Chapter 3

The classification of the Transeurasian languages

Martine Robbeets

Abstract

Even if the hypothesis of Transeurasian affiliation is gradually gaining acceptance, supporters do not coincide on the internal structure of the family. Over the last century, a range of different classifications has been proposed. While these proposals show some remarkable overlap, the position of the Tungusic branch in the family tree remains a recurrent issue. Here I infer the best supportable tree for the Transeurasian family, notably a binary topology with a Japano-Koreanic and an Altaic branch, in which Tungusic is the first to split off from the Altaic branch. To this end, I combine the power of classical historical-comparative linguistics with computational Bayesian phylogenetic methods. In this way, I introduce a quantitative basis to test various competing hypotheses with regard to the internal structure of the Transeurasian family and to solve uncertainties associated with the application of the classical historical-comparative method.

Keywords: Classification, Bayesian phylolinguistics, lexicostatistics, historical comparative linguistics, Transeurasian, Japanese, Korean, Tungusic, Mongolic, Turkic

3.1 Introduction

Throughout this reference guide, we use the term "Transeurasian" to refer to a group of geographically adjacent and structurally homogeneous languages across Eurasia that consists of five uncontroversial families: the Turkic, Mongolic, Tungusic, Koreanic and Japonic languages. Johanson and Robbeets (2010a: 1–2) coined the label to complement the traditional term "Altaic", which we reserve for the unity of the Turkic, Mongolic and Tungusic languages only. Figure 3.1 shows the distribution of 23 contemporary Turkic languages, 10 Mongolic languages, 10 Tungusic languages, 6 Japonic languages in addition to Korean. The abbreviations for languages are explained in the list of abbreviations in the preface of this guide.
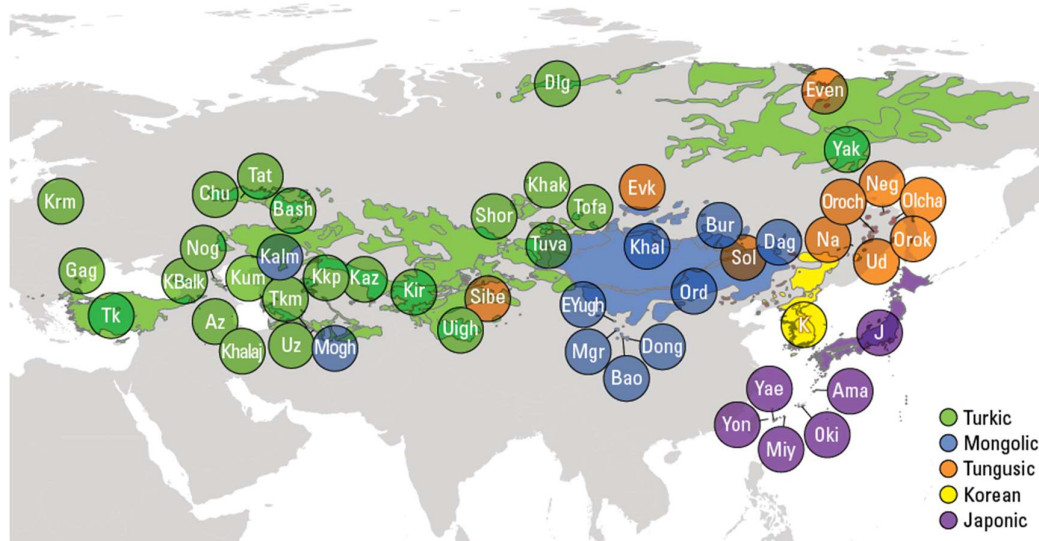

Figure 3.1 The distribution of the Transeurasian languages (Robbeets and Bouckaert 2018).

The question of whether these five groups descend from a single common ancestor has been the topic of a longstanding debate. As early as 1692 the first known

reference to a unity of Turkic, Mongolic and Tungusic languages was made by Nicolaes Witsen, but the idea was probably derived from Abu al-Ghazi Bahadur's *Shajare-i Türk* 'Genealogy of the Turks' from 1661, a Russian translation of which was circulating in Europe by the end of the 17th century. The work was translated into French and published in Leiden in 1726.

Since then, the concept of this linguistic unity was comprised under different labels such as 'Tatar', 'Scythian', 'Ural-Altaic' or 'Altaic', it became expanded with various languages and reduced again, in turns fervently supported and heavily criticized among others by Strahlenberg (1730), Abel-Rémusat (1820), Klaproth (1823), Castrén (1850), Schleicher (1850), Grunzel (1895) and Bang (1895). However, Ramstedt is usually considered the founder of Transeurasian linguistics because he established a modern linguistic framework for Transeurasian comparison, supported by regular sound correspondences (1957) and morphological cognates (1952). While, until the late sixties, the field focused on the comparison of Turkic, Mongolic and Tungusic on the one side (e.g. Poppe 1960b, 1965, 1975) and of Korean and Japanese on the other (e.g. Martin 1966), in the seventies, Miller's (1971) monograph "Japanese and the other Altaic languages" increased the scholarly interest in the overall comparison of these languages. Clauson (1956) and Doerfer (1963–1975) raised substantial criticism against the genealogical relatedness of these languages, which was mainly based on the alleged lack of shared basic vocabulary and the explanation of all correlations by borrowing; see Robbeets, this volume: Chapter 36. Starostin et al. (2003) resurrected scholarly interest in the Transeurasian unity, accumulating a body of evidence that was far more impressive in quantity and rich in empirical material than the number and scope of etymologies proposed previously.

However, these new matches were, in their turn, criticized for reason of phonological, morphological or semantic overpermissiveness, among others by Robbeets (2005), leaving room for a reduced core of reliable etymologies and by Vovin (2005c, 2009c, 2010) and Georg (2007b), completely rejecting all evidence advanced so far.

In my past research (Robbeets 2005, 2015), I suggested that even if the majority of support provided in the past is questionable, there is nonetheless a core of reliable evidence for the classification of Transeurasian as a valid genealogical grouping. In line with the requirements of the classical comparative method of historical linguistics, the evidence consists of regular sound correspondences, lexical etymologies including common basic vocabulary (see Robbeets, this volume: Chapter 10 and 36) and shared verb morphology (see Robbeets, this volume: Chapter 30). As a result, the hypothesis that the Transeurasian languages are related is gradually gaining acceptance in the field (Gözaydin 2006; Rozycki 2006; Büyükmavi 2007; Décsy 2007; Dybo 2016; Kara 2007).

Whereas supporters of Transeurasian affiliation basically agree about the unity of the family, they do not necessarily coincide on its internal structure. In this chapter, I present the different classifications previously proposed for the Transeurasian family and show how the application of different methods can result in different topologies for the Transeurasian tree. The next Section introduces previous classifications. Section 3.3 presents the classification inferred on the basis of the historical comparative linguistic method, while Section 3.4 provides the Bayesian classification. Section 3.5 discusses the correlations and discrepancies between both trees and argues that the Bayesian analysis yields the tree with the best support for the observed data.

3.2 Previous classifications

Over the last century, various classifications of Transeurasian have been suggested on the basis of either the classical historical-comparative method (Vladimircov 1929: 44–47; Poppe 1965: 147; Street 1962: 95; Miller 1971: 44; Baskakov 1981: 14; Tekin 1994: 82; Robbeets 2015: 506), Bayesian phylolinguistics (Robbeets and Bouckaert 2018) or lexicostatistics (Starostin et al. 2003: 236; Blažek and Schwarz 2014). The first two methods are character-based approaches, which estimate the relationship between two languages by inferring the pathways by which they developed from their common ancestor, while the latter is a distance-based approach, which estimates the relationship between two languages from the amount of difference in their shared cognate proportion (Dunn 2015). In lexicostatistics the distance metric is the diverging cognate proportion between two languages in a basic vocabulary list. The Transeurasian lexicostatistic classifications all use the mathematical algorithm proposed by Starostin (1989). The main criticism against the lexicostatistic approach is not so much in the way it derives the internal tree structure, but rather in its assumption of a constant rate of language change in dating the nodes of this tree (McMahon and McMahon 2006; Campbell and Poser 2008; Greenhill 2015). When limited to the internal classification of the Transeurasian family, however, the method seems to yield acceptable results that show a serious amount of overlap with typologies inferred through character-based approaches.

The classical comparative method, a character-based approach, which relies on the so-called "parsimony method", seeks the tree that explains a dataset by minimising the number of evolutionary changes required to produce the observed state. It is based

on the principle of shared innovations: the best tree is the one that places the innovations where they create the greatest amount of diversity. The strength of this method is that it involves data from various levels of language structure such as phonology, morphosyntax and lexicon, but its limitation is in the cherry picking of features underlying the proposed innovations.

A character-based method for tree-building that is quantitatively better underpinned is the Bayesian method, which seeks to explain a set of observed data by quantifying how likely it is that they have been produced by a certain model of the evolution of cognates along a tree. As the method can score different trees according to how well they explain the data under the chosen model of cognate change, it can sample the best tree from a large number of possible trees and it can give us an idea about how sure we really are that a certain branch is placed at a certain position in the tree. The weakness of the method is that it abstracts detailed comparative datasets in the basic vocabulary into a mere sequence of 1 's and 0 's and that the "mathemagic" it involves is not made sufficiently transparent for classically trained historical linguists.

Even if the existing Transeurasian classifications are based on different datasets and derived by different methods by scholars from different theoretical backgrounds at different times, they display a remarkable overlap. Indeed, they all agree that, first, if a Japonic branch is postulated, Koreanic and Japonic are more closely related to one another than to any of the other branches concerned and, second, that Mongolic forms a binary unity with either Turkic or Tungusic, distinct from the Japano-Koreanic branch. The main difference in the proposals so far has to do with the position of Tungusic vis-à-vis the other branches: Does Tungusic represent a first-order split, which separated simultaneously with Japano-Koreanic and Mongolo-Turkic? Does

Tungusic cluster with Japano-Koreanic or does it rather belong with Mongolic and Turkic? And, if the latter is the case, does Tungusic stand in a binary unity with Mongolic or not?

Figure 3.2 displays the first set of proposals, notably a polytopology whereby Tungusic separated simultaneously from Japano-Koreanic and Mongolo-Turkic. It is the classification supported by the so-called "Moscow School", which was first proposed by Vladimircov (1929: 44–47) and lived on in the view of Baskakov (1981: 14), both scholars using the classical historical-comparative method. More recently, however, the same tree was replicated on the basis of lexicostatistic methods by Starostin and his colleagues. Both approaches conceive of the Transeurasian family as consisting of three principal groups, Turko-Mongolic, Tungusic and Japano-Koreanic, but, contrary to the classical conception, in Starostin's view, Turko-Mongolic and Japano-Koreanic separated around the same time, in the fourth millennium BC.
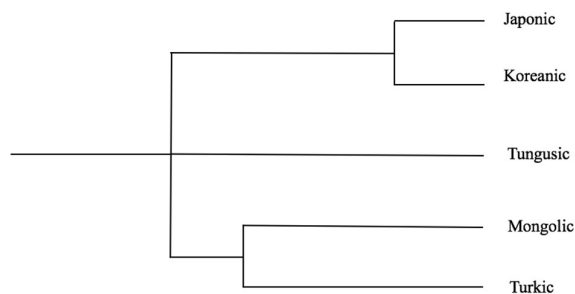

a) Baskakov (1981: 14)
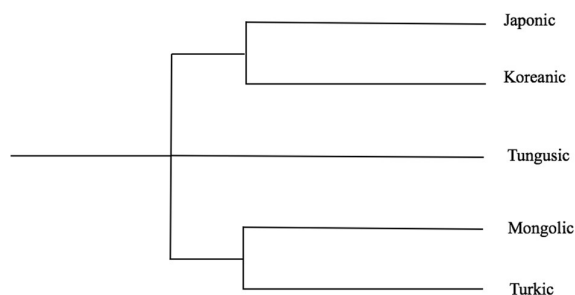
b) Starostin et al. (2003: 236)

Figure 3.2 Previous classifications suggesting a polytopology for the Transeurasian family

The second set of representations involves a binary topology in which Tungusic clusters with the Japano-Koreanic branch, separately from the Mongolic and Turkic branches, as shown in Figure 3.3. It is favoured by some Transeurasian linguists in the West, especially by specialists in Japanic and Koreanic languages. Using the historical-comparative method as a basic tool, Miller (1971: 44) proposed a unity between Tungusic, Koreanic and Japonic, which recalls the suggestion made by Unger and the Altaic panel (1990: 481) to limit the Transeurasian reconstructions to a "Macro-Tungusic" perspective, consisting of Tungusic, Koreanic and Japonic languages only. However, unlike Unger's proposal, Miller conceived of the position occupied by the Ryukyuan languages as independent from Mainland Japanese. The view of a separate unity between Tungusic, Koreanic and Japonic, which was initially reached by using the classical historical-comparative method, was recently supported by Blažek and Schwarz's (2014: 90) application of lexicostatistic methods. However, they conceive of Mongolo-Turkic as the second unity making up the Transeurasian
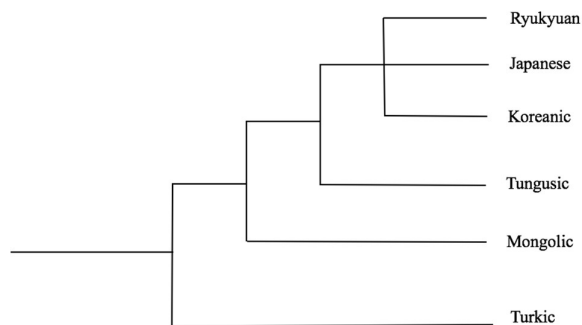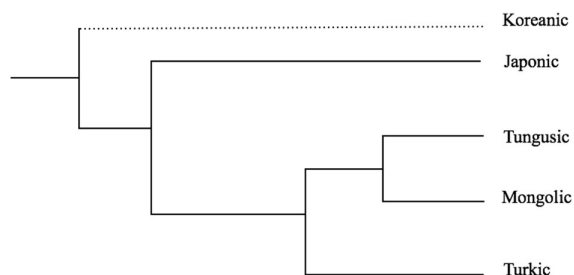
family.

a) Miller (1971: 44)
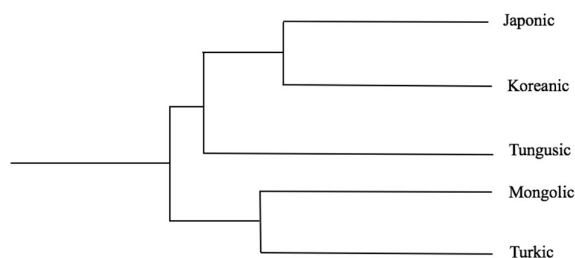
b) Blažek and Schwarz (2014: 90)

Figure 3.3 Previous classifications suggesting a binary topology for the Transeurasian family, whereby Tungusic clusters with the Japano-Koreanic unity

Figure 3.4 shows the third set of conceptions, which represents the consensus view among most western scholars, classifying the Transeurasian languages on the basis of the historical-comparative method. It also reflects a binary topology, but here Tungusic clusters with the Mongolic and Turkic branches to form a separate "Altaic" unity, bifurcated from the Japano-Koreanic unity. In this view, Tungusic stands in a
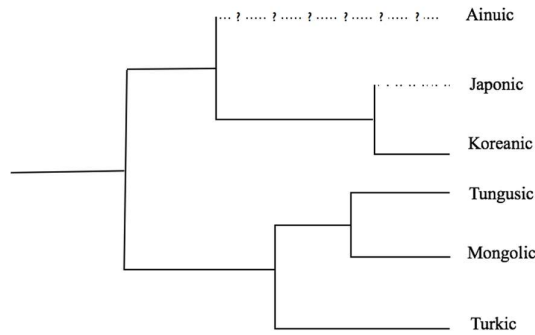
binary unity with Mongolic, and Turkic is the first to branch off from the Altaic unity. Poppe (1960b, 1965: 147) included Korean as a separate branch of Altaic but later he remarked that "Korean is a language only partly belonging to the field of Altaic studies" (Poppe 1975: 172), referring to the possibility that Korean could be a non-Transeurasian language imposed on a Transeurasian substratum. This possibility is indicated with a dotted line in Figure 3.4. In his review of Poppe (1960b), Street (1962: 95) suggested a different configuration for the Japanese and Korean branches, speculating that the Japano-Koreanic branch could eventually cluster with Ainu. The dotted line with the question marks in Figure 3.4 represents Street's uncertainty about the inclusion of Japanese and Ainu. Tekin (1994: 82) included Koreanic in the classification, assuming that Proto-Koreanic was first to branch off from the Transeurasian unity, but he did not accept the inclusion of Japonic into the family.
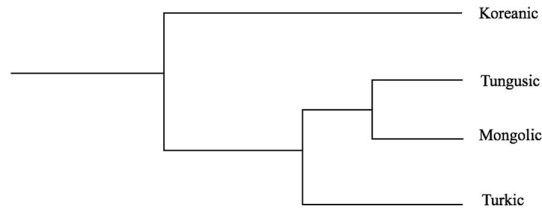

a) Poppe (1965: 147)

```
... ? ..... ? ..... ? ..... ? ..... ? ..... ? ....   Ainuic

                              . . .. ... ..   Japonic

                                              Koreanic

                                              Tungusic

                                              Mongolic

                                              Turkic
```

b) Street (1962: 95)

```
                                              Koreanic

                                              Tungusic

                                              Mongolic

                                              Turkic
```

c) Tekin (1994: 82)

Figure 3.4 Previous classifications suggesting a binary topology for the Transeurasian family, whereby Tungusic clusters with the Altaic unity and Turkic branches off first

Finally, there is a fourth proposal, whereby Turkic stands in a binary unity with Mongolic and Tungusic is the first to branch off, as shown in Figure 3.5. Although Ramstedt never drew a concrete classification, he seems to support this topology in his research. In addition to his concept of the Mongolo-Turkic languages, i.e. "die mongolisch-türkischen Sprachen" (Ramstedt 1912), he described the relations between the Transeurasian languages as follows: "Old Turkish and Old Mongolian were no more remoted from each other than are English and German today. Both

were descendants of an older language, from which also the Tunguses of today have inherited their speech..." (Ramstedt 1924: 438); "Japanese ... is of course not directly related to Turkish or Mongolian or Tungus, but with the oldest possible source of pre-Altaic." (Ramstedt 1924: 439); and "If Japanese is supposed to be genetically related to Altaic, we must presume that also Korean is related to Altaic." (Ramstedt 1924: 440).
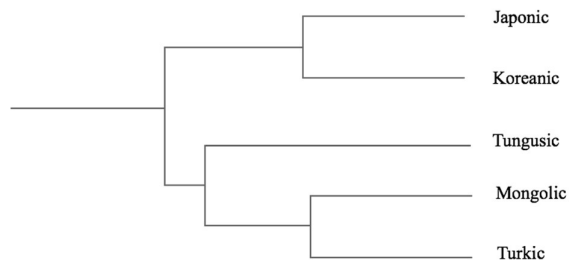

Figure 3.5 Ramstedt's suggestion of a binary topology for the Transeurasian family, whereby Tungusic clusters with the Altaic unity and branches off first.

3.3 Classification on the basis of the historical comparative linguistic method

The only criterion for subgrouping available within the historical comparative linguistic method is the principle of shared innovation. A shared innovation is the result of a change, which took place in a single daughter language, which then subsequently diversified into daughters of its own. When several languages share a particular innovation, this suggests that they descend from a common source, which split up, leaving evidence of this innovation in its daughters. Whereas shared innovations in morphosyntax are just as diagnostic as those in phonology, shared retentions cannot tell us much about subgrouping because they do not reveal which

languages share a period of common history after the break-up of the proto-language.

3.3.1 Shared innovations in Japano-Koreanic

Among the phonological innovations that group the Japonic and Koreanic languages together we find (1) the loss of voicing distinction for stops perhaps in connection with the development of a register distinction (see F4 in Robbeets, this volume: Chapter 10); (2) the neutralization of the Proto-Transeurasian velar fricative *x to PJK *k; (3) the merger of the two Proto-Transeurasian liquids *l and *r into a single liquid phoneme PJK *r; and (4) the Proto-Transeurasian heterorganic clusters *-$m^{(P)}T$-, *-$n^{(T)}K$-, *-$ŋ^{(K)}T$- losing their final stop in Proto-Japano-Koreanic (see Robbeets, this volume: Chapter 36).

Among the morphosyntactic innovations we find (1) the addition of anticausative meaning in Japonic and Koreanic to the originally causative meaning of the Proto-Transeurasian suffix *-gi- (Robbeets 2015: 315–329; e.g., K *cicc-ki*- 'be torn', OJ *ware*- < *$wara$-(C)i- 'be split'); (2) the reduction of allomorphy of the Proto-Transeurasian suffixes *-xA ~ *-kA resultative nominalizer and *-xU ~ *-kU nominalizer/infinitive into PJK *kA and *ku (Robbeets, this volume: Chapter 30; Robbeets 2015: 396–416, 466–480); (3) the insertion of a copula PKJ *a- in the converb construction PJK *a-ku, derived from the Proto-Transeurasian converb *-xU ~ *-kU (Robbeets 2015: 467–470; e.g., OJ *tir-aku* 'the falling', K *kwul-ek* 'the acting'); (4) the insertion of a copula PKJ *wo- in sentential nominalizations with PJK *-mA and -rA, derived from the Proto-Transeurasian nominalizers *-mA and *-rA (Robbeets 2015: 339–378); and (5) the development of a three-way contrast in Proto-Japono-

Koreanic demonstratives from an original two-way proximal-distal distinction in Proto-Transeurasian (see F6 in Robbeets, this volume: Chapter 10).

3.3.2 Shared innovations in Altaic

A phonological innovation shared between the Altaic languages is the loss of the initial nasal in Proto-Transeurasian heteroganic clusters of the shape $*-m^{(P)}T-$, $*-n^{(T)}K$, $*-\eta^{(K)}T-$. Among the morphosyntactic innovations we find (1) the assimilation of the initial labial voiced stop $b-$ to the nasal oblique suffix $-n$ and the resulting *mi-Ti* opposition in the first and second pronoun paradigms (see F11 in Robbeets, this volume: Chapter 10); (2) the merger of mixed encoding of property words in Proto-Transeurasian into mainly nominally encoded property words in Altaic (F7 in Robbeets, this volume: Chapter 10); and (3) the replacement of the original Transeurasian negative *ana-* by the Altaic negative *ə-* (Robbeets 2015: 177–200).

In addition to 9 innovations in support of the Japano-Koreanic subgrouping, we thus find 4 innovations in support of the Altaic one. Therefore, the primary branching in the Transeurasian family is well supported. The observation that the Altaic languages are more conservative while the Japano-Koreanic are more innovative, may result from association of the primary break-up with a migration of Japano-Koreanic speakers to the Liaodong Peninsula, where substratum inference may have induced change, whereas the Altaic speech community is thought to have remained in the original Transeurasian homeland (see Robbeets, this volume: Chapter 44).

3.3.3 Shared innovations between Turkic and Mongolic

A potential instance of a shared phonological innovation between the Turkic and Mongolic languages is the reflex of the medial velar fricative PTEA *-x-, which yields an alternation between *-g- and *-k- in Mongolic and Turkic languages, but is retained as *-x- in Tungusic (Robbeets 2015: 438–440), e.g., in PTg *daxa- 'to follow' (e.g., Ma. daha-, Neg. daxaw-, Na. daxa- 'to follow') against PMo *daka- ~ *daga- 'to follow' (e.g., Khal. daga-, Bur. daxa- 'to follow') and PTK *yak- ~ *yagu- 'to become near' (e.g., OTk yak- ~ yagu- 'to approach, become near').

Possible cases of morphosyntactic innovations include (1) the development of deverbal anticausative meaning in Turkic and Mongolic on the denominal fientive suffix PTEA *-dA- (e.g., WMo. dugul-da- 'to be heard, be audible', OTk uya-d- 'to be ashamed'; Robbeets 2015: 301–308); and (2) the development of converbial and adverbial meaning in Turkic and Mongolic on the instrumental and action/state nominalizer PTEA *-i (e.g., OTk tak-ï 'in addition', WMo. daru-i 'thereafter'; Robbeets 2015: 455–466). However, as these morphosyntactic innovations are also present in Japonic and/or Koreanic, we cannot exclude that they reflect shared retentions of the ancestral Transeurasian state.

3.3.4 Shared innovations between Mongolic and Tungusic

Although there are many shared retentions, there are no straightforward examples of shared phonological innovations in Mongolic and Tungusic languages. Nevertheless, we find several shared morphosyntactic innovations, which relate to the shared derivation and grammaticalization of compound morphemes. These include (1) the grammaticalization of plural or collective markers to express an inclusive/exclusive distinction on the first person pronoun (e.g., MMo bi-da (1SG-PL) 'we (inclusive)',

Ma. *mu-se* (1PL-COLL) 'we (inclusive)'; see F 20 in Robbeets, this volume: Chapter 10); (2) the development of the Proto-Transeurasian nominalizer *-mA* to include converbial and finite use in Tungusic and Mongolic languages in addition to the compounding of this suffix with a possessive reflexive marker to derive a co-referential converb (e.g., Khal. *-mAAn* converb < *-m-AAn* NMLZ-REFL; Evk. *-mi* converb < *-mA-wi* NMLZ-REFL; Robbeets 2015: 361–378); (3) the compounding of the reflexes of the Transeurasian inchoative *-xa* with the resultative nominalizer *sa*- to derive the deverbal perfective nominalizers PTg *-xsan* and PMo *-gsAn ~ -ksAn* (e.g., Evk. *-ksA* perfective converb; MMo. *-ɣsan ~ -gsen / -qsan ~ -ksen* perfective nominalizer; see Robbeets 2015: 424, 431); and (4) the formation of a converb on the basis of the reflexes of the Transeurasian nominalizer *-nA* and the diminutive suffix *-kAn* (e.g., Neg. *-nAxAn* converb < *-nA-kA:n* NMLZ-DIM; Khal. *-n-xAn* CVB-DIM; Robbeets 2015: 385, 392).

3.3.5 Shared innovations between Turkic and Tungusic

Except for the development of head marking in the noun and verb phrase, I was unable to find potentially shared innovations between Turkic and Tungusic languages. Whereas Transeurasian was originally dependent marking, we find (1) person agreement on the verb in the Turkic and Tungusic languages; and (2) head-marking on the noun, occasionally in Turkic, e.g., *köl tegin atï-sï* [Köl Tegin nephew-3SG.POSS] 'Köl Tegin's nephew', and generally in Tungusic, except in Manchu, e.g., Even *svinija ulrə-n* [swine meat-3SG.POSS] 'swine's meat, pork' (See F 13/14 in Robbeets, this volume: Chapter 10).

3.3.6 Shared innovations in Turkic

The Turkic languages share a number of specific innovations that set them apart from the Tungusic and the Mongolic languages. From a phonological perspective these include (1) the loss of voicing distinction *t-/d- and *k-/g- in word-initial position (e.g., Even deg- 'go up', WMo. degere 'higher than', OTk yeg 'better than'); (2) the denasalization of the initial nasals *n-/*m- into *y-/ *b- (e.g., Evk. murume 'round', WMo muru- 'be curved', OTk bür- 'wind round'), (3) the loss of final short vowels in verb stems (e.g., WMo. saɣa- 'to milk', OTk saɣ- 'to milk'); and (4) the development of RTR harmony into palatal harmony (see F1 in Robbeets, this volume: Chapter 10). Morphosyntactic innovations include (1) the development of the perfective nominalizer PTEA *-xA into a perfective non-past finite form in Turkic, while its reflex in Tungusic and Mongolic holds past finite meaning (Robbeets 2015: 396–416); (2) the replacement of the Proto-Altaic negative *ə- by a productive negative *ma- in Proto-Turkic (Robbeets 2015: 177–200); and (3) the development of double-marking in possessive noun phrases, while the other Transeurasian languages are generally dependent-marking (e.g. OTk ton-nuŋ bit-i (clothes-GEN louse-3SG.POSS) 'clothes' louse', see F14 in Robbeets, this volume: Chapter 10).

3.3.7 Shared innovations in Tungusic

One phonological innovation that sets Tungusic apart from the Turkic and Mongolic languages is the development of a velar nasal in word-initial position (see F3 in Robbeets, this volume: Chapter 10). Morphosyntactic innovations include (1) the development of a distinction between alienable and inalienable possession (e.g., Evenki dili-v (head-1SG.POSS) 'my head (sitting on my shoulders)' vs. dili-ngi-v

(head-ALN-1SG.POSS) 'my head (e.g., of an animal that I killed and have in my possession)'); (2) the development of suffixes to express the imperative rather than using the bare verb stem as in Mongolic and Turkic (see F9 in Robbeets, this volume: Chapter 10); (3) the extension of the number of semantic classes denoted by collective numeral suffixes, distinguishing between people and things (e.g., Evk. -*kt(e)* and -*ni* for counting people, Evk. -*gdA*/ -*ngnA* for counting objects, Evk. -*llA* for counting the number of days, Evk. -*nu* / -*pu* for counting the number of tents and Evk. -*musa* denoting the number of places or directions) and (5) the development of pluractional meaning on original causative suffixes (e.g., causative in Evk. *koyi-* 'to be confused (intr.)' -> *koyi-ci-* 'to deceive', but with pluractional meaning relating to he object in Evk. *bu:-* 'to give (tr.)' -> *bu:-t-* 'to distribute, give out, hand out (tr.)' and relating to the number of actions in Evk. *wa:-* 'kill (tr.)' -> *wa:-t-* 'massacre, slay (tr.)'; see Robbeets 2015: 276–290).

3.3.8 Resulting classification

Using shared innovations as our sole criterion, the primary split between Japano-Koreanic and Altaic is well supported, but it remains unclear whether Tungusic or Turkic was the first to split off within Altaic. As shown in Table 3.1, the support for a separate Mongolo-Tungusic branch is only slightly better than that for a Turko-Mongolic branch: I find 4 innovations in support of Mongolo-Tungusic and 7 in support of Turkic as a separate branch, while there are 3 innovations in support Turko-Mongolic and 6 in support of Tungusic as a separate branch. Thus, although the historical comparative method would favor the tree shown in Figure 3.6, Ramstedt's classification shown in Figure 3.5 is about as likely. In addition, especially

in the case of Turko-Mongolic, we cannot exclude that some of the features that we counted as shared innovations in reality are shared retentions, inherited from the Transeurasian ancestor. Besides, the shared innovations proposed above are based on my own understanding, while other linguists may propose more or different innovations. In sum, the approach taken here fails to give us an idea of the robustness of the different branches in the tree, it cannot unequivocally distinguish shared retentions from innovations and it is based on random "cherry picking" of supporting evidence. Therefore, in order to overcome these shortcomings, Bayesian analysis is a welcome addition to our methodological apparatus, which can help us to verify the historical comparative linguistic classification proposed in Figure 3.6.

Table 3.1 Number of potentially shared innovations in phonology and morphosyntax discussed here

	Phonological innovations	Morphosyntactic innovations	Total
Japano-Koreanic	4	5	9
Altaic	1	3	4
Turkic and Mongolic	1	2	3
Mongolic and Tungusic	0	4	4
Turkic and Tungusic	0	2	2

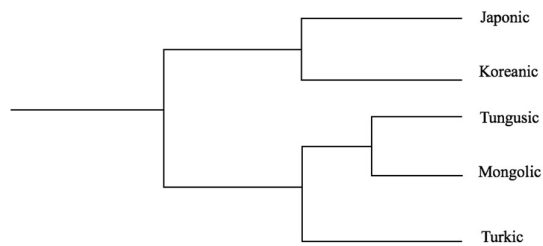| Turkic	4	3	7
Tungusic	1	5	6

Figure 3.6 The classification of Transeurasian according to the historical comparative method (Robbeets 2015: 506)

3.4 Classification on the basis of Bayesian inference

In a recent study with Remco Bouckaert, I applied Bayesian phylogenetic methods using BEAST to infer a phylogeny of the Transeurasian languages (Robbeets and Bouckaert 2018). Our dataset consisted in lexical etymologies supporting the reconstruction of Proto-Transeurasian forms with meanings that belong to the Leipzig-Jakarta 200 basic vocabulary list (see Robbeets, this volume: Chapter 36 for a full documentation of the first 100 items). Having coded our data for presence or absence of a cognate in each language, we tried to find a model of how these data

could change on a tree. Through a statistic procedure, we determined that the pseudo Dollo covarion model with relaxed clock was the most suitable evolutionary model for our data. When tracking the history of a cognate from the root of the tree to a leaf, the pseudo Dollo covarion model posits that a cognate can be gained multiple times, but that it can be lost only once. Contrary to the problematic assumption of a constant rate of language change in lexicostatistics, Bayesian methods can account for variation in the rate of change, using the notion of a 'molecular clock'. Our model thus implied that over time there is a certain rate of change along the branches of the tree. Given our model of cognate change and our basic vocabulary data, the Bayesian algorithm allowed us to compute a score for how well our data fit onto any given tree. In this way, we could score different trees according to how well they explained our data under the chosen model of cognate change. After running the computations millions of times, the analysis essentially sampled the best trees from the space of all possible trees. The resulting classification is shown in the DensiTree in Figure 3.7. The blue line traces the maximum clade credibility among the sample of best trees indicated in green. The numbers in the tree indicate the likelihood of the proposed branching: Japano-Koreanic has 98.7% support, Altaic 90.3% and Mongolo-Turkic 100%. In this way, the Bayesian analysis provides a quantitative basis to test the various competing hypotheses with regard to the internal structure of the Transeurasian family proposed above and selects Ramstedt's original proposal in Figure 3.5 as the classification with the highest credibility.
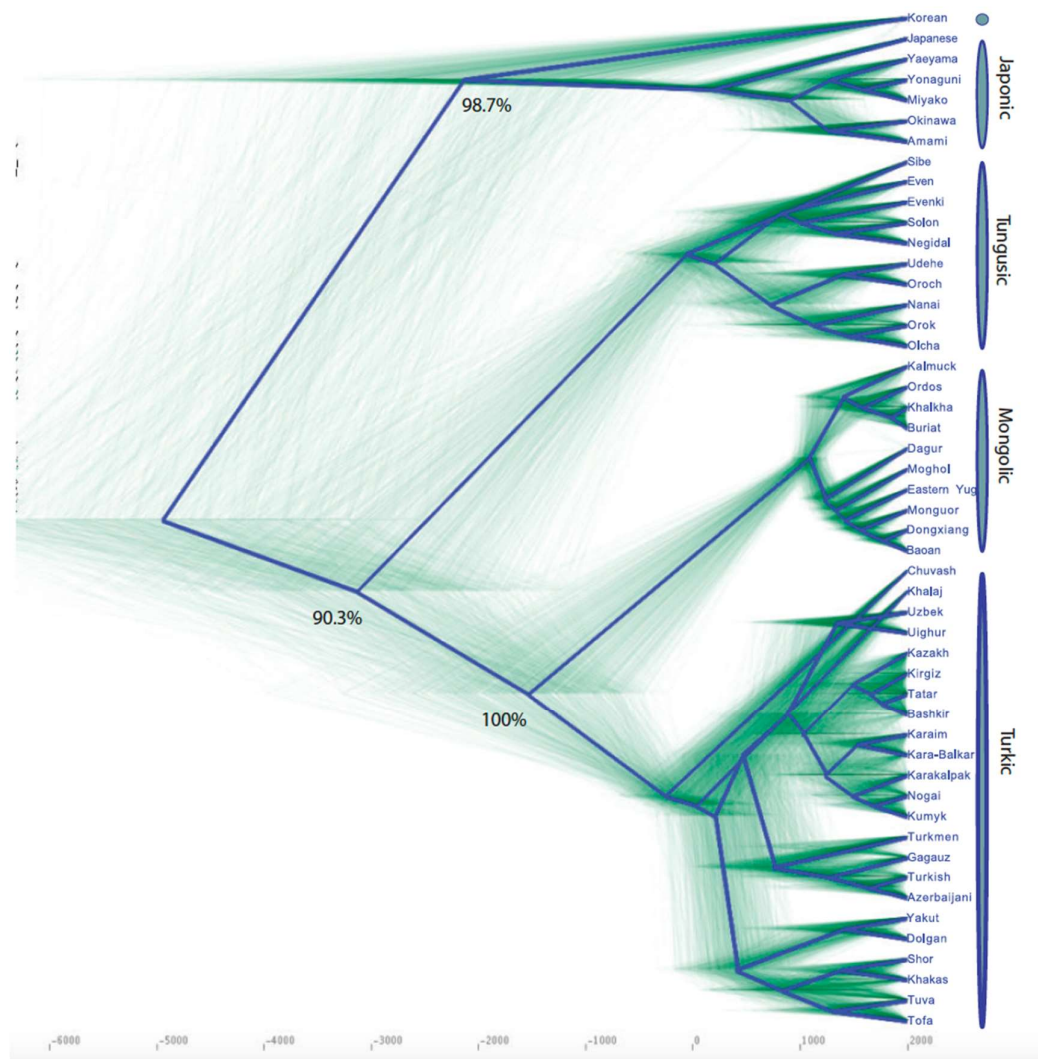
Figure 3.7 DensiTree of Transeurasian based on Bayesian analysis

3.5 Conclusion

While the hypothesis that the Transeurasian languages are genealogically related is gradually gaining acceptance in the field, supporters do not coincide on the internal structure of the family. In this chapter, I presented different classifications proposed for the Transeurasian family in the past and I suggested introducing Bayesian phylolinguistics as an additional tool for inferring and verifying phylogenies.

The main issue in the classifications so far has to do with the position of Tungusic vis-à-vis the other branches, more specifically whether it represents a separate branch in a polytopology, clusters with either the Japano-Koreanic or the Altaic branch and—if Altaic—whether it is most closely connected to Mongolic. Applying the classical historical comparative method, which is based on the principle of maximum parsimony, results in a tree in which Tungusic clusters with Mongolic, but leaves uncertainty whether the classification in which Mongolic clusters with Turkic should not be preferred. Introducing Bayesian phylolinguistics, we can quantify this uncertainty by indicating the likelihood of the proposed branching. We find that the likelihood that Tungusic splits off first in the Altaic branch is 90,3% and that the probability that Mongolic closely clusters with Turkic is near to 100%. In this way, Bayesian phylolinguistics can give us an idea about the robustness of the branches. It also avoids misinterpreting shared retentions as shared innovations and cherry picking particular shared innovations to support our preconceived ideas of classification, dangers of applying the classical comparative linguistic method of classification.

As a result, the best supportable tree for the Transeurasian family is the one in which Tungusic is the first to split off within the Altaic branch. This outcome seems to keep pace with recent developments in interdisciplinary research, related to the genetic and archaeological past. In Chapter 50, Jeong, Wang and Ning indicate that contemporary speakers of Tungusic languages are genetically continuous with ancient individuals of the Neolithic Boisman culture (4825–2470 BC) in the Southern Primorye. Contemporary Turkic and Mongolic speakers share this ancestral eastern lineage, albeit with an increasing western admixture from the Bronze Age onwards. In addition, in Chapter 49, Tao Li shows that agriculture was adopted in the Southern

Primorye from the Liao River region during the Neolithic, while Mark Hudson, in chapter 51, argues that pastoralism was gradually adopted in the homelands of Turkic and Mongolic speakers west of the Liao River and on the Eastern steppes. Thus, genetics and archaeology converge on a scenario whereby the linguistic ancestors of the Tungusic speakers separated at an early time from the linguistic ancestors of Turkic and Mongolic speakers, who only from the Bronze Age onwards started to share an increasing degree of genetic admixture with Western steppe herders and to adopt a common pastoralist subsistence strategy. As such, the classification proposed over 100 years ago by Gustav Ramstedt, the pioneer of Transeurasian linguistics, can be confirmed.

Acknowledgement