University. doi: https://doi.org/10.15144/PL-579

Sidwell, Paul (2005) *The Katuic languages: classification, reconstruction and comparative lexicon*. München: Lincom Europa.

Sidwell, Paul (2010) *Preliminary Notes on Proto Palaungic*. Unpublished manuscript. http://sealang.net/monkhmer/database/

Sidwell, Paul (2011) *Proto Bahnaric*. Unpublished manuscript. http://sealang.net/monkhmer/database/

Sidwell, Paul (2012) *Proto Khasic*. Unpublished manuscript. http://sealang.net/monkhmer/database/

Sidwell, Paul (2013) *Proto Khmuic*. Unpublished manuscript. http://sealang.net/monkhmer/database/

Sidwell, Paul (2014) Austroasiatic classification. In Mathias Jenny and Paul Sidwell (eds) *The Handbook of Austroasiatic languages*, 144–220. Leiden: Brill. doi: https://doi.org/10.1163/9789004283572_004

STEDT = The Sino-Tibetan Etymological Dictionary and Thesaurus. Database. Online: http://stedt.berkeley.edu/search/

Takeuchi, Tsuguhito (2021) History of the Tibetan language. In Yasuhiko Nagano and Takumi Ikeda (eds) *Grammatical phenomena of Sino-Tibetan languages 4: Link languages and archetypes in Tibeto-Burman*, 303–323. Kyoto: Research Institute for Humanities, Kyoto Univeristy. URI: http://hdl.handle.net/2433/263983

Tamura, Suzuko (2000) *The Ainu language*. Tokyo: Sanseido.

Thurgood, Graham (2017) Sino-Tibetan: Genetic and areal subgrouping. In Graham Thurgood and Randy J. LaPolla (eds) *The Sino-Tibetan languages*, second edition, 3–39. London and New York: Routledge. doi: https://doi.org/10.4324/9781315399508

Tōdō, Akiyasu (ed.) 藤堂明保 (1978) *Gakken Kanwa Daijiten* 学研漢和大辞典 [Gakken Chinese-Japanese Character Great Dictionary]. Gakushū Kenkyūsha 学習研究社.

Tournadre, Nicolas and Hiroyuki Suzuki (2022) *The Tibetic languages: An introduction to the family of languages derived from Old Tibetan* (with the collaboration of Xavier Becker and Alain Brucelle for the cartography). Villejuif: LACITO Publications (CNRS).

Turner, Ralph Lilley (1985) *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

Utsumi, Atsuko (2022) 'Horse' in Austronesian. [In this volume]

VanBik, Kenneth (2009) *Proto-Kuki-Chin: A reconstructed ancestor of the Kuki-Chin languages.* Berkeley: University of California.

Watters, John R. (2018) East Benue-Congo. In John R. Watters (ed.) *East Benue-Congo: Nouns, pronouns, and verbs.* Berlin: Language Science Press. doi: https://doi.org/10.5281/zenodo.1314337

Williamson, Kay (1989) Niger-Congo overview. In Bendor-Samuel, John (ed.) *The Niger-Congo languages.* Lanham: University Press of America.

Williamson, Kay (2000) Did chickens go west? In: Roger M. Blench and Kevin C. MacDonald (eds.) *The origins and development of African livestock: Archaeology, genetics, linguistics and ethnography*, 368–448. London: Routledge. doi: https://doi.org/10.4324/9780203984239

Williamson, Kay and Roger Blench (2000) Niger-Congo. In Bernd Heine and Derek Nurse (eds.) *African Languages: An introduction*, 11–42. Cambridge: Cambridge University Press.

Wurm, S. A., Rong Li, Theo Baumann, Mei W. Lee, Australian Academy of Science, and Zhongguo Shehuikexueyuan (1987) *Language atlas of China.* Hong Kong: Longman.

Yonezawa, Takahiro and Tsuyoshi Sasaki (2016) The way the chicken came is etched in its genes 遺伝子に刻まれた鶏の来た道. *Society for Livestock Studies* 15: 4–13.

Zhan, Bohui 詹伯慧 et al. (2017) *Hanyu fangyanxue dacidian* 汉语方言学大词典. Guangzhou 广州: Guangdong jiaoyu chubanshe 广东教育出版社.

Zhang, Hanzhi, Ting Ji, Mark Pagel & Ruth Mace (2020) Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Scientific Reports* 10(20792): 1–8. doi: https://doi.org/10.1038/s41598-020-77404-4

Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin (2019) Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569(7754): 112–115. doi: https://doi.org/10.1038/s41586-019-1153-z

❦

Linguistic atlas of Asia and Africa
Cumulative index of chapters

❧