

Chapter 11

Typological profile of the Transeurasian languages from a quantitative perspective

Nataliia Hübler

Abstract

This chapter provides an overview of the typological features of the Transeurasian (Turkic, Mongolic, Tungusic, Japonic, Koreanic) languages, including brief descriptions of the phonology and morphosyntax of these languages. By applying phylogenetic comparative methods, I delimit a set of structural features with a high phylogenetic signal. These features can be assumed to be genealogically stable. I compare the trees achieved by Bayesian tree-sampling based on all 226 features and on the 97 structural features with a high phylogenetic signal and come to the conclusion that the data set with presumably stable structural features does not provide a tree that is compatible with the language history assumed by classical historical linguists. Neither full nor reduced feature set provides a reliable internal classification of the Turkic, Mongolic, Tungusic and Japonic language families.

Keywords: Transeurasian languages, typological features, phylogenetic signal, Bayesian tree-sampling

11.1 Introduction

It is common knowledge that most languages of Northeast Asia exhibit similarities in their structure, among them verb-final word order, strong head-marking, agglutinative suffixing morphology, lack of gender distinctions. The main discussion concerns the

question whether all these similarities can be attributed to areal dispersal or whether some are residue of inheritance from a proto-language.

Although there is still no full consensus on the status of the Transeurasian unity as a Sprachbund or a language family, the genealogical relatedness of the Transeurasian languages is gradually gaining acceptance in the literature. See Robbeets (this volume: Chapter 10) for the view that the Transeurasian languages are related and Vajda (this volume: Chapter 41) for the view that Transeurasian languages represent an area of diffusion. Moreover, scholars that agree on the relatedness of Transeurasian languages suggest different topologies for the Transeurasian macrofamily (see Robbeets, this volume: Chapter 3).

Classical comparative linguists rely on basic vocabulary and cognate grammatical morphemes when postulating language relationships. There is basic vocabulary (Robbeets, this volume: Chapter 36) and cognate grammatical morphemes (Robbeets, this volume: Chapter 30) in support of Transeurasian genealogical affiliation. Among the reasons why historical linguists do not wish to take abstract grammatical features into account are the following. First, structural features are more prone to borrowing than basic vocabulary or form-function matches in morphology. Second, the number of states structural features take (namely two: absent or present) facilitates convergent evolution (Heggarty 2006: 187, 193; Greenhill 2017: 5). Third, the possible functional dependencies between features may lead to non-informative branch lengths (Heggarty 2006: 186). Fourth, a high rate of change leads to frequent switches between the states and the impossibility to predict the states for the latest common ancestors (Greenhill 2017). The answers to the questions such as “Do structural features change faster than basic vocabulary?” and “How easily are structural features borrowed?” differ drastically. Some scholars state that structural features contain a deeper phylogenetic

signal than basic vocabulary (Dunn et al. 2005), others add that it is impossible to disentangle genealogical signal from the one coming from ancient contact events (Wichmann and Holman 2009: 221) or that a group of features cannot define a genealogical unit (Reesink et al. 2009: 8).

In this chapter, I will not use structural evidence to establish language relatedness, but examine whether a set of stable structural features can replicate a topology of the individual Transeurasian families based on basic vocabulary and phonological correspondences and compare the performance of structural features in providing tree structures that represent true language relationships to that of basic vocabulary (as in Savelyev, this volume: Chapter 9; Whaley and Oskolskaya, this volume: Chapter 6).

Robbeets (this volume: Chapter 10) delimits a core of structural features that are shared between the Transeurasian languages and seem to be more easily explainable by inheritance than by borrowing. My approach is different from that of Robbeets, as I apply Bayesian inference to reach the topology of the Transeurasian languages and calculate the phylogenetic signal in the structural features along this topology.

Bayesian inference and phylogenetic comparative methods have not yet been applied to the structural features of the Transeurasian languages to find historical signal in them and build the topology of the Transeurasian languages. Among the studies that applied Bayesian methods to structural data cross-linguistically, we find Dunn et al. (2008), Dediu and Levinson (2012), Reesink et al. (2009). Wichmann (2015) and Greenhill et al. (2017) concentrate on the rate of change of structural features.

The chapter is structured as follows. In Section 11.2 I present the language sample used for the typological description of the languages in question and the phylogenetic analysis in the following sections. Section 11.3 provides an overview of the typological similarities and differences between 38 Transeurasian languages. In

Section 11.4 I apply phylogenetic comparative methods to delimit a set of structural features with a high phylogenetic signal and compare the topology of the Transeurasian languages based on all the features to the one based on the delimited set of structural features with a high phylogenetic signal. I summarize the findings in Section 11.5.

11.2 Language sample

The language sample is heterogeneous in terms of geography (see Figure 1 for the distribution of the languages) and genealogical affiliation. The sample covers 13 Turkic languages, 10 Tungusic, 5 Mongolic, 9 Japonic languages and Korean.

< Insert Figure 1 here >

Figure 11.1 Geographical distribution of the languages of the samples (EvB=Even (Beryozovka dialect); EvD=Even (Dogdo-chebogalahskiy dialect); Evk=Evenki; Nan=Nanai; Neg=Negidal; Oroc=Oroch; Orok=Orok; Udi=Udihe; Ulch=Ulch; Soln=Solon; Azer=Azerbaijani; Bash=Bashkir; Chu=Chuvash; Crim=Crimean Tatar; Gag=Gagauz; Khak=Khakas; Khal=Khalaj; Shor=Shor; Trk=Turkish; Tuv=Tuvan; Yak=Yakut; Tat=Tatar; Tuk=Turkmen; Jap=Japanese; Ogm=Ogami; Shu=Shuri; Tar=Tarama; Hat=Hateruma; Ike=Ikema; Oki=Okinoerabu; Yon=Yonaguni; Yuw=Yuwan; Bon=Bao'an; Halh=Khalkha; Mang=Mangghuer; Kalm=Kalmyk; Bur=Buriat; Kor=Korean)

Table 11.1 Language sample (classification according to Johanson, this volume: Chapter 8 (Turkic); Whaley and Oskolskaya, this volume: Chapter 6 (Tungusic); Heinrich et al. 2015 (Japonic))

Turkic	Bulgharic: Chuvash
	Oghuzic: Turkmen, Azerbaijani, Gagauz, Turkish, Khalaj ¹
	Siberian: Yakut, Tuvan, Khakas, Shor
	Kipchak: Crimean Tatar, Tatar, Bashkir
Mongolic	Khalkha, Kalmyk, Buriat, Mangghuer, Bao'an
Tungusic	Northern: Even (Beryozovka dialect), Even (Dogdo-chebogalahskiy dialect), Evenki, Solon, Negidal
	Southern: Udihe, Oroch, Nanai, Ulch, Orok
Japonic	Northern Ryukyuan: Shuri, Yuwan, Okinoerabu
	Southern Ryukyuan: Ogami, Yonaguni, Hateruma, Tarama, Ikema
	Mainland Japanese: Japanese
Koreanic	Korean

11.3 Typological overview of the Transeurasian languages

In the description of the typological type of the Transeurasian languages that follows I will refer to the doculects of the sample². Any generalizations about the overall presence or absence of a feature will only take into account the doculects mentioned in Table 11.1. Cases, where the feature value is debatable or unknown, will also be excluded from generalisations.

11.3.1 Phonology

11.3.1.1 Vowels

Japonic (apart from Yonaguni), Tungusic, Mongolic languages as Buriat, Kalmyk, Khalkha, Siberian Turkic languages and Khalaj exhibit the vowel length distinction,

e.g., Buriat (Mongolic, Poppe 1960a: 6): *tohon* ‘fat, butter’ - *to:hon* ‘dust’, *dara* ‘press (imperative)’ - *da:ra* ‘freeze (imperative)’

The most common type of vowel harmony synchronically is palatal harmony, which is present in all Turkic, some Mongolic, some Tungusic languages and Korean (1) (for a detailed discussion on the vowel harmony and beyond, see Joseph et al., this volume: Chapter 29, for the distribution of other types of vowel harmony, see the online supplementary materials for this chapter).

(1) Korean (Koreanic, Sohn 1999:181)

- a. *cwuk-essta*
die-PST
‘died’
- b. *nol-assta*
play-PST
‘played’

Tungusic and some Mongolic languages (for a broader discussion, see Janhunen (2012a: 78–79 and Svantesson, this volume: Chapter 21 for Khalkha; Oskolskaya, this volume: Chapter 19; Robbeets, this volume: Chapter 10) also exhibit tongue root vowel harmony (2).

(2) Even (Tungusic, Kim 2011: 40)

- a. *nɔŋan-do*
3SG-DAT
‘to him/her’

- b. *min-du*
1SG-DAT
'to me'

11.3.1.2 Positional constraints

Initial velar nasals are not permitted word-initially in Transeurasian languages, apart from most Tungusic languages (except for Solon) and Bao'an. Initial trill /r/ in native words is restricted to Bao'an and Mangghuer. Initial consonant clusters are only permitted in some Japonic and Mongolic languages, and even if so, then most commonly the second consonant is a glide.

11.3.1.3 Phoneme inventories

Two separate liquid phonemes are present in all Mongolic and Turkic languages as well as in some Tungusic languages. They are absent in Japonic, Korean, Negidal, Orok, Oroch and Udihe.

Typical of Tungusic languages is presence of voicing distinctions in stops, but not in fricatives (apart from Oroch, where there is a voicing distinction in dorsal fricatives). Most Turkic languages have both distinctions, apart from Chuvash (has none) and Yakut (has no voicing distinction in fricatives). Among Mongolic languages, at least Mangghuer is a special case with the voicing distinctions neither in plosives nor in fricatives, whereas most Mongolic languages have this distinction in plosives. Japonic languages mostly exhibit a voicing distinction in stops (apart from Ogami), but only some have it in fricatives. Korean has no voicing distinction in plosives or fricatives.

Transeurasian languages have two laryngeal contrasts for stops: voiced and

voiceless. The only Transeurasian language exhibiting three laryngeal contrasts for stops (voiced, voiceless, aspirated) is Korean.

11.3.2 Agglutination and position of bound morphemes

Transeurasian languages in the sample are languages with agglutinative morphology, with the bound morphology being mostly suffixing.

11.3.3 Noun

In all Mongolic, Turkic languages and Korean nouns can be marked for plural.

Among Tungusic languages, this holds for all Northern Tungusic languages, Nanai and Ulch. In Japonic languages nouns can be marked for plural, but this is mostly restricted to animate nouns. Southern Tungusic languages have a plural marker for animate nouns (apart from Ulch), Nanai has both a productive plural marker and a plural marker for kinship terms.

The markers are typically regular, i.e. the plural form can be predicted from the singular form, with some phonological variation, e.g., plural formation in Yakut is accomplished by means of the suffix *-lar* and its allomorphs *-tar*, *-dar*, *-nar* (3). For lexicalization of plural markers, see Gruntov and Mazo (this volume: Chapter 31).

(3) Yakut, (Turkic, Kharitonov 1982: 191)

at-tar

horse-PL

‘horses’.

Transeurasian languages do not have any marking for any other number than plural, except for Bao'an, which has dual and paucal marking on nouns in addition to plural marking (4).

- (4) Bao'an (Mongolic, Fried 2010: 68)

au=kala silaŋ=da o-tɛo
 man=DU Xining=LOC go-IPFV.OBJ
 'The (two) men are going to Xining.'

The plural marker can have an associative meaning in Japonic, most Turkic languages (5) and Korean (-*tul*).

- (5) Chuvash (Turkic, Krueger 1961: 94):

ivanov-zem
 Ivanov-PL
 'members of the Ivanov's family'

Some Mongolic (Bao'an, Mangghuer), Tungusic languages, Khalaj, Shor, Yakut and Korean (-*ney*) have a special associative plural marker, e.g., compare the associative marker and the plural marker in Even (6a) and (6c).

- (6) Even (Tungusic, Lebedev 1978: 43–44):

a. *ami-ja*
 father-ASSOC
 'father and his relatives'

- b. *orïr*
 deer.SG
 ‘a deer’
- c. *orïl*
 deer.PL
 ‘deer’

Most Transeurasian languages have a pattern of derivation of action (7a), agent (7b) and object (7c) noun from a verb.

(7) Khalkha (Janhunen 2012a: 97–98)

- a. *saa-ly*
 milk-NMLZ
 ‘milking’
- b. *bic-e:c*
 write-NMLZ
 ‘scribe’
- c. *bic-ig*
 write-NMLZ
 ‘script’.

Most Transeurasian languages have morphological core case (S, A, P argument) marking. Japonic languages and Korean mark grammatical relations by clitics. In this study, they are treated as morphological case marking, given their phonological

boundness. Oblique arguments are marked either by a case suffix, by a postposition or by both.

In Transeurasian languages, noun reduplication serves the expression of a collective meaning (8), plurality (9) or distribution (10).

- (8) Kalmyk (Mongolic, Benzing 1985: 143)

ükr~mükr

cow~COLL

‘cows of different kinds’

- (9) Azerbaijani (Turkic, Shiraliev 1971: 43)

dästä~dästä çiçäk

bunch~PL flower

‘bunches and bunches of flowers’

- (10) Korean (Koreanic, Sohn 1994: 386)

cip~cip

house~DISTR

‘every house’

Diminutive derivation (11) is productive across all Transeurasian languages with only a few exceptions and missing information for some languages.

- (11) Shuri (Japonic, Shimoji 2012: 354)

taru:-gwa:

Taruu-DIM

‘a little Taruu’

The languages, where it is present, but is not a productive process, include Chuvash and Khalkha. Augmentative derivation is only found in Northern Tungusic languages, as, e.g., in Negidal (12), and Yonaguni.

(12) Negidal (Tungusic, Tsintsius 1982: 21).

bəje-xa:ja:

human-AUG

‘a huge human’.

11.3.4 Pronoun

Some Tungusic languages and Bao’an (among the Mongolic languages in the sample) exhibit an inclusive/exclusive distinction in the first person plural, e.g., Udihe (Tungusic, Girfanova 2002: 18): *minti* 1PL.INCL, *bu* 1PL.EXCL. This distinction is present in Buriat and Kalmyk diachronically only. There is no gender distinction in personal pronouns in all Transeurasian languages, apart from Japanese (Japonic, Hinds 1986: 239): *kare* 3SG.M, *kanojo* 3SG.F, which entered the Japanese language relatively late, in Middle Japanese, and increased in frequency after the 16th century under influence of Dutch.

Possessive pronouns not formed by a regular process are not well spread throughout Transeurasian languages. In most Tungusic languages oblique pronominal stems fulfill their function. In Mongolic languages they are usually formed from a stem, different both from nominative and oblique pronominal stem, and a genitive

marker: Mangghuer (Mongolic, Slater 2003a: 83): *mu=ni* 1SG=GEN, *namei=du* 1SG=DAT, *bi* 1SG.NOM. There is no synchronically detectable pattern in the spread of possessive pronouns across the Turkic languages.

Special logophoric pronouns are not common in Transeurasian languages. For Ogami Pellard (2009) reports the existence of a reflexive pronoun that is used to indicate that the subject of the subordinate clause is the same as the subject of the first clause. In the example in (13) the reflexive pronoun *naa* is used to indicate that the 3rd person reporting the speech refers to a group of people including himself, whereas the reflexive pronoun *tuu* cannot be used logophorically. In Bao'an (Fried 2010: 121), logophoric pronouns are not obligatory.

(13) Ogami (Japonic, Pellard 2009: 122)

kanu psta=a naa-ta ik-a-teen=ti
 DIST nobody=TOP REFL-PL go-IRR-ACOM=QUOT
aur-i=u
 say-CVB=IPFV
 'He says, they will not go.'

Most Transeurasian languages possess a phonologically independent reflexive pronoun. Reciprocal pronouns are only rarely mentioned in descriptive works. Tungusic, Mongolic, some Japonic and Turkic (apart from Yakut) languages all form the oblique pronominal stem with a dental nasal, e.g., Negidal (Oskolskaya p.c. 2017): *bi* 1SG.NOM, *min* 1SG.OBL; Buriat (Poppe 1960a: 50): *bi* 1SG.NOM, *mini*: 1SG.GEN; Turkish (Kornfilt 1997: 281): *o* 3SG.NOM, *on-a* 3SG-DAT (see also Schwarz, this volume: Chapter 32). This is not the case in most Japonic languages and

Korean. In northern Ryukyuan dialects, the first person pronoun uses *waa-* as the nominative and genitive base and extended *waN-* in the oblique cases (Robbeets, this volume: Chapter 10).

11.3.5 Demonstrative

Demonstratives in Mongolic and Tungusic languages have a two-way distance contrast. Japonic (Japanese, Shuri), Turkic languages (Chuvash, Shor, Turkish and Yakut) and Korean possess three demonstratives expressing a three-way distance relationship, e.g., Japanese (Hinds 1986: 232): *kono* ‘this’, *sono* ‘that’, *ano* ‘that over there’. Invisibility seems to be an accompanying meaning of the distal demonstrative in some Turkic languages. The only demonstrative with the dedicated function of expressing invisibility is present in Bao’an in the sample: *ənə* ‘this’, *nokə* ‘that’, *thər* ‘that out of sight’ (Fried 2010: 143).

In some Tungusic languages demonstratives agree with the noun in number (14), in Mongolic languages this is only the case in Buriat (see example 16 for Buriat in contrast to 15 for Kalmyk), although it had been a standard agreement in Middle Mongolian (Orlovskaya 1999: 37).

- (14) Evenki (Nedjalkov 1997: 294)

<i>tari-l-va</i>	<i>beje-l-ve</i>
that-PL-ACC.DEF	man-PL-ACC.DEF
‘those people’	

- (15) Kalmyk (Benzing 1985: 133)

<i>ter</i>	<i>ger-müd</i>
------------	----------------

this house-PL

‘these houses’

- (16) Buriat (Poppe 1960a: 110)

te-de *gern-ii:d*

that-PL house-PL

‘those houses’

All Mongolic, some Tungusic and Siberian Turkic languages and Gagauz possess a verb for content interrogation (meaning ‘do what?’). Japanese (Japonic, Hinds 1986: 29) has a compound *do:-si-te* how-do-PTCP ‘why’.

11.3.6 Article

Nouns are not obligatorily modified by definite articles in the whole area. Indefinite articles are optional in some Turkic languages (Khalaj, Khakas, Crimean Tatar, Turkish, Gagauz, Turkmen), Mangghuer, Bao’an and Oroch. Their position varies though: in Turkic languages there are only indefinite prenominal articles (18), Mangghuer (17, indefinite), Bao’an (indefinite) and Udihe (definite) have only postnominal articles.

- (17) Mangghuer (Mongolic, Slater 2003a: 99)

shuguo *beghe ge*

big tree SG:INDEF

‘a big tree’

(18) Khalaj (Turkic, Doerfer 1988: 94)

bi: ki-ni:

one day-ACC

‘on one day’

11.3.7 Adjective

In Korean and most Japonic languages adjectives can receive the same marking as verbs used both predicatively (19a) and attributively (19b), in Turkic languages adjectives in predicative position can receive the same marking as verbs (20).

(19) Japanese (Japonic, Hinds 1986: 345–346)

a. *ano eiga=wa omosiroka-tta*

that movie=TOP interesting-PST

‘That movie was interesting.’

b. *omosiroka-tta eiga*

interesting-PST movie

‘an interesting movie’

(20) Turkish (Turkic, Kornfilt 1997: 83)

termiz-di-m

clean-PST-1SG

‘I was clean.’

Reduplication of adjectives is a common process in Transeurasian languages; mostly it expresses intensification of the quality (21).

- (21) Yakut (Turkic, Kharitonov 1982: 156)

χap~χara

black~INT

‘very black’

Adjectives normally do not agree with nouns in number, with the exception of some adjectives in Buriat (22), Even (at least Dogdo-chebogalahskiy dialect) and Evenki. It was very common in Middle Mongolian, in Buriat it might, however, be either an archaism or the influence of Modern Russian (Gruntov p.c. 2018).

- (22) Buriat (Mongolic, Sanzheev 1953: 137)

hain-ü:d *mori-d*

good-PL horse-PL

‘good horses’

11.3.8 Numeral system

The only numeral system represented in Transeurasian languages in the sample is the decimal one.

11.3.9 Verb

11.3.9.1 TAME

TAME marking is accomplished by means of suffixation. Most Transeurasian languages have present (or non-past, i.e. not dedicated to marking present tense) and past tense marking (23–24).

- (23) Bashkir (Turkic, Yuldashev 1981: 273)

al-di-m

take-PST.INDEF-1SG

‘I took (it).’

- (24) Chuvash (Turkic, Andreev 1997: 485)

yurla-d-əp

sing-PRS-1SG

‘I am singing.’

Japonic languages lack dedicated future tense marking, whereas Tungusic languages, Korean, Bao’an and some Turkic languages mark it. Some Japonic and Turkic languages possess a free-standing particle for marking mood, Ogami (Japonic) and Turkmen (Turkic) for marking aspect, Yakut, Crimean Tatar (Turkic) and Bao’an (Mongolic) for marking tense. Most Transeurasian languages have a morphological distinction between perfective/imperfective aspect and morphological marking of mood. The verb form in the 2nd person imperative mood is identical to the root of the verb in Mongolic and Turkic languages, whereas Japonic, Tungusic languages and Korean have a dedicated suffix marking imperative mood (25).

- (25) Korean (Sohn 1999: 276)

mek-ela

eat-IMP

‘Please eat.’

Evidentiality marking is moderately common in Turkic languages, e.g., in Chuvash (26), Yakut, Khakas, Crimean Tatar, Tatar, Gagauz, in Japonic languages, e.g., in Hateruma, Ogami, Okinoerabu, Yonaguni, Mongolic languages (Mangghuer, Kalmyk,) and Korean.

(26) Chuvash (Turkic, Savelyev p.c. 2017)

- a. *vəʷl kay-nə*
 3SG go-EVID
 ‘He went (apparently).’
- b. *vəʷl kay-rə*
 3SG go-PST.3SG
 ‘He went.’

11.3.9.2 Valency-changing operations

The only valency-increasing strategy across Transeurasian is causativization, which is accomplished by means of suffixation exclusively. As for other strategies of adding arguments to a verb, some Transeurasian languages possess locative markers. Ogami has a “purposive” converb marked by *-ka*, which introduces an argument for a goal of motion. A motion suffix is common in Northern Tungusic languages (see Pakendorf, this volume: Chapter 18) and Oroch.

A morphologically marked passive voice (27) is available as a valency-decreasing strategy for all Transeurasian languages, excluding Chuvash, Nanai, Bao’an and Mangghuer.

- (27) Shuri (Japonic, Shimoji 2012: 376)

ari=nkai sugur-at-ta-n

3SG=DAT hit-PASS-PST-IND

‘Someone was hit by her/him.’

Some Transeurasian languages (among them Korean and Even) use the same marker for passivization and causativization (28–29). However, as this marker became lexicalized (Robbeets 2007: 235), it is not a common isomorphism in modern Transeurasian languages.

- (28) Korean (Koreanic, Sohn 1999: 367)

po-i-ta

see-CAUS/PASS-DECL

‘be seen; show’

- (29) Even (Tungusic, Lebedev 1978: 84)

a. *maa-v-daji*

kill-CAUS/PASS-PTCP

‘be killed’

b. *i:-v-deji*

enter-CAUS/PASS-PTCP

‘carry in’

The agent in a passive construction is most often marked the same way as the recipient in a ditransitive construction, i.e. either as a dative case marker or as a dative particle (30).

(30) Okinoerabu (Japonic, van der Lubbe and Tokunaga 2015: 361–362)

Mariko=ga Taroo=ni ʔabi-ra-tta-mu

Mariko=NOM Taroo=DAT to.call-PASS-PST-IND

‘Mariko was called by Taroo.’ (Masana)

Incorporation of nouns into verbs is not a common intransitivising strategy in Transeurasian languages. In all the Transeurasian languages antipassive marking is absent.

11.3.9.3 Verb morphology in subordinate clauses

Most Transeurasian languages use infinite verbal morphology to indicate subordinate clauses, with the verb marked for finiteness in the main clause, i.e. clause chaining, which is only in rare cases described as such. The converb strategy for marking the distinction between simultaneous and sequential clauses (31) is very common across all Transeurasian languages.

(31) Ulch (Tungusic, Petrova 1936: 58)

buə ʔənə-məri jaja-ha-pu

1PL walk-SIM.PL sing-PST-1PL

‘We were singing while we were walking.’

Among Transeurasian languages of the sample, three subgroups possess an existential verb that is different from the equative copula: most Turkic (Azerbaijani *var*), Japonic (Japanese *aru/iru*) languages and Korean (*issta*). In Mongolic and most Tungusic languages it appears to be identical with the copula (apart from the cases of missing data): Tungusic **bi-*, Mongolic **bu-*, **a-*.

11.3.9.4 Reduplication

Apart from Bao'an (32), which employs verb reduplication for expressing a continuous action, and Tuvan, where verb reduplication “indicates an extension of the action for a definite period of time” (Krueger 1997: 141), verb reduplication is not a common phenomenon in Transeurasian languages (note that only cases where reduplicated verbs constitute a single phonological word are taken into account).

(32) Bao'an (Mongolic; Fried 2010: 102)

atɕaŋ khəl~khəl-tɕə

3SG speak~CONT-PFV

‘He talked and talked (for a long time).’

11.3.10 Attributive possession

In cases, where the possessor is marked on the possessed, pronominal possessors follow their heads (suffixes), nominal possessors precede the possessed across Transeurasian languages. In Turkic, Mongolic and Tungusic languages the possessor is indicated on the possessed by a suffix in attributive possession. Japonic, some Tungusic, most Mongolic and Turkic languages and Korean (33) indicate the possessor with a genitive marker, which can be either a clitic or a suffix.

(33) Korean (Koreanic, Sohn 1994: 174)

na=uy yenphil

1SG=GEN pencil

‘my pencil’

In most Tungusic languages the possessor is unmarked (34).

(34) Even (Tungusic, Kim 2011: 62)

svinija ulrə-n

swine meat-3SG

‘swine’s meat, pork’

Only Tungusic languages and Chuvash (debatable, see Savelyev, this volume: Chapter 27 for Chuvash) have different marking for alienable and inalienable possession. Tungusic languages have special marking for alienable possession (-**ŋi*) in addition to the person of the possessor.

11.3.11 Predicative possession

Transeurasian languages show a variety of ways to express predicative possession:

(i) with a transitive “habeo”-verb (some Japonic languages, e.g., 35), (ii) with a locative-marked possessor (a common strategy in all subgroups of the Transeurasian unity, e.g., 36), (iii) with a dative-marked possessor (available in Japonic, Mongolic, Tungusic languages, e.g., 37), (iv) with a possessor coded as an adnominal possessor (Korean, Tungusic, Mongolic, Turkic languages, Yuwan, e.g., 38), (v) with a

possessor coded as a comitative argument (the least common strategy, available in Yakut, some Mongolic and Tungusic languages, e.g., 39)³.

- (35) Japanese (Japonic, Hinds 1986: 138)

watasi=wa kuruma=o motte iru
1SG=TOP car =ACC possess.PTCP be
'I have a car.'

- (36) Korean (Koreanic, Sohn 1999: 284)

halapeci=kkey chayk=i manh-ayo
grandpa=LOC book=NOM many-POL
'Grandpa has many books.'

- (37) Evenki (Tungusic, Bulatova and Grenoble 1999: 9)

bəjətkə:n-du: kniga bisi-n
boy-DAT book be-3SG
'The boy has a book.'

- (38) Azerbaijani (Turkic, Mehraliev, p.c.)

män-im pişiy-im var
1SG-POSS.1SG cat-POSS.1SG exist
'I have a cat.'

- (39) Kalmyk (Mongolic, Benzing 1985: 56)

surhulc nain denš-tä

student eighty kopecks-COM

‘The student has 80 kopecks.’

11.3.12 Alignment

All Transeurasian languages have accusative (S/A P) alignment of marking of core arguments (40).

(40) Japanese (Japonic, Ishizuka 2012: 3, 192)

a. *keisatu=ga ken=o tukamae-ta*

police=NOM Ken=ACC catch-PST

‘The police caught Ken.’

b. *kondo=wa kiji=ga tonde-ki-mashi-ta*

next=TOP pheasant=NOM fly-come-POL-PST

‘Next a pheasant came flying down [to them].’

Parallel to it, all Turkic and Mongolic languages also have neutral S/A/P alignment of marking due to their differentiation between definite and indefinite objects: indefinite objects do not receive accusative marking and are thus unmarked (the same way as the S/A arguments), e.g., (41).

(41) Shor (Turkic, Dyrenkova 1941: 59)

aŋčï kaŋdus aŋnapča

hunter otter hunt

‘A hunter hunts otter.’

Korean allows omission of all case-marking particles. Some Tungusic languages also exhibit neutral marking, as for them the accusative marking is optional. In Udihe, it can be omitted i) for phonological reasons, ii) if the object is non-specific (42), iii) if the verb is in the imperative.

(42) Udihe (Tungusic, Nikolaeva and Tolskaya 2001: 123)

ŋene-mi ogbö wa:-ni
 go-INF elk kill.PST-3SG
 ‘On the way, he killed an elk.’

The A/S argument is often indexed on the verb by a suffix across Transeurasian languages (43).

(43) Chuvash (Turkic, Andreev 1997: 484)

pïradə-p
 go-1SG
 ‘I go.’

There is variation in the alignment of marking the recipient of a ditransitive construction and the patient of a transitive verb. Chuvash, Mangghuer, Japanese and Korean allow the same marker for both constructions. Other Transeurasian languages employ different markers for these roles.

11.3.13 Negation

Most Transeurasian languages mark negation on the verb by means of a suffix (44).

(44) Bashkir (Turkic, Poppe 1964a: 94)

min unī kyr-mä-ne-m

1SG 3SG see-NEG-PST-1SG

‘I didn’t see him.’

Some Mongolic languages, such as Bao’an, Kalmyk and Mangghuer, do not have inflectional morphology for negation, they mark it by a particle instead. For one of the strategies in Kalmyk, see example (45).

(45) Kalmyk (Mongolic, Benzing 1985: 90)

es bosna:

NEG stand.up.PRS

‘He doesn’t stand up.’

All Tungusic and some Japonic languages (Tarama and Yonaguni) possess an auxiliary for marking standard negation (46).

(46) Evenki (Tungusic, Nedjalkov 1997: 96)

bejumimni homo:ty-va e-če-n va:-re

hunter bear-ACC.DEF NEG-PST-3SG kill-PTCP

‘The hunter didn’t kill the bear.’

It is possible to mark prohibitive and declarative negation (transitive declarative clauses) in the same way in Mongolic and some Turkic languages. Japanese uses the

marker *-nai* for some types of declarative negation as well as for prohibitive negation (prohibitive also requires the infiniteness marker *-de*), but in general different markers for both negation types are employed in Japonic languages and Korean. Most Transeurasian languages employ different negation markers for verbal vs. locative/existential/nominal negation (47a–c), apart from Korean, Nanai and several Japonic languages.

(47) Turkish (Turkic, Kornfilt 1997: 123–125)

- a. *hasan kitab-ï oku-ma-dï*
 Hasan book-ACC read-NEG-PST
 ‘Hasan didn’t read the book.’
- b. *ben hasta deyil-im*
 1SG sick NEG.COP-1SG
 ‘I am not sick.’
- c. *ben ev-de yok-tu-m*
 1SG home-LOC NEG.EXIST-PST-1SG
 ‘I was not at home.’

11.3.14 Word order

In most cases in Transeurasian languages modifiers precede their heads, thus adjectives, numerals and demonstratives usually precede the noun (see Sections 11.3.5 and 11.3.7 for examples). In most Japonic and Mongolic languages a numeral can both precede and follow the noun. In Korean, the standard position for the numeral is the one after the noun. The modifier-head structure also holds for relative clauses: in all Transeurasian languages, apart from Azerbaijani and Khalaj, the

relative clause precedes the noun it modifies. In simple pragmatically unmarked clauses the word order is verb-final both for transitive and intransitive clauses for Transeurasian languages. Clausal objects typically appear in the same position as nominal objects in Transeurasian languages, apart from some Turkic (Azerbaijani borrowed this construction from Persian) and Tungusic languages. The order of main arguments in transitive declarative clauses is rigid in some Mongolic and almost all Turkic languages (apart from Gagauz), whereas Tungusic, Japonic languages and Korean allow variation in the order of A and P, as long as these are appropriately marked for their function. Content interrogatives most often occur in situ in Transeurasian languages.

11.3.15 Interrogation

Marking interrogation by a clause-final question particle is the most common strategy across Transeurasian languages (48).

(48) Shor (Turkic, Dyrenkova 1941: 244)

ol taiga-da aŋ köp-pe
 that taiga-LOC animal many-Q
 ‘Are there many animals in taiga?’

A minor strategy is marking it by intonation, which is present in some Turkic, Tungusic and Japonic languages.

11.3.16 Comparative construction

Comparison is mostly accomplished by means of one kind of locative marking of the standard of comparison. The most common case used in this function is ablative (50).

(50) Azerbaijani (Turkic, Shiraliev 1971: 47)

bakı kirovabad-dan böyükiür

Baku Kirovabad-ABL big

‘Baku is bigger than Kirovabad.’

All Japonic languages, Ulch, Nanai, Oroch and Korean have a marker that has neither locational meaning nor the meaning ‘surpass/exceed’ (51).

(51) Japanese (Japonic, Kaiser et al. 2013: 42)

gyu:niku=ga butaniku yori yasui

beef=NOM pork COMP cheap

‘Beef is cheaper than pork.’

The adjective in a comparative construction is unmarked in most Transeurasian languages or marked optionally, apart from a number of Turkic languages, both Even dialects and Evenki (52).

(52) Even (Lebedev 1978: 55)

bii hin-duk egdže-tmır

1SG 2SG-ABL high-COMP

‘I am higher than you.’

11.3.17 Coordination and conjunction

Conjunction vs. coordination marking has internal discrepancies among Transeurasian languages. Nanai, Oroch, Evenki, Buriat, Khalkha, Kipchak Turkic languages, Yakut, Khakas and Azerbaijani use different morphemes to express conjunction and comitative (e.g., 53).

(53) Evenki (Tungusic, Bulatova and Grenoble 1999: 12, 56)

- a. *bi: əkin-nu:n-mi: təwlə:-m*
1SG sister-COM-REFL.SG collect.berries-1SG
‘I went with my sister to pick berries.’
- b. *bi: taduk girki-w ollo-mo:-čo:-wun*
1SG and.then friend-POSS.1SG fish-go-PST-1PL.EXCL
‘My friend and I went fishing.’

11.3.18 Obligatoriness of S/A argument

Most Transeurasian languages allow omission of the S/A argument (54).

(54) Mangghuer (Mongolic, Slater 2003a: 124)

- ning ge khuba di ge-jiang*
this do divide eat do-OBJ:PFV
‘Like this (they) divided and ate (him).’

11.3.19 Derivation of adpositions

Adpositions are often derived from place nouns by locational suffixes, e.g., (55–56).

- (55) Buriat (Mongolic, Sanzheev 1962: 301–304)

bäe-hä:n

body-ABL

‘from the side’

- (56) Evenki (Tungusic, Bulatova and Grenoble 1999: 13)

amut daga-la:-n

lake close-LOC-POSS.3SG

‘closer to the lake’.

11.3.20 Classifiers

Numeral classifiers are the only type of classifiers present in Transeurasian languages.

These are common in Japonic languages, some Tungusic languages, such as Evenki, Ulch, Negidal (the latter probably under the influence of Ulch, Oskolskaya p.c. 2017), Nanai, Turkic (Crimean Tatar), Mongolic (Mangghuer) and Korean. In Evenki there are numeral classifiers differentiating human and non-human counted entities (57a–b).

- (57) Evenki (Tungusic, Nedjalkov 1997: 283)

a. *nadan-i:*

seven-CLF:HUM

‘seven people (together)’

b. *nada-ngna*

seven-CLF:NHUM

‘seven objects (together)’

11.4 Phylogenetic analysis

11.4.1 Coding procedure

The current study encompasses a heterogeneous language sample consisting of 38 doculects and 226 binary structural features. I use 189 formulations of the features from the Grambank database (Hammarström et al. 2017), 10 binarised versions of Grambank features on word order and 27 features relevant for Transeurasian languages (partly from Robbeets 2017c). I coded the features based on descriptive works, dictionaries and personal correspondence with language experts. The dataset with the coding for each individual language for each feature as well as the description of the structure of the data set can be found in the online supplementary materials.

The feature set provides an extensive coverage of morphosyntax of the language (e.g., person and number marking on nouns, possessive constructions, interrogation, negation, derivation patterns, valency operations, numeral systems, comparison, argument marking, deixis) as well as phonology (voicing distinction in plosives and fricatives, l/r distinction, constraints on initial consonants, availability of initial consonant clusters, vowel harmony, vowel length).

The four main criteria for feature selection are: i) stability, ii) informativity, iii) codability, iv) logical independence. The first criterion is fulfilled by the preselection of the features for their being stable cross-linguistically in Grambank. The “Transeurasian” features are assumed to be stable by Robbeets (this volume: Chapter 10). The second criterion foresees informativity of the features. The features, which are not part of Grambank, were added based on their variation in the language sample. This aims at resolving the internal relationships between the languages in question.

The third criterion takes into account the coverage of the respective topic by reference grammar. In this way, languages with extensive descriptions available can be included as well as those with only grammar sketches. Grambank features have been preselected to meet this criterion. Some “Transeurasian” features were excluded a-posteriori due to the low coverage of the respective topics in the descriptive works. According to the fourth criterion, the value of one particular feature has to be independent of the value of another feature, i.e. neither triggered nor predicted by it.

11.4.2 Stability of structural features

I delimit structural features stable in Transeurasian languages by extracting the features with a high genealogical signal. To avoid circularity, I calculate the signal along the tree based on lexical data and phonological correspondences (see Figure 11.2). For each feature with moderate variation (149 features in total), I calculate the phylogenetic signal with the metric Fritz and Purvis’ D using the function “phylo.d” from the package “caper” in R. This method takes into account the distribution of the feature values in sister branches: if sister languages have the same feature value, the D value will be low and thus the phylogenetic signal will be high.

<Insert Figure 11.2 here>

Figure 11.2 The tree used for the estimation of the phylogenetic signal

To set a cut-off point for the stability of the features, I compare the distribution of the D values for the real data and the randomized data (see Figure 11.3). I set this point to the two standard deviations from the mean of the randomized data, i.e. 0.53.

<Insert Figure 11.3 here>

Figure 11.3 Estimated D values for real data (blue) and randomized data (red). The vertical line indicates the two standard deviations from the mean threshold.

65% of the features have a D value smaller than 0.53 and can thus be considered relatively stable. The impact of the language domain, which the feature covers, the genealogical attribution of the languages in question and the proportion of 0's and 1's for a particular feature can impact the estimated phylogenetic signal in the feature. We will have a closer look at the influence of the part of speech on the amount of the phylogenetic signal in the feature and its interaction with other factors.

The phylogenetic signal differs across features covering different parts of speech (see 11.3). The differences in the distribution of D values across parts of speech can be explained by at least two factors. First, the number of the features that correspond to a particular language domain differs (6 features on adjectives, 2 on articles, 3 on demonstratives, 23 on nouns, 2 on numerals, 17 on pronouns, 42 on verbs). Second, extremely low values of D are often due to high uniformity of features, e.g., 36 out of 38 languages have the same value for a particular feature and this leads to underestimation of D values. This is particularly the case for pronouns, where most of the features have the same value 0 for all the languages except one or two.

<Insert Figure 11.4 here>

Figure 11.4 Estimated D values across parts of speech. The horizontal line indicates the 0.53 threshold.

11.4.3 Bayesian approach to the classification of the Transeurasian languages

I use the whole data set and the data set with only the stable features delimited according to the procedure described previously in Section 11.4.2 to build two topologies of the Transeurasian languages (compare Figure 11.5 and Figure 11.6). The underlying Bayesian analysis derives a distribution of trees instead of a single tree. The more often a particular clade (a language grouping) appears in this distribution, the higher is the credibility of the clade and the lower the uncertainty within the clade.

<Insert Figure 11.5 here>

Figure 11.5 Topology of the Transeurasian languages based on the whole data set

<Insert Figure 11.6 here>

Figure 11.6 Topology of the Transeurasian languages based on stable structural features

The traditional affiliation of languages to the respective language families is replicated in the topology based on the whole data set, except for Yakut. As the Mongolo-Yakut branch is short and the posterior probability for the clade is low, Yakut must have split from the Turko-Mongolic ancestor at approximately the same time as Mongolic and Turkic split into two branches. The posterior probabilities for the individual language families are moderately high: 1.00 for Koreano-Japonic, 0.83 for Altaic, 0.98 for Tungusic, 0.77 for Mongolo-Turkic, 0.7 for Mongolic and 0.71 for Turkic excluding Yakut.

The internal structure of each smaller-level language family is replicated to a different extent, which is reflected in the high uncertainty (i.e. low posterior

probability estimates) in the clades. There might be several explanations for this. First, it may be a result of horizontal transmission, i.e. a high number of borrowing events between the languages. Second, the branches may be so closely related that it is difficult for the algorithm to resolve them (Dunn et al. 2008). Both explanations are valid for some branches in the Transeurasian topology. For example, Yakut appears outside the Turkic cluster due to its known history of contact with Tungusic languages. Turkic languages are structurally too similar to each for the algorithm to reliably establish the individual groupings. The same explanation might also be valid for Japonic languages. On the positive side, the known close interrelatedness of Ulch and Nanai and two Even dialects respectively is replicated in the high posterior probabilities in the tree.

Some of these relationships are replicated if the features with the higher D values, i.e. the ones assumed to be stable, are excluded. The main structure of the tree, Japonic-Koreanic vs. Altaic branch, Tungusic languages splitting off first from the Proto-Altaic ancestor, remains intact. High posterior probabilities for Japonic-Koreanic and Tungusic branches are also preserved. After the reduction of feature number to 97, Korean does not appear as a separate branch anymore, Yakut disrupts the structure of the Mongolic language family by appearing inside the Buriat-Khalkha-Kalmyk cluster and separating it from Baoan and Mangghuer, posterior probabilities for Altaic, Turkic and Mongolic branches drop. This result might be due to the following methodological restrictions. First, the metric functions reliably for language samples with 50 languages and more (Fritz and Purvis 2010: 1050). Second, a different strategy for the delimitation of the features could lead to a better topology.

The exclusion of “unstable” structural features does not improve the internal classification of the individual language families in terms of making it more similar to

the classification based on lexical data and phonological correspondences as expected. Instead, the structure of the Mongolic language family is disrupted by the interference of Yakut and the probabilities for the clades drop in most cases.

Comparing the Bayesian classifications of the individual language families (Savelyev, this volume: Chapter 9; Whaley and Oskolskaya, this volume: Chapter 6), I come to a conclusion that structural features perform worse in terms of confidently disentangling the internal structure of the lower-level branches, but well enough to replicate the affiliation of most languages within their respective language families, if the number of features is sufficiently high (or at least as high as the vocabulary lists used in this volume, i.e. around 200).

11.5 Conclusion

The chapter addressed the following research questions: What does a typological profile of the Transeurasian languages look like? Do structural features provide a reliable tree of the Transeurasian languages? Are there differences in structural features in terms of their phylogenetic signal? Do structural features with a high phylogenetic signal provide a “better” tree than the whole set? These questions aimed at filling the gap in the debate on the internal structure of the Transeurasian unity, on the suitability of structural features for building trees that represent true language history (ideally genealogical relationships between languages) and on the stability of structural features.

There is a consensus on the fact that the Transeurasian languages share structural similarities, but no quantitative approach has been applied to the structural data to address the issue of the exact interrelationships between these languages. The previous research suggested a genealogical relationship between Japonic and

Koreanic languages and a genealogical grouping of Turkic, Mongolic and Tungusic languages. The results of the study show that the distribution of the structural features among Transeurasian languages supports a division between Altaic and the Japonic-Koreanic unities. This split is also reflected in the results of the Bayesian analysis through the binary structure of the Transeurasian tree with the Altaic and the Japonic-Koreanic branches. There is also a noteworthy tendency of Tungusic languages to follow either an Altaic or a Japonic-Koreanic type in a number of features.

The features in Table 11.2 can be assumed to constitute the Transeurasian language type synchronically, according to the frequency of their occurrence across Transeurasian languages. Features that are common only in Altaic languages are still listed in Table 11.2, as they are frequent in 3 of 5 branches of the Transeurasian unity.

Table 11.2 Typological profile of the Transeurasian languages

Language domain	Features
Phonology	vowel length distinction; vowel harmony; no word-initial velar nasals and consonant clusters; two-fold division of the distribution of the distinction in liquids
Nominal morphosyntax	regular plural marking; either the same marker for associative plural or an additional one; rich derivational morphology; morphological core case marking; nominal reduplication; oblique pronominal stem with a nasal; no agreement in number between the noun and adjective/demonstrative; no plural marking on the noun in numeral-noun phrases; GEN marking of the possessor; possessor indicated on the possessed by a suffix; accusative

	alignment of marking of main arguments; ABL case marking of the standard of comparison in a comparative construction; adpositions derived from place nouns marked with locative cases; NP word order: modifier-head;
Verbal morphosyntax	passivization and causativization by morphological means; clause chaining; morphological marking of negation; verb agreement with the S/A argument in person and number
Word order and other clause-level construction	SOV word order; pro-drop languages; clause-final particle for marking interrogation; LOC/DAT marking of the possessor (lit. 'The cat is on/to me.') or coding of the possessor as an adnominal possessor (lit. 'My cat exists.') in a predicative possession construction

There is an ongoing debate on the stability of structural features. Despite the discrepancies in the results achieved in previous studies, most scholars agree upon the fact that there is at least a set of genealogically stable structural features. This study has measured the phylogenetic signal in the structural features of the Transeurasian languages by applying the metric Fritz and Purvis's D. The most commonly discussed range of the D values is between 0 (strong phylogenetic signal) and 1 (the feature is distributed randomly on the tree). The analysis of the stability of the Transeurasian structural features has shown that the features vary in terms of the phylogenetic signal and more than a half of the features with moderate variation have a high genealogical signal.

The current study has thus provided a summary of the typological profile of the Transeurasian languages, suggested an internal structure of the Transeurasian unity

based on the structural features and calculated the phylogenetic signal in the structural features. The internal structure of the Transeurasian unity achieved in this study goes in line with the proposal of Robbeets and Bouckaert (2018) on the Transeurasian tree consisting of the Japonic-Koreanic and Altaic branches, with Altaic splitting further into Mongolo-Turkic and Tungusic branches. The structural features with a high phylogenetic signal do not point to a tree of Transeurasian languages suggested by historical comparative linguists. In order to account for the source of the similarities between languages, a further study is needed, where the geographical location of languages (synchronically) and the nodes (diachronically) is controlled for. Neither the tree based on the whole set of structural features nor the tree based on the stable set of features provide a reliable structure of Turkic, Mongolic, Tungusic and Japonic language families.

Acknowledgements

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 646612) granted to Martine Robbeets.

¹ The position of Khalaj is debatable. Here its position is in line with the classification by Savelyev, this volume: Chapter 9.

² The information on a language in this chapter refers to a particular language as it is documented in the language description. The current state of the language can therefore deviate from the form described in the language grammar, which was used for this

study. By referring to a particular “language” I thus mean a “doculect”, if not noted otherwise.

³ In cases, where the same marker is used for several functions (e.g. dative and locative are identical), the languages were included double. E.g. in Mongolic and Japonic languages the same marker is used for dative and locative, thus these two groups are implied both in 12.3.11 (ii) and (iii).

This is a draft version of a chapter that appears in Robbeets, M. and A. Savelyev (eds). *The Oxford Guide to the Transeurasian Languages* (OUP, 2020), see <https://global.oup.com/academic/product/the-oxford-guide-to-the-transeurasian-languages-9780198804628>. The research leading to these results has received funding from the European Research Council under the Horizon 2020 Program/ ERC Grant Agreement n. 646612 granted to Martine Robbeets.