

## Chapter 10

### **The typological heritage of the Transeurasian languages**

Martine Robbeets

#### **Abstract**

This chapter addresses the question of whether some structural features shared by the Transeurasian languages can be explained by inheritance. For this purpose, I first establish 20 structural features shared by the Transeurasian languages that are shared to a lesser extent by neighboring non-Transeurasian languages. Next, I propose a number of diachronic typological criteria, which indicate that the Transeurasian languages can be characterized as a “typological heritage”, even if this still leaves room for the concepts of “diffusion” and “language area”.

**Keywords:** typological heritage, diachronic typology, language area, diffusion, inheritance, Transeurasian, Japanese, Korean, Tungusic, Mongolic, Turkic

#### **10.1 Introduction**

In the present chapter, I will suggest a set of typological features that delimit the Transeurasian languages vis-à-vis the neighboring non-Transeurasian languages and I will argue that some of these features are more easily explainable by inheritance from a common source than by areal diffusion. This view goes against a common belief among the opponents of Transeurasian affiliation, notably “that speaking of ‘Altaic’ instead of ‘Ural-Altaic’ is a misconception, for there are no

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.

areal or typological features that would be specific only to ‘Altaic’ without Uralic” (Janhunen 2014b: 13).

In spite of the strong polarization in the Transeurasian field between so-called “retentionists”, who view the similarities as arising from common descent, and “diffusionists”, who view them as arising from areal interaction, detailed characterizations of Transeurasian as a “typological heritage” or a “linguistic area” are surprisingly rare in the linguistic literature.

Poppe (1964b) analyzed Altaic as a “language type” on the basis of a list of structural parallels shared between Korean, Tungusic, Mongolic and Turkic languages and Rickmeyer (1989) elaborated on this research, adding data from Japanese. Even if these contributions provided an impressive list of shared features, they do not strictly identify Transeurasian as a language area or a typological heritage because they do not delimit the language type in relation to its non-Transeurasian neighbors and neither do they consider the extent to which the features in question are common or rare across the world as a whole or attempt to distinguish contact-induced from genealogically motivated features.

More recently, Anderson (2006) proposed to consider Siberia as a linguistic area, cross-cutting the genealogical diversity exhibited among the different indigenous Siberian languages, which belong to a variety of language families such as Uralic, Altaic, Amuric, Yeniseic, Eskimo-Aleut and Chukotko-Kamchatkan. His idea has been called into question by Comrie (2013), who pointed out that, in contrast with the languages of the Southeast Asia linguistic area that share up to 82% of WALS features, the languages of the Siberian area only share 40–50% of

these features. In line with Comrie's concern, Wichmann (forthcoming) found that the statistical support in a random sample of 107 features is not sufficiently significant to delimit the Siberian from the non-Siberian languages and therefore concluded that Siberia is not a linguistic area.

The term “typological heritage” implies a set of structural features shared by a group of genealogically related languages due to inheritance from a common ancestor, whereby the features are significantly less shared by languages belonging to other language families, while a “linguistic area” is a geographical region in which neighboring languages show a set of structural features in common due to language contact and where the features are significantly less shared by languages immediately outside the area; see Campbell (2006) for various similar definitions proposed for “linguistic area”. For both concepts we first need to establish a set of common structural properties and delimit that set from neighboring and/or demonstrably unrelated languages. Only in the next step, we must consider the way in which the languages may have come to share these features, considering areal diffusion, genealogical relationship or an interaction of both factors as possible explanations. Therefore, I will first try to establish a concentration of linguistic features shared more intensely between the Transeurasian languages than between Transeurasian and non-Transeurasian languages, independent of how these features developed historically. Only after evaluating 20 structural features especially shared across the Transeurasian languages, will I consider how the insights from the data are relevant for historical statements about the distinction between “linguistic area” and “typological heritage”.

The organization of this chapter is as follows. In Section 10.2, I will introduce the dataset underlying this study and discuss the methods used in defining a set of typological features specific to Transeurasian and in distinguishing inherited from diffused features. In Section 10.3, I will set up a typological profile of the Transeurasian languages in relation to that of the selected languages outside the proposed family and/or area. The linguistic levels discussed will include phonology, lexicon and semantics, morphology and syntax. In Section 10.4, I will show that even if we agree with the massive diffusion of linguistic features across the Transeurasian languages, there is reason to assume that a part of the shared features constitutes a structural heritage rather than a linguistic area. Finally, I will conclude this chapter in Section 10.5.

Some of the data underlying this paper are drawn from Robbeets (2017g/h), but the approach taken here differs from these papers in that it is focused on delimiting the Transeurasian languages in relation to their non-Transeurasian neighbors and on distinguishing between “linguistic area” and “typological heritage”.

## **10.2 Dataset and methods**

In this study, I will use a dataset of 17 languages as horizontal comparison points and 20 selected structural features as vertical comparison points. The presence of a given feature in a particular language is indicated with plus (+), absence with minus (-) and uncertainty with (+/-). The overall aim is to add up the number of plus values for each language and to calculate the percentage of prototypical structures present in that language. This will enable us to draw inferences about the

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

typological coherence of the Transeurasian languages vis-à-vis neighboring non-Transeurasian languages.

The linguistic levels discussed will include phonology, lexicon and semantics, morphology and syntax. I treat grammaticalization patterns as a distinct level of analysis because, rather than representing a static feature value, they are concerned with a dynamic force, leading in languages to change from a less to a more grammatical status.

The features are chosen to maximize positive values for the Transeurasian languages and to minimize such values in the neighboring non-Transeurasian languages. Selecting the most desirable features with regard to the delimitation of the Transeurasian languages, my approach could be accused of “cherry-picking”, i.e. of focussing on just the features that confirm the idea. However, for the purpose of the present study, this is legitimate because the challenge precisely is to gather features that are uniquely shared across the Transeurasian languages. Moreover, a concentration of features is a requirement for both definitions, “linguistic area” as well as “typological heritage”. Since the diffusion of linguistic features across the Transeurasian languages is not denied here, but the basic question rather amounts to whether it is sensible to assume the inheritance of features in addition to diffusion, it is legitimate to focus on features that are potentially diagnostic of inheritance. Therefore, I start from a selection of inherently Transeurasian typological features and let appropriate linguistic criteria decide whether there is support for the idea of “typological heritage” or not.

As representatives of the contemporary varieties of Transeurasian, I will use Turkish (Turkic), Khalkha Mongolian (Mongolic), Evenki (Tungusic), Korean (Koreanic) and Japanese (Japonic). For retrieving linguistic data underlying the feature values, I consulted Göksel and Kerslake (2005) for Turkish; Janhunen (2012a) for Khalkha Mongolian; Bulatova and Grenoble (1999) and Nedjalkov (1997) for Evenki; Martin (1992) and Sohn (1994) for Korean, Martin (1988), Kaiser et al. (2001) and Iwasaki (2013) for Japanese. To allow for a diachronic perspective, I supplemented the contemporary languages with the oldest reliable historical varieties, consulting Erdal (2004) for Old Turkic, Street (1957), Weiers (1966) and Rybatzki (2003b) for Middle Mongolian, Poppe (1954) for Written Mongolian, Gorelova (2002) for Manchu, Martin (1992) and Lee and Ramsey (2011) for Middle Korean, and Vovin (2005, 2009a) and Frellesvig (2010) for Old Japanese. In order to delimit external boundaries, I have included Ainu (Ainuic), Nivkh (Amuric) and Rukai (Austronesian) as adjacent languages to the east, Mandarin Chinese (Sino-Tibetan) to the south, and Kolyma Yukaghir (Yukaghiric), Ket (Yeniseian) and Eastern Khanty (Uralic) to the north; see Figure 10.1. For retrieving linguistic data underlying the feature values, I use Gruzdeva (1998) for Nivkh; Maslova (2003a) for Kolyma Yukaghir; Werner (1997), Vajda (2004) and Georg (2007a) for Ket; Filchenko (2007) for Eastern Khanty; Li and Thompson (1989) for Mandarin; Zeitoun (2007) for Mantaoran Rukai; and Shibatani (1990) and Tamura (2000) for Ainu.

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.



Figure 10.1 The distribution of the contemporary Transeurasian languages and neighboring non-Transeurasian languages included in the sample (generated with WALS tools)

Structural similarity is traditionally considered inconclusive evidence of genealogical relatedness because it is difficult to distinguish areal diffusion, universal tendencies and coincidence from inherited correlations (Meillet 1948; Hock 1991: 561; Campbell and Poser 2008: 309–316). Nevertheless, recently some linguists (a.o. Bickel 1999, Johanson 2002b, Nichols 2003, Dunn et al. 2005, Wichmann and Holman 2009, Greenhill et al. 2017) have argued that it is possible to identify a core of stable structural features that is particularly resistant to change, loss or borrowing.

In Section 10.4, I will rely on a set of diagnostic criteria to examine whether the entire set of shared features identified in Section 10.3 is indicative of a linguistic area, or whether structural heritage may be involved as well. Indications that a set

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

of shared structural features is more easily explainable by inheritance from a common source than by areal diffusion may come from (i) geography, i.e. the absence of the features in neighboring unrelated languages in spite of their geographical proximity vs. the sharing of features within the language group in spite of the geographical remoteness of certain members; (ii) history, i.e. historical varieties of the language group display more structural cohesion than contemporary varieties within the same group; (iii) distribution: the features are better distributed in peripheral languages of the language grouping than in central languages; (iv) areal dominance, i.e. prototypical features of the language grouping are dominant in family-external contact situations; (v) recurrent grammaticalization, i.e. prototypical features of the language grouping are the result of grammaticalization processes, that are recurrent within the grouping but rare outside the grouping and (vi) isomorphism; i.e. a prototypical feature is shared in addition to a formal correspondence of the particular marker reflecting that feature.

Complementing these qualitative criteria based on diachronic typology, in the next chapter Hübler (this volume: Chapter 11) takes an alternative quantitative approach to the question of whether some Transeurasian structural features may be inherited. To this end, she applies Bayesian inference methods to synchronic data, capturing the tree that best explains the correlations between the structural features and she shows that the concrete shape of the tree is more suggestive of inheritance than of areal influence. Even if her analysis of purely synchronic data at times yields different values in comparison with the diachronic approach applied here, the overall outcome of her study confirms the findings in this chapter.



### 10.3 Typological profile of the Transeurasian languages

#### 10.3.1 Phonology

##### (i) F1 Presence of tongue root vowel harmony

The most general harmony phenomenon in Turkic is palatal harmony, which prompts all vowels within a domain to be exclusively front or back (e.g. Tk. *ip-ler* [rope-PL] ‘ropes’ vs. *pul-lar* [stamp-PL] ‘stamps’); see Johanson (1993). Palatal harmony is also found in most Uralic languages, such as in Khanty. Since the western Mongolic languages Oirat and Kalmuck display palatal harmony as well, it has been proposed that the original system of Mongolic harmony was palatal (Poppe 1955, Svantesson 1985).<sup>1</sup> However Ko (2012) demonstrated that the original vowel harmony in Mongolic was in fact based on the opposition between the advanced vs. retracted position of the tongue root, rather than on a palatal contrast. He argued that the tongue root retraction system in Khalkha (e.g. *od-o:s* (feather-ABL) vs. *ɔd-ɔ:s* (star-ABL)) represents retention rather than innovation. Furthermore, he supported the view that Tungusic vowel harmony is RTR based, as it is in Manchu and Evenki, and that the reduced vowel harmony in contemporary Korean derives from a tongue-root based system in Middle Korean. Contemporary Japanese and Ryukyuan languages do not have vowel harmony.

In Old Japanese, however, there is a restriction on the shape of root morphemes, whereby the vowel *o*<sub>2</sub> cannot occur in a root together with the vowels *u*, *o*<sub>1</sub> or *a*. This phenomenon, known as Arisaka’s law, has been taken as a kind of vowel harmony, but it has been excluded from comparisons with other Transeurasian

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

languages because it does not reflect palatal harmony, the type of harmony which was attributed to the Transeurasian languages until recently (Frellesvig 2010: 44). In the light of the 7-vowel system reconstructed for Proto-Japonic (Frellesvig and Whitman 2008), the harmony-like opposition in Old Japanese implies an underlying opposition between PJ *\*ə*, *\*i* and *\*a*, *\*u*, *\*o*, in which an RTR based contrast seems to be preserved for PJ *\*ə* ~ *a*; see Robbeets 2015: 124–126. As Joseph et al. (this volume: Chapter 29) suggest, the attrition of RTR harmony systems commonly starts with the loss of the contrast for high vowels in the direction of the non-retracted alternant, which may explain why the contrast in Proto-Japonic is only detectable for the low vowel. In addition, petrified verb suffixes, close to the verb root, leave traces of an RTR harmony-like alternation, whereby PJ *\*ə* occasionally harmonizes with PJ *\*u* and PJ *\*ə* in the root, while PJ *\*a* is the default suffix vowel (Robbeets 2015); e.g. OJ *udunop-* ‘to prize, value’ < PJ *\*untu-nə-pə-* (treasure-VBLZ-ITER), OJ *susurop-* ‘slurp’ < *\*susu-rə-pə-* (SOUND-VBLZ-ITER), OJ *urupop-* ‘get damp’ < *\*uru-pə-pə-* (*be.wet*-FIEN-ITER), OJ *noꝛoꝛp-* ‘to curse’ < PJ *\*nə-rə-pə-* (SOUND-VBLZ-ITER); see Joseph et al. (this volume: Chapter 29) for other Old Japanese alternations in support of this interpretation.

Given traces of RTR harmony in four out of five Transeurasian branches, it appears that RTR harmony may have been a prototypical structure of Proto-Transeurasian, with Proto-Turkic perhaps having shifted to a palatal contrast under the influence of the neighboring Uralic languages. The vowel reconstructions for

Proto-Transeurasian vowels (Robbeets, this volume: Chapter 36) suggest an original opposition between non-retracted PTEA *\*ə, o, u* and retracted *\*a, \*ɔ, \*ɔ̄*.

Whereas Vovin (1993a: 50–51) and Bugaeva (2015: 465–467) reconstruct palatal harmony in Ainu, Shibatani (1990: 15) speculates that the Ainu opposition between *o* and *u, a* might have its origin in tongue root harmony, but here the indications are even weaker than in the Japanese case. According to Maslova (2003a: 35), Yukaghir might be more appropriately described as having tongue root harmony than palatal harmony. Chukchi also displays tongue-root harmony.

Although Gruzdeva (1998: 10) suggests that Nivkh leaves traces of height harmony, Janhunen (1981b) and Joseph et al. (2014, this volume: Chapter 29) interpret the system in terms of tongue root harmony.<sup>2</sup> Since RTR harmony extends beyond Eurasia into North America (e.g. in Eskimo-Aleut, Sahaptian, and Salishan), Joseph et al. assume that it must be reconstructed for many proto-languages of unrelated families in North East Asia and that its spread must be ancient. Therefore, RTR harmony may have spread from and into Proto-Transeurasian preceding the primary split of the family. Cross-linguistically, tongue root harmony seems to be rather rare (Ko 2012: 11–12). A rough estimate would be that less than 10% of the world's languages have a tongue root vowel harmony system.

(ii) F2 Absence of *r-* in initial position in native words

Across the Transeurasian languages, the consonant *r-* is not allowed to occur word-initially, except in copies (e.g. J *rajio*, K *latiwo*, Even *radio*, Khal. *radio*, Tk. *radyo*

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.

‘radio’); see also Hübler (this volume: Chapter 11) for the coding of this feature in a broader sample of Transeurasian languages. The feature is also true for Kolyma Yukaghir. Ket lacks a phoneme /r/ altogether. Although initial \*r- is not reconstructed for Proto-Uralic, Khanty is atypical in this sense, e.g. *rayta* ‘to drop, slide’ and *räy* ‘garbage’. Nivkh, Ainu, Mandarin and Rukai also have native words in initial r-. Outside Uralic and Transeurasian, a fair number of languages such as Proto-Indo-European, Basque, some Melanesian and Caucasian languages, Efik (Niger-Congo), Susu, Diyari (Australian), and Piro (Arakawan) lack r- in word initial position. A rough estimate would be that less than 30% of the world’s languages lack r- in initial position.

### (iii) F3 Absence of velar nasal restricted to initial position

In most Turkic languages, as well as in Mongolic languages and Korean, the velar nasal *ŋ*- cannot appear in word-initial position. Contemporary Japanese and Ainu lack a velar nasal phoneme all together, but Unger (2008b) and Francis-Ratte (2016: 29) argue for the reconstruction of non-initial \**ŋ* in pre-Old Japanese, among other reasons pointing to the allophony of the voiced stop *g* with *ŋ* in contemporary Japanese dialects, e.g. the pronunciation of Tokyo dialect *kagi* ‘key’ as [kaŋi]. In the Tungusic languages, with exception of Manchu, however, *ŋ*- can appear word-initially, but generally it is restricted to a specific phonological environment, notably when it is followed by the sonorants *n*, *r*, *l*, *m*, *y*, e.g. Evk. *ŋene-* ‘to go’, Ma. *genu-* ‘to go together’, Evk. *ŋe:le-*, Ma. *gele-* ‘to fear’.

According to Poppe (1964b: 4) the initial velar nasal in Tungusic is the result of

secondary assimilation of PTg \*g-, which implies that originally \*ŋ- was absent in Tungusic as well. The assimilation was probably triggered by influence from languages in the Siberian area, such as Nivkh, which allow initial velar nasals (Anderson 2006). It is under the same influence that initial *ŋ* became allowed in Dolgan (Turkic), e.g. *ŋassa* ‘pipe’. Khanty, Ket, Kolyma Yukaghir and Mandarin have a velar nasal, but it does not occur in word-initial position. Rukai allows an initial velar nasal, e.g. *ŋa|ai* ‘saliva’. In Anderson’s (2005: 42) sample of 468 languages, 69% lack an initial velar nasal. Among the languages of the world that have a velar nasal phoneme, as is the case with most Transeurasian languages, only 35% do not use it in word-initial position.

(iv) F4 Presence of voicing distinction for stops

Turkic, Mongolic and Tungusic languages share a voiced-voiceless opposition for stops, and a voicing distinction can be reconstructed for Proto-Transeurasian. In Contemporary and Middle Korean, stops display an opposition between lax (*p*), aspirated (*ph*) and tensed (*p'*). Even if the lax stops become lightly voiced between voiced sounds, there is no phonemic voicing distinction. The Japanese and Ryukyuan voicing distinction for stops is a secondary development, as voiced stops derive from prenasalized voiceless stops. Therefore, the ancestor of Japanese lacked voicing distinction. Khanty lacks voicing distinction for stops, a feature characteristic of Proto-Uralic, although many contemporary Uralic languages have developed an original singleton-geminate contrast into a voicing distinction. Ket, Yukaghir and Nivkh display a voicing distinction, but Ainu and Chukchi do not. Mandarin,

has a distinction between aspirated and unaspirated stops, but lacks a voiced-voiceless opposition. Characteristic of most Austronesian languages, Rukai also displays voicing distinction for stops. In Maddieson's (2005: 24) sample of 566 languages, 61% display a voicing distinction for stops.

### 10.3.2 Lexicon and semantics

#### (i) F5 Preference for non-verbal strategy for (extra-family) verbal copies

As far as the mechanisms of loan verb accommodation are concerned, most recipient languages can be categorized into two distinct groups: one where copied verbs arrive as verbs needing no formal accommodation, and one where they arrive as non-verbs and need formal accommodation (Wohlgemuth 2009). Most Transeurasian languages can be assigned to the second category because their loan verbs need formal accommodation by a suffix or a light verb; e.g. Tk. *klik-le-* ~ *klik-et-* << English *click* (with a mouse), Khal. *zee-l-* << Mandarin *zhài* 'borrow, lend'; Ud. *tancewa-la-* << Russian *tancewa-t'* 'to dance'; K *coking ha-*, J *zyogingu suru* 'to jog' << English *jog*; J *demo-r-* << English *demonstrate*. The northern Tungusic languages, however, prefer to borrow verbs through direct insertion, e.g. Evk. *vypolñaj-* << Russian *vypolnja-t'* 'to fulfill, carry out'. Since we have no information about verb borrowing in the historical stages, I mark them with +/- . In contrast to the Transeurasian languages, Ainu, Sinitic languages such as Mandarin, Uralic languages such as Khanty, and Austronesian languages such as Rukai show a strong preference for direct insertion (Wohlgemuth 2009: 158, 161; Tamura 2000: 267). Yukaghir and Nivkh did not integrate any recognizable verbal borrowings

from Russian or other foreign languages into their lexicons. In Wohlgenuth's (2009: 157) sample, 55% of languages worldwide are found to use direct insertion, while the remainder prefer non-verbal strategies such as indirect insertion and the light verb strategy.

(ii) F6 Presence of a two-way proximal-distal distinction in demonstrative pronouns

Although Old Turkic displays a two-way distinction in its demonstratives, i.e. OTk *bol/bun-* 'this' vs. *ol/an-* 'that', many contemporary Turkic languages such as Turkish make a three-way distinction, e.g. Tk. *bu* 'this', *şu* 'that', *o* 'that (over there)'. Demonstrative pronouns in earlier and contemporary varieties of Mongolic and Tungusic exhibit a proximal-distal distinction: MMo. *ene* 'this' vs. *tere* 'that', Khal. *e-* 'this' vs. *te-* 'that', Ma. *ere* 'this' vs. *tere* 'that' and Evk. *er(i)* 'this' vs. *tar(i)* 'that'. Demonstrative pronouns in Contemporary and Middle Korean, however, show a proximal-mesial-distal opposition: K *i* 'this', *ku* 'that', *ce* 'that over there' and MK *i* 'this', *ku* 'that', *tye* 'that over there'. This is also true for Contemporary Japanese: J *ko-* 'this', *so-* 'that', *a-* 'that over there'. In contrast to most accounts of Old Japanese demonstratives, which posit a three-way contrast between OJ *ko<sub>2</sub>* 'this', *so<sub>2</sub>* 'that' and *ka* 'that over there', Frellesvig (2010: 139–142) argued that OJ *ka* was not a productive member of the demonstrative system and that pre-Old Japanese had a simple proximal-distal distinction. Similarly, in Yaeyama Ryukyuan the opposition between demonstratives is restricted to proximal *kuri* 'this' vs. distal *uri* 'that' (Aso 2015: 429). While Khanty distinguishes between proximal *timi* 'this (here)' and distal *tomi* 'that (there)',

Yukaghir, Ket and Ainu have a three-way opposition, with each demonstrative pronoun denoting a different degree of proximity: Yukaghir *tij* ‘this’ (proximal), *adiŋ ~ ediŋ* ‘that’ (mesial), *taŋ* ‘that’ (distal); Ket *tu-* ‘this, that’ (neutral), *ki-* ‘this, that’ (proximal), *qa-* ‘this, that’ (distal), and Ainu *ta an* ‘this’ (distal), *ne an* ‘that’ (mesial), *to an okai* ‘that over there’ (distal). Nivkh makes as many as five distinctions: *tyd’* ‘this’ (near and visible), *hyd’* ‘this, that’ (distant), *ad’* ‘that’ (more distant and visible), *aixnt* ‘that’ (most distant), *kud’* ‘that’ (absent).<sup>3</sup> Rukai distinguishes four demonstrative pronouns in terms of visibility and distance: ‘*ina* ‘this’ (proximal), *ana* ‘that’ (mesial), *ona* ‘that over there’ (distal but visible), *dhona* ‘that over there’ (distal and invisible). Mandarin has a two-way distinction between proximal *zhè(ge)* ‘this’ and distal *nà(ge)* ‘that’, which developed from a three way-distinction in Classical Chinese between neutral, proximal and distal. In Diessel’s (2005: 170–173) sample of 234 languages, 54% exhibit a two-way distance contrast in demonstratives, while 38% exhibit a three-way contrast.

(iii) F7 Property words are verbally and nominally encoded such that some property words exhibit switched encoding

Cross-linguistically, adjectives have no prototypical encoding strategy of their own; they will align themselves either with verbs or with nominals. The large majority of property words in the contemporary Turkic languages are nominally encoded.

Originally, in Proto-Turkic, the encoding of property words appears to have been mixed because, at least in Old Turkic, both the nominal and the verbal strategy is used. There seems to be a tendency to apply the verbal strategy in the case of time-



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.

unstable properties such as OTk *bädü-* ‘to be(come) big, great’, OTk *isi-* ‘to be hot’, OTk *kat-* ‘to be hard, firm, tough’, OTk *kiz-* ‘to be red’, OTk *tumli-* ‘to be cold’, OTk *tinči-* ‘to be(come) putrid, smell foul’, OTk *us-* ‘to be thirsty’, OTk *yeni-* ‘to be(come) light’, OTk *tigra-* ‘to be tough’, OTk *iglä-* ‘to be(come) ill’. Contemporary Turkic languages maintain a few reflexes of these verbal property words, for instance for ‘to be(come) big’, Tk. *büyü-*, SUig. *pezi-*, Az. *böyü-*, Khalaj *bidi-*, Tuva *bedi-*, Gag. *bü:-*, Karaim *büyü-* and for ‘to be(come) red, red-hot’, Turkm. *Giz-*, Tur. *kiz-*, Yak. *kī:s-*, but in the majority of cases, the earlier verbal property word has been derived with a deverbal noun suffix and became lexicalized as a nominal property word, e.g. OTk *bädük* ‘big, great; greatness’ and Tk. *büyük* ‘big’; see also Johanson (2006).

As in Turkic, most property words in Mongolic and Tungusic languages are nominally encoded, but some are verbally encoded (e.g. WMo. *qala-* ‘to be(come) warm’; Khal. *ayu:-* ‘to be afraid’, Ma *aka-* ‘to be sad’, Evk. *buli:-* ‘to be sad’). In Japanese and Korean many property words are verbally encoded, but others are nominally encoded (J *sizuka*, OJ *siduka* ‘quiet’, J/OJ *taka-* ‘to be high’, K *kippu-* ‘to be happy’, *phikon ha-* ‘be tired’). In line with most Uralic languages, property words in Khanty are exclusively nominally encoded. This is also true for Ket. In Yukaghir, Ainu and Nivkh, however, property words are exclusively verbally encoded. As in the case of most Transeurasian languages, Ainu property verbs express both the property and the process leading to the property, e.g. *pirka* ‘to be(come) good’. In line with Mainland Southeast Asian and Austronesian languages, Mandarin and Rukai use verbal encodings for property words.

Generally, the mixed encoding of adjectives in the Transeurasian languages is split in the sense that most property words have only a single encoding option. Turkish *güzel* ‘beautiful’, for instance, has nominal encoding and cannot be inflected as a verb. However, in Old Turkic, some doublets such as OTk *ač* ‘hungry’ / *ač-* ‘to be hungry’, OTk *keč* ‘late, slow’ / *keč-* ‘to be late, slow’, OTk *köp* ‘abundant’ / OTk *köp-* ‘to swell, boil over’ / OTk. *kari* ‘old’ and *kari-* ‘to become old’ exhibit traces of original switching, whereby the same property word can have both nominal and verbal encoding; see Doerfer’s (1982: 104–112) list of so-called ‘Nomenverba’. Similar traces of switching are found in the other Transeurasian languages, especially in the earlier varieties, e.g. MMo. *bulqa* ‘hostile; hostility’ / *bulqa-* ‘to be hostile’; Ma. *jalu* ‘full’ / *jalu-* ‘to be full’, Ma. *sula* ‘loose, free’ / *sula-* ‘to be loose, be free’; MK *toso-* vs. MK *toso ho-* ‘to be warm’; OJ *taka* ‘high’ / *taka-* ‘to be high’, OJ *opo* ‘big’ / OJ *opo-* ‘to be big’. None of the neighboring languages exhibits such behavior. In Stassen’s (2005b: 478–481) sample of 386 languages, 27% have mixed encoding in predicative adjectives. Therefore, the proportion of languages exhibiting mixed and switched encoding will be lower.

### 10.3.3 Morphology

(i) F8 Inflectional morphology is predominantly suffixing

Bound units in Transeurasian languages are postponed; i.e. they are suffixes rather than prefixes or infixes. Across the strongly suffixing Transeurasian languages, prefixation is rare and is restricted to derivational morphology, such as partial emphatic reduplication in Altaic (e.g. Tk. *bem-beyaz* ‘snow white’, *up-uzun*

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.

‘extremely long’, Otk *kap-kara* ‘quite black’, Khal. *xob-xoldu*: ‘frozen through’, WMo. *ub-ulayan* ‘completely red’, Evk. *ab-aya* ‘very good’; Whaley and Li 2000) and some derivational prefixes in Korean (e.g. K *yel-* ‘young, new’ in *yel-cwungi* ‘a chick out of its shell’) and in Japanese (e.g. *ma-* intensive in *ma-siro* ‘snow white’). As is the case for most Uralic languages, Khanty is strongly suffixing, as is Yukaghir. Nivkh is considered to be weakly suffixing. In Ket, nominal inflectional morphology is strongly suffixing, whereas verb inflection is predominantly prefixing. In Ainu and Rukai, inflection makes use of both prefixes and suffixes. Probably due to Transeurasian influence, Mandarin is hard to assign unequivocally to either the isolating or weakly suffixing type, but Sinitic varieties in general tend towards the isolating pole. In Dryer’s (2005a: 110–113) sample of 894 languages, 43% are strongly suffixing.

(ii) F9 The imperative is expressed by a bare verb stem

Across the Transeurasian languages, imperatives can be built with special endings, such as desiderative, optative, volitional or exclamatory suffixes, but most languages seem to share an original imperative that is formed on the basis of the bare verb stem alone. In most Turkic languages, the absence of verb suffixes indicates that the form is to be understood as an imperative, e.g. Otk. *kel* ‘come (here)!', *kèt* ‘go!', Tk. *gel* ‘come (here)!' . The imperative in the Mongolic languages is expressed by a bare verb root or stem, e.g. Khal. *ir*, WMo. *ire* ‘come!', Khal. *soo*, WMo. *sayu* ‘sit down!'. Among the Tungusic languages, only Manchu and Sibe have preserved a zero imperative, e.g., Ma. *gene* ‘go!', *te* ‘sit down!', *ala*

‘tell!’, the other languages use imperative suffixes, e.g. Evk. *eme-kel* (come-2SG.IMP) ‘come!’.

In Contemporary Japanese, plain imperatives are commonly expressed by a suffix *-e* following consonant-final stems and *-ro* following vowel-final stems, e.g., J *ik-e* (go-IMP) ‘go!’, *mi-ro* (see-IMP) ‘look!’. However, in Old Japanese the vowel-final stems simply use the bare verb stem, e.g., OJ *ko<sub>2</sub>* ‘come!’, *ake<sub>2</sub>* ‘open!’, *mi* ‘look!’, whereas consonant-final stems add a suffix OJ *-e<sub>1</sub>* to form imperative form, e.g. OJ *kak-e<sub>1</sub>* ‘write!’, *sin-e* ‘die!’, *ar-e* ‘exist!’. Most zero imperatives can be reinforced by the exclamatory particle OJ *yo<sub>2</sub>*, e.g., OJ *ake<sub>2</sub> yo<sub>2</sub>* ‘open!’, *mi yo<sub>2</sub>* ‘look!’, *oki<sub>2</sub> yo<sub>2</sub>* ‘arise!’, which was in the course of changing to an inflectional ending marking of the imperative itself (Vovin 2009a: 655; Frellesvig 2010: 101). According to Vovin (2009a: 647), the suffix OJ *-e<sub>1</sub>* can be derived as an instance of monophthongization of the converb suffix PJ *\*-i* plus an auxiliary *\*a* in the imperative. It is, therefore, inviting to identify PJ *\*a* as the bare imperative form of the copula PJ *\*a-* ‘to be’. As such, bare verb stems seem to represent the original strategy to form imperatives in Japanese.

The Korean imperative ending in the plain style is K *-e/a*, MK *-e/a* (e.g. K *mek-e* ‘eat!’). In addition to the observation that Middle Korean verbs unexpectedly retain their accent before the ending *-e/a*, violations of vowel harmony in favor of the shape *-a* in early texts and dialects suggest that the suffix should be derived from an original auxiliary, most probably the bare stem of the copula PK *\*a-* (Martin 1992: 70, Robbeets 2015: 154–157). This formation parallels that of the imperatives derived from Old Japanese consonant final stems.

In Khanty, Nivkh and Rukai the imperative is expressed by attaching a suffix to the verb base, e.g., Khanty *qolənt-a* (listen-IMP.2SG) ‘listen!’, Nivkh *marid ye-ja* (second take-IMP) ‘Take the second [dog]!’ and Rukai *vilivili-a* (pull-IMP) ‘pull!’, while Ket uses a prefix, e.g. *úl-d-al-gay* (water-ACC-3.M.OBJ-IMP-wash) ‘wash him!’. In Yukaghir, the second person imperative contains a zero suffix followed by the regular second person marker, e.g., *wie-Ø-k* (make-IMP-2) ‘make!’. In Ainu, both a bare stem and a free-standing imperative marker are used, e.g. *nu* ‘listen!’, *nu yan* (hear IMP.PL/POL) ‘listen!’. In Mandarin, imperatives are formed with bare verb stems and a second person marker, which is often omitted.

Imperatives always imply straightforward commands to second person and are cross-linguistically commonly expressed by the bare root, or stem of the verb.

### (iii) F10 Absence of obligatory numeral classifiers

Although in Turkic and Mongolic some nouns of low countability may be accompanied by a unit of measure by means of which they can be counted, e.g. Tk. *sekiz bardak su* (8 glass water) ‘eight glasses of water’, OTk *yeti tutum talkan* (7 handful parched.grain) ‘seven handfuls of parched grain’, Khal. *gourben debter nom* (3 volume book) ‘3 volumes of books’, these languages do not make use of sortal numeral classifiers. The same is true for the Tungusic languages, except Manchu. Under Chinese influence, Manchu has developed about 70 sortal numeral classifiers, such as *fesin*, which is used for objects equipped with a handle, e.g. *ilan fesin loho* (3 CLF sword) ‘three swords’. However, the use of these classifiers is not obligatory in Manchu. *Loho ilan* (sword 3) ‘three swords’, for instance, is

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

equally possible. Whereas the standard pattern in Middle Korean was to modify a noun with a preposed numeral, e.g. *twu kalh* (2 knife) ‘two knives’, under Chinese influence Contemporary Korean increased its use of classifiers, e.g. *pus se:k calwu* (writing.brush three CLF) in which *calwu* denotes long objects with handles.

However, the original pattern surfaces in expressions such as K *twu nala* ‘two countries’ and the use of classifiers remains optional in Korean, e.g. *kalh hana-ka issta* (knife one-NOM be.present) ‘there is one knife’.<sup>4</sup> While there is an extensive list of obligatory classifiers in Contemporary Japanese, e.g. *enpitu san-bon* (pencil three-CLF) ‘three pencils’, the use of classifiers is much less developed and is not obligatory in Old Japanese, where Chinese influence is restricted to a minimum.

Numeral classifiers are absent in Uralic languages such as Khanty, as well as in Yukaghir and Ket. Ainu and Nivkh make use of a set of obligatory classifiers. The obligatory use of classifiers is a widespread feature shared by Mandarin and the languages of Southeast Asia, but the use of classifiers in Classical Chinese was the exception rather than the rule. In Rukai, the use of classifiers is optional in the sense that it uses a set of unaffixed numerals without classifiers as well as a set of bound numerals, which combine with five different sortal classifiers to form verbs. In Gil’s (2005: 226–229) sample of 400 languages, 80% lack obligatory numeral classifiers.

(iv) F11 Presence of mi-Ti opposition in first vs. second singular personal pronouns  
Nichols (2012) observes that *m-T* pronominal paradigms with first person labial nasal *m* and second person apical or palatal obstruent *t, c, s*, etc. are much more common in northern Eurasia than elsewhere in the world. Janhunen (2013: 213)

adds that there is a smaller group of *mi-Ti* languages extending from Uralic in the west, to Turkic, Mongolic and Tungusic in the east, and Yukaghir in the north, in which not only the initial consonant but also the root vowel of the singular stems shows a basic similarity in that it contains a non-low unrounded front vowel *i* or *e*. Although *m* is absent in the nominative first person singular in the Turkic, Mongolic and Tungusic languages, e.g. Tk *ben*, OTk *ben*, Khal. *bii*, MMo. *bi*, Ma. *bi*, Evk. *bi:*, it has developed in oblique forms in assimilation to the nasal oblique suffix *-n*, e.g. OTk *min-*, Khal. *min-ii* (I-GEN), MMo. *mi-nu* (I-GEN), Ma. *min-*, Evk. *min-*. The second person singular forms all reflect a voiceless dental *T*, e.g. Tk. *sen*, OTk *sen*, Khal. *cii*, MMo. *ci*, Ma. *si*, Evk. *si:*. The Korean pronouns are first singular K/MK *na* and second singular K/MK *ne* among others. In Japanese, J *watasi* and OJ *wa* among others are used in the first singular, while a variety of contemporary pronouns and OJ *na* are used in the second singular. Although the Proto-Uralic first and second singular pronouns *\*mun* and *\*tun* reflect not a *mi-Ti* though still an *m-T* distinction (Janhunen 1982: 35), Khanty is deviant in having first singular *mä* and second singular *nöŋ*. In Yukaghir, however, the *mi-Ti* opposition is present in first singular *met* vs. second singular *tet*. In Nivkh, the distinction is absent in the singular pronouns, first person *n'i* vs. second person *či*, but it is present in the opposition between the first plural inclusive *mir/mer* and the second plural pronoun *čij*. The opposition is not found in Ket, Ainu, Chinese and Rukai. In Nichols and Peterson's (2005: 546–551) sample of 230 languages, 13% display an *m-T* opposition in first vs. second person pronouns. Languages reflecting a *mi-Ti* opposition will represent an even smaller proportion.

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

(v) F12 Formation of a secondary nasal oblique stem in personal pronouns

In most contemporary Turkic languages, the nominative and oblique stems of the personal pronouns have merged, e.g. Tk. *ben* for the first singular nominative and oblique, but in Old Turkic the first singular nominative *ben* is distinguished from the oblique stem *min-*, which can be derived from an original PTK *\*bi-n-* (1SG-OBL). Similarly, the Mongolic and Tungusic languages derive oblique pronominal stems from the nominative roots through a nasal suffix, for instance in the first person plural pronouns MMo. *ba* (NOM) vs. *man-* (OBL) and Khal. *bid* (NOM) vs. *bidn-* (OBL) and in the first person singular pronouns Ma. *bi* (NOM) vs. *min-* (OBL), Evk. *bi:* (NOM) vs. *min-* (OBL). There are no oblique pronominal stems in Contemporary Japanese, but in Old Japanese traces remain of an oblique nasal suffix in some case forms, e.g. in the Eastern OJ first person singular dative *wa-nu-ni* in alternation with Western OJ *wa-ni*. Vovin (2005: 229–230) further found that an original Japonic pronominal oblique *\*-n-* is well supported by Northern Ryukyuan dialects where the first person pronoun uses *waa-* as the nominative and genitive base and extended *waN-* in the oblique cases. Hence, with the exception of Korean, the Transeurasian languages share a tendency of forming a secondary oblique stem of the personal pronouns by means of a suffix, which can be identified phonologically as the dental nasal *-n-*. The oblique nasal suffix is an important element in the Uralic pronominal paradigm as well, e.g. the Khanty first person pronoun *mä* (NOM) vs. *män-* (OBL). Ket, Yukaghir, Ainu and Mandarin, however, do not derive secondary oblique stems. The third person singular pronoun in Nivkh



has both regular and suppletive case forms, e.g. *if-øn* (3SG-NOM) vs. *if-tox ~ e-rx* (3SG-DAT/ADD), but here the oblique form is not derived from the nominative base. Rukai personal pronouns have different shapes for nominative, topic, genitive and oblique cases, e.g. the first person singular *-lrao* (NOM), *ilrae* (TOP), *-li* (GEN) vs. *-iae* (OBL), in which the oblique seems to be formally derived from the nominative base by means of the same *i- ...-e* marking as in the topic form.

#### 10.3.4 Syntax

##### (i) F13 Dependent-marking of clause arguments

In the clause, the verb is the head and the arguments are dependents. Morphological marking, reflecting the syntactic relations in the clause, may be located on the head, on the dependent, on both, or on neither. Even if they may have subject-verb agreement on the verb, the Turkic languages are dependent-marking because they tend to mark agreement and case government more on dependents than on verbs; see the subject-verb and case agreement in the Turkish example in (1). Most Mongolic languages are strongly dependent-marking, as they mark case and lack verb agreement; see the Khalkha example in (2). Tungusic languages have case and verbal agreement with the subject; see the Evenki example in (3). Having case and lacking verb agreement, Korean and Japanese are strongly dependent marking, as illustrated in (4) and (5).

##### (1) Turkish

*Bu ev-i Ahmet-e yap-tı-m.*

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.

this house-ACC Ahmet-DAT make-PFV-1SG

‘I built this house for Ahmet.’ (Göksel and Kerslake 2005: 146)

## (2) Khalkha

*Öwgön-iig ger-t-ee ury-jee.*

old.man-ACC home-DAT-REFL invite-PFV

‘He [the tiger] invited the old man to his home.’ (Janhunen 2012a: 296)

## (3) Evenki

*Nungan eri gule-ve o:-ra-n.*

he this house-ACC make-PFV-3SG

‘He built this house.’ (Nedjalkov 1997: 83)

## (4) Korean

*Minca-nun halapeci-kkey ton-ul tuli-ess-eyo.*

Minca-TOP grandfather-DAT money-ACC give-PST-POL

‘Minca gave her grandfather some money.’ (Sohn 1994: 84)

## (5) Japanese

*Miki-ga Yamada sensei-ni tegami-o*

Miki-NOM Yamada teacher-DAT letter-ACC

*mise-ru.*

show-NPST

‘Miki shows the letter to professor Yamada.’ (Iwasaki 2013: 122)

Having subject-verb agreement and case marking, Khanty is weakly dependent-marking. It illustrates the tendency of gradually changing from double-marking to dependent marking in Uralic languages (Nichols 1986: 89). Yukaghir and Mandarin Chinese are dependent-marking. However, Nivkh, Ainu and Ket, the isolates in Eurasia, are all head-marking, as is Rukai. In Nichols and Bickel’s (2005a: 98–101) sample of 235 languages, 27% are dependent-marking in the clause.

(ii) F14 Dependent-marking in possessive noun phrases

In possessive noun phrases, the possessed noun is the head and the possessor is the dependent. Morphological marking, reflecting the syntactic relation between the possessor and the possessed, may be located on the head, on the dependent, on both, or on neither. The Turkic languages are double marking, e.g. *Ali-nin oğl-u* (Ali-GEN son-3SG.POSS) ‘Ali’s son’, *oda-nin kapı-sı* (room-GEN door-3SG.POSS) ‘the door of the room’. Old Turkic is commonly double-marking, e.g. OTk *ton-nuñ bit-i* (clothes-GEN louse-3SG.POSS) ‘clothes’ louse’, but there are also cases in which it is head-marked, e.g. *köl tegin atı-sı* (Köl Tegin nephew-3SG.POSS) ‘Köl Tegin’s nephew’. For seemingly unmarked cases such as OTk *balık kapag* (city gate) ‘city gate’, it is important to distinguish true possessive constructions from compound nouns. The Mongolic languages are dependent marking, e.g. Khal. *min-ii eej* (I-GEN mother) ‘my mother’, WMo. *šabi-yin nom* (pupil-GEN book) ‘the pupil’s book’.

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.

Except for Manchu, where possessive relations are dependent-marked, e.g. *ama i bo*: (father GEN house) ‘father’s house’, genitive case is absent in most Tungusic languages, since possessive relations are head-marked, e.g. Even *svinija ulrə-n* (swine meat-3SG.POSS) ‘swine’s meat, pork’. Both Contemporary and Middle Korean are dependent-marking, e.g. K *na-uy yenphil* (I-GEN pencil) ‘my pencil’, MK *mo-l-oy hyang* (horse-GEN scent) ‘the scent of horses’. This is also true for Contemporary and Old Japanese, e.g. J *anata no atama* (you-GEN head) ‘your head’, OJ *Yamato-no kuni* (Yamato-GEN land) ‘the land of Yamato’. Proto-Uralic was originally head-marking like Khanty, e.g. Khan. *qul-əm* (fish-2SG.SG) ‘your fish’.<sup>5</sup>

The loss of head-marking patterns and extension of dependent-marked ones in western Uralic languages result from the influence of the Indo-European languages to the west. Given the fact that Turkic is double marking in the possessive noun phrase, while it has the proto-typical Transeurasian pattern of dependent marking in the clause, it is likely that it acquired double marking under the influence of the proto-typical Uralic pattern. Yukaghir and Mandarin are dependent marking, e.g. Yuk. *tude kerewe-d ugurce* (3SG cow-GEN leg) ‘the legs of his cow’ and Man. *bāba de máma* (father GEN mother) ‘the mother of father’. However, the Eurasian isolates are prototypically head-marking, e.g. Ainu *eci-siki-hi* (2PL-eye-GEN) ‘your eyes’, Nivkh *vit-yan* (2SG-book) ‘your book’ and Ket *b-a:m* (1SG-mother) ‘my mother’. Rukai is head-marking as well, e.g. Ruk. *tolropongo-ni dhipolo* (hat-3SG.GEN Dhipolo) ‘Dhipolo’s hat’. In Nichols and Bickel’s (2005b: 98–101) sample of 235 languages, 42% are dependent-marking in the possessive noun phrase.

(iii) F15 Extensive use of converb

Converbs, also known as gerunds or adverbial participles, can be defined as nonfinite verb forms whose main function is to mark adverbial subordination (Haspelmath 1995: 3). Originally coined by the Altaic scholar Ramstedt, the term converb was adopted from Transeurasian linguistics to denote a cross-linguistic category. All Transeurasian languages are converb-prominent languages in the sense that they use converbs rather than adverbial subordinators as found in many European languages; see the examples below.

(6) Turkish

<i>Ali</i>	<i>gel-ince</i>	<i>şaşır-dı</i>
Ali	come-CVB	be.surprised-PST3SG

‘When Ali came, he was surprised’ (Johanson 1995: 314)

(7) Khalkha

<i>Ger-ees-ee</i>	<i>gar-aad</i>		
house-ABL-REFL	exit-PFV.CVB		
<i>deuc-en</i>	<i>jil-iin</i>	<i>daraa</i>	<i>ol-d-lao.</i>
forty-ADNZ	year-GEN	after	find-PASS-FIN

‘She went away from home and was found forty years later.’ (Janhunen 2012a: 280)

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

(8) Evenki

*əmə-mme:n*                      *iri-l-i-m.*

come-CVB                      cook-INCH-NPST-1SG

‘As soon as I arrive, I will start cooking.’ (Bulatova and Grenoble 1999: 44)

(9) Korean

*Kiho-nun*      *nol-ko*                      *ca-ss-eyo.*

Kiho-TOP      play-CVB                      sleep-PST-POL

‘Kiho played and then slept.’ (Sohn 2009: 300)

(10) Japanese

*Taroo-ga*      *bangohan-o*      *tabe-te*                      *furo-ni*                      *hai-ta.*

Taroo-NOM      dinner-ACC      eat-CVB                      bath-DAT                      enter-PST

‘Taroo took a bath after he ate dinner.’ (Alpatov and Podlesskaya 1995: 473)

Although the Uralic languages are characterized by extensive use of converbs, Khanty is rather atypical in this sense because it has only a single converb in *-min*, which is the least frequent nonfinite verb form. Yukaghir and Nivkh also use a variety of converbs to link clauses. Ainu, however, employs subordinating conjunctions. Ket has no converbs or serial verb constructions of any kind. In Mandarin, verbs or verbal phrases are merely juxtaposed, the relation between the items being largely unmarked. Rukai marks adverbial subordination through a

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.

variety of means such as subordinating conjunctions, changes in word order and nominalized verb forms.

(iv) F16 Use of locative existential constructions to encode predicative possession

The Transeurasian languages show a clear preference for expressing the concept “X has Y” on the basis of an existential sentence, whereby the possessed noun phrase functions as the grammatical subject of the ‘exist’-predicate, while the possessor noun phrase is in a dative-locative case form. Although locative possessive constructions were standard in Old Turkic, Turkish uses genitive existential sentences as well as locative existential sentences. ‘I have a book’, for instance, can be expressed by *Ben-de bir kitab var* (I-LOC a book exist) or by *Ben-im bir kitab-im var* (I-GEN a book-1SG.POSS exist). Middle Mongolian and Khalkha make use of either a conjunctive possessive which construes the possessor noun phrase as the grammatical subject of the copula and marks the possessed with the comitative *-tai*, e.g. Khalkha *Bi nom-tai bai-n*’ (I book-COM be-DUR), or a locative possessive, e.g. *Nad-ed nom bai-n*’ (I-DAT book be-DUR). As is the case for most Tungusic languages, Manchu and Evenki employ locative existential constructions, e.g. Evk. *Min-du: kniga bisi-n* (I-DAT book be-3SG); Ma. *Min-de bithe bi* (I-DAT book be). Korean uses a locative existential construction, e.g. K *Na-hanthey chayk-i issta* (I-LOC book-NOM exist), but the possessor can also be construed as the topic of the noun phrase, e.g. *Na-nun chayk-i issta* (I-TOP book-NOM exist). This is also true for Japanese, e.g. *Watashi-ni hon-ga aru* (I-DAT book-NOM exist) and *Watashi-wa hon-ga aru* (I-TOP book-NOM exist). Topic possessives may have developed under the

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.

influence of Chinese, since they represent the standard strategy in Mandarin.

Among the strategies used to encode predicative possession in the Uralic languages, we find locative possession, as in Finnish and Hungarian, genitive possession, as in Nenets, and possession encoded by a transitive verb ‘to have’, as in Khanty. Whereas Yukaghir employs a conjunctive possessive and Ainu a ‘have’-possessive, Ket and Nivkh use locational possessives. Although many Austronesian languages employ topic possessives, Rukai makes use of locative and genitive possessive constructions. In Stassen’s (2005c: 474–477) sample of 240 languages, 20% use a locative existential construction to encode predicative possession.

(v) F17 Use of the ablative case form to encode predicative comparison

The Transeurasian languages all form comparative constructions in which the standard noun phrase is constructed in the ablative case form, e.g. Tk. *bu araba-dan daha büyük* (this car-ABL more big) ‘bigger than this car’, OTk *barča-da üzä-räk* (everything-ABL high-COMP) ‘higher than anything else’, Khal. *ene xun-ees iluu* (this person-ABL good) ‘better than this person’, MMo. *qola-sa qola* (far-ABL far) ‘farther than far’, Evk. *oron-duk gugda-tmar* (deer-ABL tall-COMP), Ma. *ere niyalma ci sain* (this person ABL good) ‘better than this person’, OJ *ware-yo<sub>1</sub>ri mo<sub>2</sub> mantusi-ki<sub>1</sub> pi<sub>1</sub>to<sub>2</sub>* (I-ABL PTCL be.poor-ADNZ person) ‘people poorer than me’ and J *chikyu:-yori omoi* (globe-ABL be.heavy) ‘heavier than the globe’. In literary Korean, the ablative marker *eyse* ‘from’ can be used in comparative constructions, e.g. K *i eyse te khu-n salang* (this ABL more be.big-ADNZ love) ‘a greater love than



This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

this’, but it is more common to use a comparative particle *pota* ‘than’, e.g. K *kicha pota ppaluta* (train PTCL be.fast) ‘faster than a train’, MK *nyey pwota thak.wel hota* (past PTCL superior be) ‘superior to the past’. The Uralic languages differ from one another with regard to comparative constructions; languages to the west, such as Finnish and Hungarian, use more particle comparatives as in European languages, languages to the east, such as Nenets and Udmurt, mark the comparative standard with the ablative case ending, as in the Transeurasian languages. In Khanty, the marker of comparison is a postposition *niŋə* ‘since, from’, which has ablative-like semantics but differs from the standard ablative case ending *-oy* or the ablative-relative ending *-i*. Yukaghir and Ket mark the comparative standard with the ablative case ending. In Nivkh, the comparative suffix *-yk* is traditionally considered a separate case form, as there is no evidence to relate it to the formally similar locative-ablative suffix *-(u)ye; -(u)x* (Gruzdeva p.c.). Ainu forms comparative constructions by means of the particle *kasuno* ‘than’. In comparative constructions in Mandarin, the standard noun phrase is constructed as the direct object of a verb *bǐ* ‘to compare’. In Rukai, a comparative construction is formed through partial reduplication (CVV) of the descriptive verb stem. In Stassen’s (2005a: 490–493) sample of 167 languages, 47% use locational comparatives, but the proportion of languages that specifically use the ablative case form to encode predicative comparison is expected to be lower.

### 10.3.5 Grammaticalization patterns

### (i) F18 Direct insubordination

The Transeurasian languages display a recurrent tendency to reanalyze non-finite suffixes as finite ones without the omission of a specific matrix predicate, a tendency which I call “direct insubordination” (Robbeets 2015, 2016a; this volume: Chapter 30). Comparative evidence indicates that these markers originated as deverbal noun suffixes, marking a derivational process at the lexical level, which were then extended to function as (ad)nominalizers in dependent clauses at the syntactic level, and eventually—through a pragmatic role in discourse—were extended still further to mark finite forms in independent clauses. For instance, deverbal noun suffixes such as OTk *-(A)r* in OTk. *tug-* ‘to be born, to rise (of sun) (intr.)’ → *tugar* ‘sunrise, east’; MMo. *-m* in MMo. *quri-* ‘to come together (intr.)’ → *qurim* ‘feast’; Ma. *-rA* in *mute-* ‘to be able’ → *mutere* ‘ability’; MK *-(u/o)m* in *yel-* ‘to bear’ → *yelum* ‘fruit’ and OJ *-sa* in *naga-* ‘to be long’ → *nagasa* ‘length’ develop over intermediate stages of clausal nominalizers and relativizers into finite suffixes, as illustrated in examples (11) to (15).

#### (11) Old Turkic

<i>Ölüm-tä</i>	<i>oz-upan</i>	<i>ögir-ä</i>	<i>savin-ü</i>
death-ABL	escape-CVB	rejoice-CVB	be.happy-CVB

*yori-r.*

go.on-FIN

‘Having been saved from death it happily goes on with its life.’ (Erdal 2004:

325)

(12) Middle Mongolian

*Udurit-basu ber ulu busire-m.*

guide-COND PTCL NEG believe-FIN

‘Even if you guide them, they don’t believe.’ (Weiers 1966: 144)

(13) Manchu

*Si nene-me isinji-ci uthai sin-de bu-re.*

you be.first-CVB come-CVB at.once you-DAT give-FIN

‘If you come first, I shall give [it] to you straight away.’ (Gorelova 2002: 256)

(14) Korean

*Onul-un swuep-i eps-um.*

today-TOP class-NOM not.exist-FIN

‘No class today.’

(15) Old Japanese

*Punapi<sub>1</sub>to<sub>2</sub>-wo mi<sub>1</sub>-ru-ga to<sub>2</sub>mo<sub>2</sub>si-sa.*

boat.people-ACC see-NMLZ-GEN be.enviable-FIN

‘How enviable it is to see the boat-people!’ (Wrona 2008: 206)

The Uralic languages also display a recurrent tendency toward direct insubordination. Deverbal noun suffixes such as Proto-Uralic *\*-k*, *\*-pÄ*, *\*-mə* and *\*-śÄ* are thought to have developed into finite markers for present-day (*\*-k*, *\*-pÄ*) and past (*\*-mə*, *\*-śÄ*) tense, either in Proto-Uralic or after the separation of the daughter languages (Collinder 1965a: 110–115; Janhunen 1982: 36–37). Eastern Khanty preserves only a faint trace of this development since the finite form of the negative verb can be marked with the perfective participle *-əm*, as illustrated in example (16).

(16) Eastern Khanty

<i>Məta</i>	<i>wajay</i>	<i>lök</i>	<i>ənt-im.</i>
some	animal	track	NEG-FIN

‘There is not a single animal track.’ (Filchenko 2007: 429)

In Nivkh the deverbal action noun and infinitive suffix *-dʹ* has developed into a finite form *-dʹ*, as illustrated in example (17). However, rather than being a case of “direct insubordination”, Gruzdeva (2016: 196) attributes this development to the lexicalization of a modally marked form of the copular verb *ha-* ‘to do so’. When the modally marked copula was lexicalized into a modal particle, *-dʹ* was reanalyzed as a finite form.

(17) Nivkh

<i>If</i>	<i>hum-dʹ</i>	<i>hyjm-dʹ.</i>
-----------	---------------	-----------------

he live-NMLZ know-FIN

‘He knows the living one/(his) life.’ (Malchukov 2013: 200)

As in Nivkh, clausal nominalization in construction with a copula is the main source for developing new finite constructions in Yukaghir and Mandarin (Yap and Matthews 2008: 20; Malchukov 2013: 192–195). Ket displays yet another strategy for developing finite markers, namely to reduce the matrix predicate to an affix on the former dependent verb (Malchukov 2013: 196–197). In Ainu, deverbal noun suffixes appear to function as both derivational suffixes and syntactic clausal nominalizers, but there is no indication that they have developed into finite endings. Ainu lacks other nonfinite markers such as participial or converb affixes that could be open to developing into finite markers. Similarly, Rukai does not exhibit traces of direct insubordination.

(ii) F19 Grammaticalization from negative verb to verbal negator over a construction comprising an inflected negative auxiliary and an invariant lexical verb

In Turkic, we find indications that the verbal negative suffix OTk *-mA-* originated as an inflecting negative auxiliary verb plus an invariant derivationally complex lexical verb. These include the stem-internal position of negation in Turkic, sandwiched between derivation and inflection; the occurrence of the adnominal negative suffix OTk *-mA-z*, which seems to have developed from an aorist *-r* in unaccented position; the observation that Chuvash *mar* acts as an independent negative verb, taking a nominal argument in petrified constructions such as (18);

and the analysis of the Chuvash optative first person singular as an auxiliary negative form, whereby inflection has shifted to the lexical verb, e.g. *vula-m mar* (read-1SG NEG) ‘I will not read’. In addition, we can find independent lexical cognates for PTK *\*ma-* ‘not to exist’ in the other Transeurasian languages (Robbeets 2015: 203–204).

(18) Chuvash

<i>Man-ăn</i>	<i>kil-melle</i>	<i>mar.</i>
I-GEN	come-DEB	NEG
	<i>*Kel-me-lle</i>	<i>ma-r.</i>
	come-NMLZ-DIR	NEG-FIN

‘I don’t have to come.’

This proto-typical Transeurasian grammaticalization cycle is more explicitly recoverable in the Tungusic languages; see examples (19) and (20). In (19a) Evenki *e-* is an independent negative verb ‘not to exist, not to live’. In example (19b), the negative verb acts as a finite auxiliary to the lexical verb, which assumes an invariant adnominal form, and in (21c) the negation has shifted to a postposed position. The Nanai example in (20) represents the final stage of the negative cycle, i.e. fusion, whereby the auxiliary negative verb has assumed the status of derivational suffix on the lexical verb and its phonological form is reduced to lengthening of the stem-final vowel.

(19) Evenki

- a. *Esile e-dyeli-m tadu-gla.*  
now NEG-FUT-1SG there-ENCL

‘Now I will not be (live) there.’ (Nedjalkov 1994: 27)

- b. *Nungan nekun-mi e-ce-n*  
he younger.brother-POSS.REFL NEG-PST-3SG  
*suru-v-re.*  
go.away-CAUS-ADNZ

‘He did not lead his younger brother away.’ (Nedjalkov 1994: 11)

- c. *Nungan songo-ro e-ce-n.*  
he cry-ADNZ NEG-PST-3SG

‘He did not cry [—what’s the use of crying?].’ (Nedjalkov 1994: 8)

(20) Nanai

- Xola:-ci-si.*  
read.NEG-PST-2SG  
‘You didn’t read.’

Similarly, Middle Mongolian *ese-* acts as an independent negative verb, meaning ‘not to be, not to exist’ inflected with past *-be* in (21a), but gradually the negative auxiliary came to be used as an invariant form, transferring its entire inflection to

the lexical verb; e.g. the past marker *-be* is attached to *ire-* ‘to come’ in example (21b).

(21) Written Mongolian

a. *Ükü-be-iiü*                      *ese-be-iiü*.

die-PST-INTER                      NEG-PST-INTER

‘Did [he] die or did [he] not?’ (Poppe 1954: 175)

b. *Manu*                      *baγši*                      *ese*                      *ire-be*.

our                      teacher                      NEG                      come-PST

‘Our teacher did not come.’ (Poppe 1954: 175)

Old and Contemporary Japanese use an independent negative existential adjective *na-* ‘to be non-existent, not to exist’, which is thought to derive from the same origin as the Old Japanese negative suffix *-(a)n-*. The Korean verbal negator MK *a-ni*, K *an(i)* can also be derived from an original negative verb *\*an-* and the suffix MK *-i* that derives both nouns and adverbs from verbs (Robbeets 2014a).

Similar to the Transeurasian languages, one of the characteristics of the Uralic languages is the expression of negation by means of a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb (Comrie 1981; Janhunen 1982: 37; Payne 1985: 215–221; Honti 1997; Suihkonen 2002: 173). In the case of the Khanty negative particle *əntə*, the negative auxiliary has become totally free of inflections and turned into an invariant verbal negator, which recalls



This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

the situation in Mongolic in (21b). However, there are no examples in Uralic in which the negative auxiliary ultimately becomes a suffix, as it does in Turkic, Tungusic and Japanese. In Yukaghir there are no language-internal indications that the proclitic clausal negation *el-* originated in a negative verb or auxiliary. For Nivkh and Chinese, we find indications that the verbal negator originated as an independent verb, but this did not follow the same pathway, through a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb. For Khanty, Nivkh and Chinese, we find indications that the verbal negator originated as an independent verb, but only in Uralic did this happen along the same pathway, through a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb. Ainu uses a negative particle that precedes the verb and cannot be derived from a verb. Neither are there indications that the Ket negative particle *bə:n* or the Rukai negative suffix *-ka* originated from a negative verb. Worldwide, the expression of negation via negative auxiliaries is a minor type to begin with, being found in only 40 (17%) out of 240 languages in Dahl' s (1979) sample, which is areally biased towards Uralic and Altaic languages, in 45 (4%) out of 1011 languages in Dryer' s (2005b) sample, and in 16 (5%) out of the 297 languages in Miestamo' s (2005) sample. As a consequence, the particular development of negative verbs to auxiliaries to particles or suffixes is even more rare.

(iii) F20 Grammaticalization of plural/collective markers to express

inclusive/exclusive distinction on first person pronoun

The Transeurasian languages share a particular grammaticalization pattern, by which plural and collective suffixes on the first person pronoun grammaticalize into an inclusive/exclusive distinction. Among the Turkic languages, there are no unique pronominal forms that distinguish inclusive from exclusive person forms, although traces of the distinction remain in the imperative paradigms of Yakut, Tofa, Tuvan, Turkmen, Khakas, Shor, Altay and Chulym (Schönig 1987; Nevskaya 2010: 122). Nevertheless, Old Turkic and most currently spoken varieties of Turkic distinguish between a first person plural (Tk./OT *biz* ‘we’) and an augmented plural form (Tk./OT *biz-ler* ‘we (as a group)'). Nevskaya (2010: 124) argues for a collective interpretation of the augmented plural, denoting “an isolated group of people who want to oppose themselves to the others”, rather than an inclusive interpretation as suggested by Grönbech (1936: 81). However, the collective interpretation seems to be an intermediate stage on the way from augmented plural to an inclusive/exclusive distinction, because in the imperative paradigms of Khakas, Shor, Altay and Chulym, the inclusive is derived from the first plural marker augmented with the plural marker *-LAR*. As the person endings on verbs have grammaticalized from original person pronouns, it seems safe to assume that the augmented plural first person markers developed into inclusive markers at some stage in pre-Old Turkic.

The Middle Mongolian distinction between exclusive *ba* and inclusive *bida* is formally preserved in the Khalkha oblique paradigm in the variation between for-

mally exclusive *man-* and formally inclusive *bidn-*, but the functional distinction has been lost. Etymologically, the Middle Mongolian inclusive *bida*, reflected in the Khalkha formally inclusive oblique *bidn-*, derives from the first person singular MMo. *bi* ‘I’ and a plural suffix *-dA*, which also occurs in the plural demonstrative pronouns MMo. *e-de* ‘these’ vs. *te-de* ‘those’ (Doerfer 1985: 2; Domii 2006; Nevskaya 2010: 119).<sup>6</sup> Domii argues that originally *\*ba* and *\*bi-da* complemented each other as plural pronouns, and that the distinction between exclusive and inclusive meaning was a secondary development.

In the Tungusic languages, the inclusive-exclusive opposition is generally well preserved, e.g. exclusive Ma. *be*, Evk. *bu* vs. inclusive Ma. *muse*, Evk. *mut* ~ *mit*. The exclusives Evk. *bu* and Ma. *be* can be derived from the first person plural PTg *\*bö* and an augmented plural *\*bö-(x)e*, respectively (Doerfer 1978b: 81–83, 95–96; Janhunen 2013: 217), whereas the inclusive Evk. *mut* ~ *mit* may go back to PTg *\*bö* plus the collective suffix PTg *\*-ti* (Benzing 1955a: 1020), and the inclusive Ma. *muse* may be an extension of this root with the collective suffix *-sA* (Benzing 1955a: 1017–1018).

Similar to the Turkic languages, Middle and Contemporary Korean distinguish between a first person plural (K/MK *wuli* ‘we’) and an augmented plural form (K *wuli-tul*, MK *wuli-tolh* ‘we (as a group)’) in which K *tul*, MK *·tolh* is a collective marker. A similar tendency can be found in the history of Japanese, where the first person singular/plural OJ *wa-* ‘I, we’ coexists with the same form augmented by a collective marker OJ *wa-re* ‘we’, a form which in its turn was later augmented into *ware-ra* ‘we’. Like contemporary Japanese, Old Japanese lacks a real inclusive-

exclusive distinction, but the distinction is well preserved in the Ryukyuan languages; for example, in Kikai (Amami), Sesoko (Okinawa), Ikema (Miyako), Irabu (Miyako), Tarama (Miyako), Hateruma (Yaeyama), and Yonaguni (Yaeyama) (Shimoji 2014). In most cases the exclusive can be derived from the first person singular plus a plural suffix, while the inclusive is based on the same form plus a collective suffix, e.g. the Kikai exclusive suffix *-naa* is also used as a plural suffix in the second and third person plural pronouns, while the inclusive suffix *-tjaa* is also used as a collective suffix on human nouns. Therefore, it seems sensible to assume a recurrent tendency whereby plural and collective pronouns grammaticalized into an inclusive-exclusive distinction in Proto-Japonic.

As is the case in many Uralic languages, Khanty marks a dual distinction, but not an inclusive-exclusive distinction, on its person pronouns. While Ket and Yukaghir lack the distinction, Nivkh distinguishes between exclusive *n'yŋ* and inclusive *mer ~ mir*, and Ainu between exclusive *cóka* and inclusive *aoka*, but there is no indication that the distinction derives from augmented plural or collective marking.<sup>7</sup>

The inclusive-exclusive distinction found in the first person plural pronouns between exclusive *wōmen* and inclusive *zánmen* ‘we’ of Beijing and certain other northern Chinese dialects may be due to Transeurasian influence. Such a distinction was not found in Old Chinese, and it began to appear in North China during the period of Altaic rule. It is significant in this regard that both Middle Mongolian, spoken under the Yuan dynasty, and Manchu, spoken under the Qing dynasty, distinguish exclusive and inclusive forms. Rukai distinguishes exclusive *-nai ~ nai-*

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.

(NOM) from inclusive *-mita* ~ *ta-* (NOM), a feature characteristic of Austronesian languages, but the forms are In Rukai, these forms are not derivable from each other. In Cysouw's (2005: 166–167) sample of 200 languages, 31% differentiate inclusive and exclusive in independent pronouns. Therefore, the percentage of languages that developed the distinction from plural/collective marking on first person pronouns will be much smaller.

### 10.3.6 Overview

By way of overview, Table 10.1 shows the feature values for the 20 features discussed above for the selected contemporary and historical varieties of the Transeurasian languages and neighboring non-Transeurasian languages. In the first column, I added an estimate of the cross-linguistic frequency of the feature in question, if such information is available. As these cross-linguistic percentage probabilities are relatively low, they indicate that the shared properties are not due to mere universal principles in linguistic structuring. This observation contradicts the claim that the properties of the Transeurasian language type are universally so common that their parallel occurrence in several adjacent language families can be explained by coincidence. In the penultimate rows of the table, I added the plus values and in the final row, I calculated the percentage of positive feature values for a given language.

Table 10.1 Feature values for selected Transeurasian languages along with their historical stages and representative neighboring languages

This is a draft version of a chapter that appears in Robbeets, M and A. Saveljev (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.

Frequency worldwide	F	Tk	(pre-)OT	Khal	(pre-)MMo	Evk	(pre-)Ma	K	(pre-)MK	J	(pre-)OJ	Khan	Ket	Yuk	Niv	Ain	Ch	Ruk	
< 10%	01	-	-	+	+	+	+	-	+	-	+/-	-	-	+	+	+/-	-	-	
< 30%	02	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-	
35%	03	+	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	-	
61%	04	+	+	+	+	+	+	-	-	+	-	-	+	+	+	-	-	+	
45%	05	+	+/-	+	+/-	-	+/-	+	+/-	+	+/-	-	-	-	-	-	-	-	
54%	06	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	+	-	
< 27%	07	-	+	-	+	-	+	-	+	-	+	-	-	-	-	-	-	-	
43%	08	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	
	09	+	+	+	+	-	+	-	+	-	+	-	-	-	-	+	+	-	
80%	10	+	+	+	+	+	+	+	+	-	+	+	+	+	-	-	-	+	
< 13%	11	-	+	+	+	+	+	-	-	-	-	-	-	+	-	-	-	-	
	12	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	-	-	
27%	13	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	-	
42%	14	-	-	+	+	-	+	+	+	+	+	-	-	+	-	-	+	-	
	15	+	+	+	+	+	+	+	+	+	+	-	-	+	+	-	-	-	
20%	16	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	-	+	
< 47%	17	+	+	+	+	+	+	+	-	+	+	-	+	+	-	-	-	-	
	18	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	
	19	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	
	20	-	+	+	+	+	+	-	-	-	+	-	-	-	-	-	-	-	
		+	13	18	19	20	15	20	12	14	11	18	8	5	11	4	2	5	3
		%	65	90	95	100	75	100	60	70	55	90	40	25	55	20	10	25	15
Frequency worldwide	F	Tk	(pre-)OT	Khal	(pre-)MMo	Evk	(pre-)Ma	K	(pre-)MK	J	(pre-)OJ	Khan	Ket	Yuk	Niv	Ain	Ch	Ruk	

## 10.4 Discussion

### 10.4.1 Typological features specific to Transeurasian

The 20 feature values listed in Table 10.1 represent a concentration of proto-typical linguistic features that delimit the Transeurasian languages from their neighbors.

This is clear from the fact that 55% or more of the structures are represented in the Transeurasian languages, compared to 40% and less in the neighboring languages.

Moreover, in the historical varieties of Transeurasian, the concentration of

linguistic features is even more pronounced. Although Khanty (40%) and the Uralic

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.

languages in general show more typological similarity with the Transeurasian core structures than do other neighboring languages, it is still possible to delimit the Transeurasian languages in relation to their Uralic neighbors. This seems to contradict Janhunen's statement quoted in the introduction.

Among the core structures that enable us to delimit the Transeurasian languages in relation to their Uralic neighbors are (F1) presence of tongue root vowel harmony in Transeurasian vs. palatal harmony in Uralic; (F4) presence of voicing distinction for stops vs. original singleton-geminate distinction in Uralic; (F5) preference for non-verbal strategy for (extra-family) verbal copies vs. direct insertion in Uralic; (F7) mixed and switched encoding of property words in Transeurasian vs. nominal encoding in Uralic; (F9) the use of bare stems as imperatives in Transeurasian vs. the marking of imperatives with second person suffixes in Uralic; (F11) absence of a primary initial *m* in the nominative first person singular vs. presence in Uralic, as well as presence of a (secondary) *mi-Ti* opposition in first vs. second singular person pronouns vs. *m-T* distinction in Uralic; (F13) dependent-marking of clause arguments in Transeurasian vs. original double-marking in Uralic; (F14) dependent-marking in possessive noun phrases vs. original head-marking in Uralic; and (F20) grammaticalization of plural/collective markers to inclusive/exclusive distinction on first person pronoun, compared to Uralic's lack of an inclusive-exclusive distinction on its first person pronouns.

Along the margins of the Transeurasian continuum, we can observe a gradual loss of Transeurasian features in the western and eastern peripheries. Examples of original Transeurasian features in Turkic probably changing under Uralic influence

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.

include (F1) Transeurasian tongue root harmony, which aligns with the Uralic languages as palatal harmony in Turkic; (F7) gradual loss of verbal encoding of property words—mirroring Uralic nominal encoding—as one proceeds from older to contemporary varieties and from Tungusic in the east to Turkic in the west; (F11) the secondary development of *m*-initials yielding a *mi-Ti* opposition in first vs. second singular personal pronouns in Turkic, Mongolic and Tungusic; and (F14) the fact that Turkic is double marking in the possessive noun phrase, while it has the proto-typical Transeurasian pattern of dependent marking in the clause, under the influence of the proto-typical double marking in Uralic. Therefore, we can say that sandwiched between the Uralic languages to the west and the other Transeurasian languages to the east, Turkic behaves like a proto-typical Transeurasian branch under Uralic influence.

Along the eastern margins of the Transeurasian continuum, we can observe a loss of Transeurasian features, some of which may have occurred under influence of substratum interference from pre-Nivkh and pre-Ainu in Tungusic, Koreanic and Japonic (Robbeets et al. 2017). Other features seem to have diffused from Chinese into Manchu, Korean and Japanese in historical times (Robbeets 2017g).

#### 10.4.2 Structural heritage vs. linguistic area

Typological parallelism does not necessarily imply genealogical relationship, but neither does it automatically mean areal diffusion. Even if we find support for Transeurasian ancestry in the basic vocabulary and the verb morphology (Robbeets, this volume: Chapter 36 and 30), we also know that the individual language



families were once concentrated in a compact area in southern Manchuria and present-day eastern Mongolia, adjacent to the homelands of Yeniseic, Yukaghiric, Amuric (Nivkh) and Sinitic speakers (Robbeets et al., this volume: Chapter 43).

This geographical situation provided an excellent opportunity for prehistoric contact within the Transeurasian family. As a result, the sharing of the structural features identified in Section 10.3 could be the result of diffusion, inheritance or a combination of both features. Therefore, the distinction between areal and genealogical features is one of the major challenges in the Transeurasian debate. In what follows, I make some general observations that help us to distinguish between areal and genealogical signals in our structural dataset.

(i) Geography: absence of the features in spite of geographical proximity and sharing of features in spite of geographical isolation

Although the Sea of Japan and the Tsushima Strait form a strong geographical boundary separating Japanese from the other Transeurasian languages, Japanese is typologically closer to the Transeurasian languages than less geographically isolated languages such as Ket, Yukaghir, Nivkh, and Chinese. Even if we assume that Japonic was once present on the continent, this indicates that the Transeurasian characteristics in Japonic did not exclusively arise through diffusion because Ket, Yukaghir, Nivkh and Chinese also were—and still are—present in the region.

This is a draft version of a chapter that appears in Robbeets, M and A. Savel'yev (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.

(ii) History: Structural cohesion of Transeurasian decreases with contact over time

The typological coherence seems to be greater for historical than for contemporary stages of the languages investigated: 90% in (pre-)Old Turkic vs. 65% in Turkish, 100% in (pre-)Middle Mongolian vs. 95% in Khalkha, 100% in (pre-)Manchu vs. 75% in Evenki, 70% in (pre-)Middle Korean vs. 60% in Korean, and 90% in (pre)Old Japanese vs. 60% in Japanese. The strongest decrease is observable at the peripheries of the Transeurasian continuum and took place under influence of Chinese, Siberian or Uralic features over the last two millennia. For Japanese, for instance, the traceable changes in features such as (F1) loss of tongue root contrast for its high vowels; (F3) loss of its velar nasal phoneme; (F4) loss of voicing distinction for stops; (F6) development of a three-way contrast for its demonstratives; (F7) increase of verbally encoded property words; (F8) development of derivational prefixes and (F9) reinforcement of the bare stem imperatives by an exclamatory particle is probably due to substratum influence of Ainu (Robbeets et al. 2017). This observation is indicative of areal divergence rather than convergence over time.

(iii) Distribution: Sharing of features was originally lower in Korean than in Japanese

Although Korean occupies an intermediate position in the contact chain between the Altaic languages and Japanese, (pre-)Middle Korean displays less structural uniformity (70%) than (pre-)Old Japanese (90%). Within a scenario of gradual areal diffusion of features, we would expect the positive values in Japanese to be

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.

significantly lower than in Korean. It is further difficult to explain how some Transeurasian features, such as the nasal oblique pronominal stems in F12 could show a gap in Korean, the nasal suffix having diffused into Japanese without a Korean intermediary.

(iv) Areal dominance: Transeurasian core structures are borrowed into other families

In his description of genealogically stable structures of Turkic, Johanson (2015: 586–587) argued that “these specific core structures have turned out to be dominant in all the numerous family-external contact situations without being overruled by copying. They have all maintained a high genealogical stability and resisted areal influence”. Interestingly, as previously noticed by Hashimoto (1986), Norman (1988: 10–12, 20) and Comrie (2008), some of the Transeurasian features discussed here have left a clear mark on the linguistic structure of Sinitic. Examples include (F6) the development of a two-way distinction in Chinese demonstratives as compared to the three way-distinction in Classical Chinese; (F8) the weak suffixing tendency of Mandarin as opposed to other Sinitic languages; and (F20) the development of an inclusive-exclusive distinction in first person plural pronouns in Beijing and certain other northern Chinese dialects, which was not found in Old Chinese. The dominance of these Transeurasian features over proto-typically Sinitic features may be indicative of their genealogical stability within the Transeurasian family.

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.

(v) Recurrent grammaticalization: Prototypical Transeurasian structures are cyclically renewed by new morphological means

Aikhenvald (2013) characterized contact-induced grammaticalization as “change against the grain” or atypical grammaticalization, while she regarded

genealogically motivated grammaticalization as “change that reinforces similarities” because it tends to maintain uniformity between related languages.

Given that languages tend to renew their formal encodings in cyclic processes of grammaticalization while maintaining their inherited grammatical categories, new forms are thus expected to grammaticalize along shared conceptual pathways to restore old categories (Heath 1998: 729; Robbeets 2013; Johanson 2015: 588).

Consequently, genealogically motivated grammaticalization is expected to recur on different formal encodings at various points in time, while contact-induced grammaticalization is expected to be restricted to a single formal encoding (or to a very limited number of encodings) during a certain period of contact. The repeated waves of grammaticalization and replacement by new forms involved in features 18 to 20 imply that the parallel grammaticalization patterns are genealogically motivated.

(vi) Isomorphism: core structures combine with formal correspondences

The observation that some core structures shared among the Transeurasian languages combine with a formal correspondence of the marker reflecting the particular feature is also indicative of genealogical retention. This is for instance the case for (F5) the non-verbal strategy of verbal borrowing employing a deverbal

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.

noun suffix of the common shape *\*-lA-* (Tk. *-lA-*, Khal. *-l-*, Ud. *-lA-*, J *-r(a)-*) to accommodate verbal borrowings (Robbeets 2015); (F12) the formation of a secondary oblique stem of personal pronouns through a common suffix *\*-n-* in all Transeurasian languages except Korean; and in the features (18) to (20), the grammaticalizations of finiteness, negation and inclusive-exclusive distinctions (Robbeets 2014a, 2016a). In such cases, the structural feature is likely to be genealogically motivated because it is complemented by a formal correspondence of a grammatical morpheme.

This set of diagnostic criteria suggests that at least part of the shared features identified in Section 10.3 are the result of structural heritage rather than being due to areal diffusion.

## 10.5 Conclusion

In this chapter, I aimed at filling a gap in the current literature about structural features that could potentially define and delimit the Transeurasian languages. In contrast to previous research in this field, I delimited the Transeurasian language type in relation to its non-Transeurasian neighbors and tried to distinguish a so-called “typological heritage” from areally diffused features. In order to separate the effects of inheritance from areal diffusion, I took a diachronic typological approach, which included the structural profile of the earliest written varieties of the Transeurasian languages. My argumentation is based on a number of qualitative criteria, but it is complemented by an alternative quantitative approach to distinction between inherited and diffused features by Hübler in the next chapter.

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.

First, I demonstrated that, against a commonly held belief, it is legitimate to speak of “Transeurasian” instead of “Ural-Altai” for there indeed are typological features that are specific only to “Transeurasian” without “Uralic”. My contribution did not settle the recent debate about the existence of a Siberian area, but it established a set of prototypical Transeurasian features shared by the languages of Siberia, which are barely present in the non-Transeurasian languages of Siberia.

Second, I argued that it is legitimate to characterize the Transeurasian languages as a “typological heritage”, even if I accept that other linguistic features may characterize it as a “linguistic area” as well. Transeurasian is a textbook example of “contact in the family” and as expected, these two complementary aspects of language change, borrowing and inheritance, are also reflected in its structural characteristics.

A fuller study would need to take more feature values into account and to insert a larger variety of Transeurasian languages as comparative points. Neighboring languages should also be more diversified and adjacent languages in the west such as Indo-European languages or languages of the Caucasus region should be included. Nevertheless, I hope to have contributed here to the understanding of the structural heritage of the Transeurasian languages.

### **Acknowledgement**

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.

---

<sup>1</sup> Joseph et al. (this volume: Chapter 29) point out that Kalmyk and Oirat articulatorily have Tongue Root rather than palatal harmony, as generally claimed in the literature.

<sup>2</sup> All tongue root oppositions have implications for vowel height, and since many systems have not been adequately investigated articulatorily, “vowel height” seems a reasonable term in such cases, where there are clear perceptual differences in terms of vowel height, but the articulatory mechanism is as yet unknown.

<sup>3</sup> Note that this analysis deviates from the feature values given for distance contrasts in demonstratives by Diessel (2005: 170–173), since he marks Ainu, Nivkh, Yukaghir and Turkish as having a two-way contrast.

<sup>4</sup> Note that my evaluation differs from Gil’s (2005: 228–229) interpretation that Korean has obligatory numeral classifiers.

<sup>5</sup> In Khanty, the possessive suffix makes reference to the number and person of the possessor, as well as to the number of the entity possessed (Filchenko 2008: 80).

<sup>6</sup> An alternative analysis, deriving the inclusive MMo *bida* from the first singular pronoun *\*bi* ‘I’ plus the second plural pronoun *\*ta* ‘you (many)’ is proposed by Janhunen (2013: 215), but the voicing of the medial dental stop would represent an irregular development.

<sup>7</sup> In Robbeets (2017), I mistakenly analyzed Ainu as lacking an inclusive-exclusive distinction on first person plural pronoun.