Chapter 36

Basic vocabulary in the Transeurasian languages

Martine Robbeets

Abstract

This chapter addresses one of the major objections raised against the genealogical relationship of the Transeurasian languages, notably the paucity of basic vocabulary in both quantity and quality. In contrast to the frequently uttered concern that the Transeurasian languages do not have enough basic vocabulary in common, it shows that there are 93 etymologies for 64 different concepts on the Leipzig-Jakarta basic vocabulary 100 list. From the viewpoint of quality of the evidence, the etymologies display regular sound correspondences and empirically supportable semantic latitude. Providing evidence for the exclusion of borrowing and coincidence as an alternative account of the similarities, this chapter shows that inheritance is the most sensible explanation of the observed correlations between the Transeurasian languages.

Keywords: basic vocabulary, Leipzig-Jakarta 100 list, inheritance, genealogical relationship, borrowing, coincidence, Transeurasian

36.1 Introduction

Here I will address one of the major objections raised against the genealogical relationship of the Transeurasian languages, notably the paucity of basic vocabulary in both quantity and quality. Indeed, critics of Transeurasian affiliation frequently

refer to the low number of correlating basic vocabulary items as a whole as well as to the weak regularity of their sound correspondences.

About a century after Nicolaes Witsen (1692) and Phillip von Strahlenberg (1730) mooted the contours of the Transeurasian language family, the idea became criticized by the French sinologist and medical doctor Abel-Rémusat (1820), who was the first to explicitly use the lack of basic vocabulary as a counter-argument. His main criticism was that the words correlating between the Turkic, Mongolic and Tungusic languages seemed to be the result of a longstanding political dominance of Turkic speakers, while the basic vocabulary shared between these languages—in Abel-Rémusat's wording "les mots qui designent des objets de première nécessité"— was radically different.[1] Following in Abel-Rémusat's footsteps but seemingly unaware of the basic vocabulary list published by Swadesh (1955) in the preceding year, Clauson (1956: 182) made his case against the so-called "Altaic theory" arguing that "[t]he basic words, that is the numerals, the basic verbs like "to say, to give, to take, to go" and so on, the basic nouns like "food, horse", and the basic adjectives like "good, bad" are all entirely different." Doerfer (1988: 155–156) rejects Clauson's concept of basic vocabulary, stating that "Mit vollem Recht haben sich ... die Vertreter der altaistischen Verwandtschaftsthese durch Sir Gerards Liste nicht erschüttern lassen...", and moves the basic vocabulary argument to a new plane. Focussing on terms for body parts, he argues that the correlations shared between the Turkic, Mongolic and Tungusic languages can be attributed to borrowing because these languages share more core than non-core body part terms. Ramer et al. (1997) reject Doerfer's conclusion by his own criteria, showing that there are more three-way cognates among the Altaic core than among the non-core body part terms.

The argument that the Altaic languages did not sufficiently share basic vocabulary was further tackled by Sergej Starostin and his colleagues, who added evidence from Japonic and Koreanic languages and created a massive etymological corpus, including 144 different etymologies for 100 basic vocabulary items (Starostin et al. 2003: 230–234). However, these new matches were, in their turn, criticized for reason of phonological, morphological or semantic overpermissiveness, among others by Robbeets (2005), leaving room for a reduced core of reliable etymologies, and by Vovin (2005c), completely rejecting all evidence advanced so far.

A basic vocabulary list is a compilation of concepts that are relatively independent of cultural context and available across the languages of the world and, therefore, can be used for testing the stability of historical-comparative linguistic evidence. The strength of the argument mainly lies in the fact that words with basic meanings tend to resist borrowing more successfully than random lexical items. The basic vocabulary list most commonly used in historical linguistics is the Swadesh 100 list (Swadesh 1955). Among the shortcomings of this list, we find the fact that it was put together on the basis of linguistic intuition and that it contains mostly nouns and too few verbs. For languages where verbs are basic to word formation and many nouns are derived from verbs, such as the Transeurasian languages, a standard Swadesh list produces too few useful comparanda.[2] Therefore, the Swadesh list has been recently updated by the Leipzig-Jakarta list (Tadmor et al. 2010), which takes a more systematic and empirical approach and partly remedies the imbalance in the vocabulary. Although 62 items on the lists overlap, the difference between both lists mainly consists in the replacement of a good number of nominal concepts by adjectival and verbal concepts in the Leipzig-Jakarta list. Moreover, on the basis of a solid cross-linguisic

investigation, the Leipzig-Jakarta list takes into account factors other than low borrowability such as the degree to which the meanings are universal, the degree to which the words are simplex and the probability of attrition.

In this chapter, I will advance Transeurasian etymologies for basic vocabulary items. In Section 36.2, I will present an overview of the evidence, paying attention to the regularity of sound correspondences. In Section 36.3, I will argue why the compared words are unlikely to be the result of borrowing. In Section 36.4, I will exclude chance similarity as a possible explanation for the correlations. Finally, I will conclude this chapter in Section 36.5. In the supplementary information to this volume, available on the following URL (www.xxx), I provided the full datasets underlying my reconstructions.

36.2 Overview of the basic vocabulary of Transeurasian

The Transeurasian basic vocabulary, summarized in Table 36.1, consists of lexical etymologies supporting the reconstruction of a Proto-Transeurasian form with a meaning that belongs to the Leipzig-Jakarta 100 basic vocabulary list. There are 93 etymologies spread over 64 distinct concepts of the basic vocabulary list. As there are no reliable cognates in Transeurasian for the following basic items, they yield empty rows and are therefore left out from Table 36.1: 2 nose, 6 tongue, 10 root, 13 rain, 15 name, 17 wing, 18 flesh/meat, 20 fly, 21 night, 23 neck, 28 bitter, 29 to say, 34 who?, 35 3SG pronoun, 38 horn, 41 yesterday, 42 to drink, 44 navel, 47 back, 48 wind, 49 smoke, 52 egg, 57 good, 59 knee, 60 sand, 61 to laugh, 64 leaf, 66 liver, 71 ant, 72 heavy, 75 to eat, 76 thigh, 83 eye, 85 tail, 89 to see and 93 bird.

The reconstructions in Table 36.1 are based on lexical data from 54 contemporary and 6 historical varieties of the Transeurasian languages, including 23 Turkic languages in addition to Old Turkic, 10 Mongolic languages in addition to Written Mongolian and Middle Mongolian, 10 Tungusic languages in addition to Manchu and Jurchen, Korean in addition to Middle Korean and Japanese and 10 Ryukyuan languages in addition to Old Japanese. In the online supplementary information (SI 2), I added a detailed documentation of all basic vocabulary comparative sets in support of Transeurasian affinity.

The reconstruction of the individual proto-languages is based on the scientific literature in the field of Turkic, Mongolic, Tungusic, Koreanic and Japonic reconstruction, combined with personal insights. In the online supplementary information (SI 1), I provided correspondence sets comparing contemporary and historical varieties of these languages, in order to infer the reconstruction of the basic phoneme inventory of the uncontroversial families. In case my personal reconstruction of an individual proto-form deviates from the standard reconstruction proposed in previous literature, I explicitly motivate my choice in the explanatory text added to the relevant etymology in SI 2.

The reconstruction of the basic phoneme inventory for Proto-Japonic is based on Martin (1987) with integration of the revision of the vowel system proposed by Frellesvig and Whitman (2008). The Ryukyuan correspondences involved in the reconstruction of Proto-Japonic are based on research by Thorpe (1983) and Pellard (2013, forthcoming). The phonological correspondences of Proto-Koreanic to Late Middle Korean and Contemporary Korean are based on Martin (1996) and Lee and Ramsey (2011). For the reconstruction of the Tungusic basic phoneme inventory, I

refer to Cincius (1949b), Benzing (1955a) and Starostin et al. (2003). The Mongolic sound correspondences and reconstructions are based on Poppe (1955), Starostin et al. (2003) and Nugteren (2011). For the phonological reconstruction of Turkic, I refer to Starostin et al. (2003). The reconstruction of proto-forms in this chapter does not necessarily confirm all reconstructions proposed elsewhere by other authors in this reference guide.

I collected the dataset by consulting dictionaries in combination with written sources against the background of etymologies proposed in the past by various linguists and evaluated in Robbeets (2005). In this monograph, I constructed a database of previous etymological proposals, which I submitted to an item-by-item evaluation in order to reach a restricted core of reliable evidence in support of the genealogical relationship between Japanese and the other Transeurasian languages. The dataset presented here is updated in the sense that I expanded my previous dataset with etymologies for basic items that do not have a Japanese cognate, consistently added cognates from the Ryukyuan languages, expanded the Tungusic cognates, carried out a detailed morphological analysis to delimit the roots more precisely and to identify petrified suffixes more accurately and, answered to some of the criticism in reviews of earlier etymologies.

Table 36.1 Summary of the basic vocabulary comparative sets in support of Transeurasian affinity

LJ item	Proto-Japonic	Proto-Koreanic	Proto-Tungusic	Proto-Mongolic	Proto-Turkic	Sound Corr. no	Sem. Corr. no

1 fire	*pɨ(r)i 'fire'	*pɨl 'fire'				1, 37, 31	68/7
3 to go	*na- 'go away'	*na- 'go out'	*-na:- 'go out to'			27, 32	56/4
4 water		*mɨl 'water'	*mö: ~ mu: 'water'	*mören 'river'		25, 37, 29	65/9
5 mouth	*kutɨ-i 'mouth, hole, opening'	*kut 'cavity, hole'				13, 38, 8	41/8 84/5
			*amga < *ʔama-g 'mouth'	*ama-n 'mouth, opening'		41, 26, 41	41/8
7 blood	*ti 'blood, spirit'			*či < *ti 'blood'	*tï:n 'spirit, breath'	7, 40	228/4 Poly
8 bone	*pəni(C)a 'bone'	*peCi 'bone'	*peni- 'knee'			1, 34, 28, 40	157/6 NO
9 2sg pronoun	*na '2 sg pronoun'	*ne '2 sg pronoun'				27, 33	250/2
			*si '2 sg pronoun'		*si '2 sg pronoun'	23, 40	250/2
11 to come	*kə- 'come'				*kel- 'come' [L] [SEP]	13, 34	161/3
12 breast	*kiki-rə 'heart'		*xökö-n 'breast'	*kökö-n 'breast'	*kökü-r₂ 'breast'	21, 37, 14	34/6 Poly
14 1sg pronoun	*wa-n- '1sg/pl pronoun'			*ba-n- '1pl excl. pronoun'		3, 32	87/4
			*bi 'l sg pronoun'	*bi 'l sg pronoun'	*bi 'l sg pronoun'	3, 40	87/4
16 louse				*sir-ke 'louse'	*sir-ke 'nit'	23,40, 18,33	205/3
19 arm/ hand	*ta(r)i 'upper limb, arm,	*tali 'lower limb, leg'				19, 32, 31, 40	38/8 NO

	hand'						
	*sune 'lower limb, leg'	*son 'upper limb, arm, hand'				23, 36, 28	38/8 NO
				*gar 'hand, arm'	*karï 'arm'	15,32, 29	38/8
22 ear				*kul 'ear'	*kul-kak 'ear'	13, 39, 31	186/4
24 far	*mara 'rare, from afar'	*melu- 'be far'				25, 33, 31	17/7
25 to do/make [1 SEP]	*-ka- iconic	*-ki- iconic	*-ki- iconic	*ki- 'do, make'	*kïl- 'do, make'	13, 40	33/6 Gram
26 house	*(y)ipi 'house, hut'	*cip(i) 'house'	*ji:b 'house'			9, 47, 4, 47	50/8
27 stone/ rock			*kada: 'rock, cliff'	*kada 'rock, cliff'		13, 32b, 10, 32b	162/4
30 tooth	*pa 'tooth'	*pal 'tooth'				1, 32	41/8
				*ari-ga 'molar, canine, fang'	*ar$_2$(-)ïg 'molar, fang'	32, 30	41/8
31 hair	*kama ~ kami 'hair of the head'	*kama 'whirl of hair on the head'				13, 32, 26	1/16
	*kara ~ ka(r)i 'hair'	*kal 'hair'				13, 32, 29	1/16
				*kilga 'coarse hair'	*kïl(k) 'hair'	13, 40, 31	1/16
32 big	*kɨkɨ- 'be large in amount'	*hiki- 'big'			*kök 'big, thick, healthy'	13, 37, 14, 37	188/3
	*mana- 'be big, many'		*mani 'large group'	*man 'big, high'	*banɨ 'big'	25, 32, 28	188/3

			*amban 'big'	*amban 'big'		41, 6, 32, 28	188/3
33 one SEP		*pili- ~ *pilʌ- 'to begin'			*bir 'one'	3, 40, 31	235/2 109/4
36 to hit/beat	*tuk- 'hit with force'	*t(ʌ)ki- 'hit, strike'	*dug- 'hit'			9, 39, 16	85/5
37 leg/foot	*panki 'lower leg, foot'	*pal 'foot, leg' PK *pʌlk 'arm'	*palgan 'foot'			1, 32, 18	15/10
39 this	*i 'you' (derogatory 2sg)	*i 'this' (proximal demonstrative)	*i 'he, she, it' (3sg)	*i 'he, she, it' (3sg)	*ï(-)n- 'that' (distal demonstrative)	40	Gram
	*kɨ- 'this' (proximal demonstrative)	*kɨ- 'that' (medial demonstrative)			*kö 'this' (proximal demonstrative)	13, 37	Gram
	*ə- 'that' (distal demonstrative)		*e- 'this'	*e- 'this'		42	Gram
40 fish	*(y)iwə 'fish'			*diɣa- 'fish'		9, 40, 4, 34	67/4
43 black				*kara 'black'	*kara 'black'	13, 32, 29, 32	63/5
45 to stand	*tata- 'stand, rise, run high'	*tʌtʌ- 'run'				7, 32b, 8, 32b	35/6 NO
		*ilʌ- / *ilɨ- 'come up, rise'	*ili- 'stand (up), rise'			40, 18	32/5
46 to bite	*kam- 'bite, chew'			*keme- 'bite, gnaw'	*kem- 'bite, chew, gnaw'	13, 33, 26	233/3
		*mɨli- 'bite'	*mödö- 'gnaw, bite'			25, 37, 10, 37	233/3
50 what?	*ka	*ka	*xa-		*ka-	21, 32	93/6

	wh-interrogative particle	interrogative particle	wh-interrogative pronoun		wh-interrogative pronoun		Gram
51 child	*wara-pa 'child'				*ba:la 'young animal, child'	3, 32, 31, 32	6/12
			*puril 'child'	*pure 'child, seed'		1, 39, 29	6/12
53 to give	*tama- 'give'		*tama- 'pay'			7, 32, 26, 32	10/13
	*(w)ura- 'sell'	*pʌlʌ-kʌ- 'sell'	*bu:- 'give'			3, 39, 30	10/13
54 new	*ara- 'new, pure'			*ari- 'be pure'	*arï- 'be(come) pure'	41, 29	6/12 Poly
55 to burn (intr.)	*tak- 'burn (tr.)'	*tʌkʌ- /*taki- 'be on / set fire'			*yak- 'burn (tr.)'	9, 32, 14	104/4
56 not SEP	*ana- negation	*an- negation	*ana- negation		*an- 'be unbecoming'	41, 28, 32	253/2 Gram
			*e- negation	*e-se negation	*e- negation	42	253/2
58 to know SEP		*alʌ- 'know'	*ala- 'make known, know'			41, 31	11/7
62 to hear	*uka- 'receive, perceive, hear'			*uka- 'understand'	*uk- 'understand , hear'	46, 14, 32	11/7
63 soil SEP	*tutɨ-i 'soil, ground'	*tutɨ 'bank, ridge, ground'				7, 38, 8	2/9
65 red		*pɨl-kɨ- 'be red'	*pula- 'be red'	*pula-ɣan 'red'		1, 39b, 31	288/4
67 to hide		*sum- 'hide, lurk in'	*sume- 'hide, conceal'			23, 38, 26	222/4

68 skin/hide	*kapa 'skin, bark, shell'	*kap(ʌ)-k 'skin, bark outer layer'			*ka:p-ïk 'bark, shell'	13, 32, 2	1/16
69 to suck			*xökö- 'to suck (breasts)'	*kökü- 'to suck (breasts)'		21 37, 14	191/5
70 to carry	*əpə- 'carry on back'	*ep- 'carry on back'	*ebe- 'carry'			34, 4, 34	23/8
73 to take	*tira- 'take, hold'	*tɨlɨ- 'hold up, lift, raise'				7, 37, 29	3/8
			*al- 'take'		*al- 'take'	41, 31	3/8
74 old	*muka-si 'be long ago, ancient'	*muk- 'be(come) old'				25, 38, 14	62/6
				*kari- 'weaken'	*karï- 'be(come) old'	13, 32, 29, 40	62/6 NO
77 thick	*puta- 'be thick'	*putʌ- 'become thick'		*büdü- 'large'		3, 38, 10	188/3
78 long	*nanka- 'be long'	*nʌlkʌ- 'be(come) old, long (in time) '				27, 32b,1 8, 32b	17/7
				*uri 'long past, former'	*ur₂a- 'be long (time/ space)'	46, 29	17/7
		*ola- 'last long'		*ora- 'be late'		43, 29, 32	17/7
79 to blow		*pulɨ- 'blow'	*pu:- 'blow'			1, 38	158/5
80 wood	*kɨ 'tree, wood'	*kɨl 'tree, wood'				13, 37	4/13
	*moro 'woods, wooded hill'	*molo 'hill, mountain'	*mo: 'wood, tree'	*mo wood, tree'		25, 35	4/13 Poly
81 to run	*pasa- 'run'	*pas- 'hurry'				1, 32, 24	80/4 NO
82 to fall	*tira- 'fall, scatter'	*ti- 'fall, scatter'				7, 40	141/5

84 ash	*papV 'ash'	*pap 'dust, waste'				3, 31, 2	66/8
86 dog	*inu 'dog'		*ina 'dog'			47, 28	230/6
87 to cry/weep [1]		*uli- 'cry, howl'		*uli- 'howl'	*u:lï- 'cry, howl'		21/9
88 to tie	*kuku- 'tie, wrap'		*xuku- 'wrap'			21, 39, 14, 39	229/3 Poly
				*boɣo- 'tie, bind, bundle'	*bog- 'tie, strangle'	3, 36, 16	229/3
90 sweet	*ma- 'to be tasty, sweet'	*ma- 'to be tasty'				25, 31	
		*tʌlʌ- 'to be sweet'	*da:l- 'to be sweet'			9, 32b, 31	60/6
91 rope	*tura 'rope, string, line'	*cul 'rope, string, line'				19, 38, 31	55/8
92 shade/ shadow	*kan(a)-ka 'reflection, shade'	*kʌnʌl-ʌk 'shadow'				13, 32b, 28, 32b	43/7
94 salt	*tura- 'bitter, unbearable'				*tu:r₂ 'salt'	7, 39, 30	60/6
95 small	*tipi- 'to be small'		*čipi- 'to be small, narrow'			19, 40, 2	99/4
96 wide	*nǝnpa- 'become wide and long'	*nelpʌ- 'be wide'	*nepte- 'become flat and wide'	*nebse- 'be wide and long'		27, 34, 6, 33	188/3
				*dalba- 'to be wide and flat'	*yalpa- 'to be wide, flat'	9, 31, 6	188/3

97 star [SEP]	*pəsi 'star'	*peli 'star'				1, 34, 20b, 40	267/2
98 in	*soko 'depth'	*soko 'depth, deep inside'				23, 35, 14	36/6
				*örü 'interior'	*öːr₂ 'interior'	44, 30	36/6
99 hard [SEP]	*kata- 'be hard'	*kata- 'be hard'		*kata- 'become hard'	*kat- 'be hard'	13, 32, 8, 32	70/6
100 to crush/ grind	*pinta- 'crush'		*pinče- 'crush'			1, 40, 12, 33	58/7
	*sura- 'grind, rub, make smooth'				*sür- 'rub, smear'	23, 38, 29	58/7 Poly
		*niki- 'crush, knead'		*niku- 'crush, knead'	yïk- 'crush'	27, 40, 22	

The evidence in Table 36.1 corresponds regularly in form and function to such an extent that it yields Proto-Transeurasian reconstructions for items that occur on the Leipzig-Jakarta 100 basic vocabulary list. The penultimate column attributes a number to each regular sound correspondence, which refers to the inventories of consonant correspondences in Table 36.2 and vowel correspondences in Table 36.3. Gray shading marks the cells in which the reconstructed word fulfills the formal requirements for at least a subsequent initial consonant, medial vowel and medial consonant, but in several cases the occasional root-final vowel is corresponding regularly as well.

The final column of Table 36.1 examines the semantic development reconstructed between Proto-Transeurasian and the daughter branches against cross-linguistically observed polysemies and semantic associations. To this end, I consulted the database

of cross-linguistic colexifications published by List et al. (2014), which brings together instances where two or more meanings are simultaneously covered by the same lexical item in a certain language. The number in the final column corresponds to the number of the relevant semantic community in List et al. (2014), while the number behind the slash is the number of nodes in that community. The tag "NO" in a few cases means that the proposed development is not supported by the semantic associations given in List et al. (2014). The tag "Gram" means that the development is a cross-linguistically well-attested grammaticalization process. This is for instance the case for the development of the verb (25) 'to do, make' in an iconic pro-verb following sound symbolic expressions (Heine and Kuteva 2002: 112–113), the development (39) from a demonstrative into a personal pronoun, (Heine and Kuteva 2002: 119–120) and the development of the interrogative pronoun (50) 'what?' into an interrogative particle.

Finally, the tag "poly" marks a semantic development that is supported by a polysemy in one or more Transeurasian languages of the dataset, even if it is not taken into account in List et al. (2014). For instance, under item (12) 'breast', I compared PJ *kiki-rə 'heart', although this semantic association is not mentioned by the authors. If this form indeed incorporates a petrified plurality marker PJ *-rə of the type found in among others OJ ko_1-ra 'children', $woto_2me_1$-ra 'young girls', ye-ra 'branches', kinu-wata-ra 'silk clothes' (Antonov 2007: 195, 197), then the plurality can be taken as indicative of a semantic shift from 'breasts' to 'heart'. Note that this semantic shift and morphological segmentation is supported by a similar development on a different etymon, namely Yonaguni ccimu 'heart, liver' that is derived as YG ccimuti 'breast', probably through addition of a petrified Ryukyuan dual/collective

suffix *-*ti*, also present in YG *khata-buru-ci* 'shoulder', Yamatohama (Amami) *xama-ci* 'hair', etc.

Similarly, for item (80) 'wood', List et al. (2014) do not mention a possible colexification between 'woods' and 'mountain' occurring in their sample of languages across the world. However, as the colexification is attested in the contemporary Miyako Ryukyuan form Irabu *muï* 'forest, mountain', the suggested semantic development of 'woods' to 'mountain' is acceptable. This is further supported by the meaning of MK *mwo·lwo* 'mountain', which is restricted to compounds with tree names, e.g. MK *·phi mwo·lwo* 'Torreya-nut Mountain' and thus suggests that the meaning really amounts to 'wooded mountain'. This can be seen as an intermediary stage in the development between 'woods' and 'mountain'. Other instances of polysemy that are taken as indicative of the proposed semantic development involve OJ *ti* 'blood, force, spirit' under (7) 'blood', OJ *ara*- 'to be rough, fresh, pure, new' under (54) 'new', J *kukum*- 'to tie up, wrap up' under (88) 'to tie' and OJ *sur*- 'to grind, rub' under (100) 'to crush, grind'. Using the principle of colexification in this way, I provide an empirical base for the degree of semantic latitude permitted in the Transeurasian comparisons.

Table 36.2 Consonant correspondences between the Transeurasian languages

	PJ	PK	PTg	PMo	PTk	PTEA
1.	*p-	*p-	*p-	*p-	*b-	*p-
2.	*-p-	*-p-	*-p-	*-ɣ-	*-p-	*-p-
3.	*p- / *w-	*p-	*b-	*b-	*b-	*b-
4.	*-p-/*-w-	*-p-	*-b-	*-b-/ -ɣ-	*-b-	*-b-
5.	*-np-	*-pC-	*-PC-	*-PC-	*-P(C)-	*-m$^{(P)}$T-
6.	*-np-	*-Rp-	-RP-	*-RP-	*-RP-	*-Rp-
7.	*t-	*t-	*t-	*t-	*t-	*t-
8.	*-t-	*-t-	*-t-	*-t-	*-t-	*-t-
9.	*t- /*y-	*t- (ci-)	*d- (ji-)	*d- (ji-)	*y-	*d-

10.	*-t-/ *-y-	*-l-	*-d- (-ji-)	*-d- (-ji-)	*-d-	*-d-
11.	*-nt-	*-c-	*-TC-	*-TC-	*-TC-	*-n$^{(T)}$K-
12.	*-nt-	*-Rc-	*-RT-	*-RT-	*-RT-	*-Rt-
13.	*k-	*k-	*k-	*k-	*k-	*k-
14.	*-k-	*-k- (-h-)	*-k-	*-k-	*-k-	*-k-
15.	*k-	*k-	*g-	*g-	*k-	*g-
16.	*-k-	*-k- (-h-)	*-g-	*-g-	*-g-	*-g-
17.	*-nk-	*-kC-	*-KC-	*-KC-	*-KC-	*-ŋ$^{(K)}$T-
18.	*-nk-	*-Rk-	*-RK-	*-RK-	*-RK-	*-Rk-
19.	*t-	*c-	*č-	*č-	*č-	*č-
20.	*-t-	*-c-	*-č-	*-č-	*-č-	*-č-
20b.	*-si	*-l(i)	*-l(č)	*-l(č)	*-l(č)~ -š	*-lč
21.	*k-	*k-, h-	*x-	*k-	*k-	*x-
22.	*-k-	*-k-	*-x-	*-g-~-k-	*-g-~-k-	*-x-
23.	*s-	*s-	*s-	*s-	*s-	*s-
24.	*-s-	*-s-	*-s-	*-s-	*-s-	*-s-
25.	*m-	*m-	*m-	*m-	*b-	*m-
26.	*-m-	*-m-	*-m-	*-m-	*-m-	*-m-
27.	*n-	*n-	*n-	*n-	*y-	*n-
28.	*-n-	*-n-	*-n-	*-n-	*-n-	*-n-
29.	*-r-	*-l-	*-r-	*-r-	*-r-	*-r-
30.	*-r-	*-l-	*-r-	*-r-	*-r$_2$-	*-r-
31.	*-r-	*-l-	*-l-	*-l-	*-l-	*-l-

Table 36.3 Vowel correspondences between the Transeurasian languages

	OJ < PJ	MK < PK	PTg	PMo	PTk	PTEA
32.	-a- < *-a-	-a- < *-a-	*-a-	*-a-	*-a-	*-a-
32b.	*CaCa	*CʌCʌ	*CaCa	*CaCa	*CaC	*CaCa
33.	-a- < *-a-	-e- < *-e-	*-e-	*-e-	*-e-	*-ə-
34.	-o- < *-ə-	-e- < *-e-	*-e-	*-e-	*-e-	*-ə-
35.	-o- <? *-o-	-wo- < *-o-	*-o-	*-o-	*-o-	*-ɔ-
36.	-u- < *-o-	-wo- < *-o-	*-o-	*-o-	*-o-	*-ɔ-
37.	-o- < *-i-	-u- < *-i-	*-ö-	*-ö-	*-ö-	*-o-
38.	-u- < *-u-	-wu- < *-u-	*-u- (gü)	*-ü-	*-ü-	*-u-
39.	-u- < *-u-	-o- < *-ʌ-	*-u-	*-u-	*-u- /-ï-	*-ʊ-
39b.	PaRu- < *PauRu-	*PʌRʌ- ~ *PiRi-	*PuRu-	*PuRu-	*PuR-	*PʊRʊ-
40.	-i- < *-i-	-i- < *-i-	*-i-	*-i-	*-i-/-ï-	*-i-
41.	a- < *-a-	a- < *a-	*a-	*a-	*a-	*a-
42.	o- < *ə-	e- < *e-	*e-	*e-	*e-	*ə-

43.	o- <? *o-	wo- < *o-	*o-	*o-	*o-	*ɔ-
44.	o- < *i-	ø < ? *i-	*ö-	*ö-	*ö-	*o-
45.	u- < *u-	wu- < *u-	*u-	*ü-	*ü-	*u-
46.	i- < *i-	i- < *i-	*i-	*i-	*i-	*i-

36.3 Arguments against borrowing

36.3.1 Basic vocabulary

How can we exclude the possibility that the comparative sets in Table 36.1 are the result of borrowing? Obviously, as is true for most argumentation in historical-comparative linguistics, we cannot provide concrete and conclusive proof that the statement "none of the correlations in Table 36.1 is borrowed" is true. Nevertheless, we can build a case against borrowing, gathering pieces of evidence that suggest that the most sensible explanation is that "not all correlations in Table 36.1 are borrowed". Traditionally, the strength of basic vocabulary lies in the fact that words with basic meanings tend to resist borrowing more successfully than random lexical items. The stability of the basic vocabulary is not based on the assumption that every single basic item is absolutely copy-free, but it consists in the low borrowability of the entire body of basic concepts as a whole. The very fact that we find 93 Transeurasian etymologies covering 64 distinct basic vocabulary concepts thus is a strong argument against borrowing by itself.

36.3.2 Borrowing hierarchy

Among the concepts of the Leipzig-Jakarta list, we find 25 actions, 15 property words, 10 deictic or grammatical items and 50 nominal concepts including natural phenomena, body parts, animal or plant parts, children of humans and animals and cultural words. Out of 40 concepts for actions and property words, we find 30 Transeurasian verbal etymologies, which means that as much as 75% of the basic verbal concepts on the Leipzig-Jakarta list are etymologized. Out of 10 concepts for deictic and grammatical items, we find 7 etymologies, which implies that 70% of the basic deictic and grammatical items on the list are etymologized. Out of 50 nominal concepts, we find 27 etymologies, indicating that only 54% of the nominal concepts are covered by a Transeurasian etymology.

   A general assumption is that a word class or a part of language structure is more likely to be borrowed if it is borrowed more frequently in cross-linguistic sampling. Empirically it is observed that languages tend to borrow lexical items more easily than grammatical ones and nouns more easily than verbs (e.g. Weinrich 1953; Moravcsik 1978; Thomason and Kaufman 1988; Muysken 2000; Wohlgemuth 2009; Matras 2009; Tadmor et al. 2010). From the seventeenth to the nineteenth century, for instance, Japanese underwent intensive contact from Dutch leading to the global copying of over 300 words and the selective copying of syntax, but Japanese did not globally copy a single verb nor a concrete grammatical marker from Dutch (Irwin 2011). The relatively low borrowability of verbs and grammatical markers is interrelated with a number of factors, such as the fact that their meanings tend to be less concrete and culturally less determined than those of nouns, that they are less perceivable as a distinct unit because they need more adaption to the morpho-

syntactic frame of the sentence, and that there simply are less verbs and grammatical markers than nouns.

In contrast to this tendency, there are significantly more correlations for verbs (75%) and deictic and grammatical items (70%) in the Transeurasian basic vocabulary than for nouns (54%). This observation indicates that it is unlikely that the comparative sets can be explained by borrowing, as borrowing would be expected to yield more correspondences in nouns than in verbs and grammatical markers.

36.3.3 Typology of verbal borrowing

As far as the mechanisms of loan verb accommodation are concerned, most recipient languages can be categorized into two distinct groups: borrowed verbs either arrive as verbs, needing no formal accommodation, or, they arrive as non-verbs and need formal accommodation. In Wohlgemuth's (2009) terminology, the first group represents "direct insertion", while the second group represents either "indirect insertion", when the formal accommodation involves a verbalizer or else, "light verb strategy", when the borrowed verb is integrated into a complex predicate. Most Transeurasian languages can be assigned to the second group because they display a clear preference for the non-verbal strategy (Wohlgemuth 2009: 159, 161); for instance, Turkish *klik-le-* and *klik et-* << English *click*; Khalkha *zee-l-* << Mandarin *zhài* 'borrow, lend'; Korean *coking ha-*, J *zyogingu suru* 'to jog' << English *jog;* J *demo-r-* << English *demonstrate.* Whereas the northern Tungusic languages prefer to borrow verbs through direct insertion, e.g. Evenki *vypolńaj-* << Russian *vypolnja-t'* 'to fulfill, carry out', the southern Tungusic languages use verbalizers, e.g. Udehe *tancewa-la-* << Russian *tancewa-t'* 'to dance' and Nanai *gotovi-la-* << Russian

gotovi-t' 'to cook'. If the 30 Transeurasian verbal etymologies in Table 36.1 would be the result of borrowing, they would represent instances of "direct insertion". This would run against the observable preference of the Transeurasian languages to apply the non-verbal strategy to verbal copies. Therefore, it is unlikely that the 30 comparative sets for actions and property words all represent borrowings.

36.3.4 Regularity and complexity of sound correspondence

The comparative sets for basic vocabulary summarized in Table 36.1 display regular correspondences for each consonant of the verb root and for each but the root-final vowel, conform to the requirements in Table 36.2 and 36.3. Phonology can help to unmask borrowing, even if extensive contact can result in strata of loanwords that exhibit systematic sound correspondences. The stratum of loanwords from Middle Chinese that has entered both Japanese and Korean during the Tang period (618–906 AD) is known as Sino-Japanese and Sino-Korean and these strata display regular sound correspondences with Middle Chinese and thus also with each other. Thus, phonological correspondences may occur between different strata of loanwords as the result of borrowing from a common model language.

As opposed to imitation of the model sounds, the phonological correspondences between cognates are expected to reflect divergence. An example of a sound correspondence that is difficult to explain by borrowing is found in the cluster correspondence in (96) 'wide'. The original cluster correspondences in Table 36.2 can be divided into homorganic and heterorganic clusters. Homorganic clusters are composed of a sonorant and a stop (PTEA *-*Rp*-, *-*Rt*-, *-*Rk*-) and merge in a nasal cluster (PJ *-*np*- > OJ -*b*-, PJ *-*nt*- > OJ -*d*-, PJ *-*nk*- > OJ *-*g*-) in Japonic. By

contrast, in heterorganic clusters, as illustrated in (96) 'wide', the nasal and the stop have a different place of articulation, which results in the insertion of a parasitic stop (PTEA *-$m^{(P)}T$-, *-$n^{(T)}K$-, *-$ŋ^{(K)}T$-). The nasal is lost in the continental Transeurasian languages (*-PT-, *-TK-, *-KT-), whereas Korean and Japanese lose the final stop (PJ *-mp- > OJ -b-, PJ *-nt- > OJ -d-, PJ *-$ŋk$- > OJ *-g-.) Such a complex correspondence cannot easily be explained by mere imitation of sounds through borrowing.

36.3.5 Broken contact chain

Borrowing is typically unidirectional and linear, progressing from one contact language into the other and then, perhaps, into the next. Genealogical divergence, by contrast, can be pictured as the rings formed when a stone is thrown into the water: innovations start in the center and push the older forms towards the periphery. This observation explains why some very conservative inherited items leave traces in remote areas, but are barely attested elsewhere in the linguistic continuum. Thus, gaps in the attestation of members of an etymology may be relevant. When the contact chain is broken, a genealogical explanation presents itself. The absence of a corresponding item in one or more intermediate contact branches can be observed in (7) 'blood', (9) '2 SG pronoun', (11) 'to come', (12) 'breast', (14) '1 SG pronoun', (32) 'big', (33) 'one', (39) 'this', (40) 'fish', (46) 'to bite', (50) 'what?', (51) 'child', (53) 'to give', (54) 'new', (55) 'to burn', (56) 'not', (62) 'to hear', (68) 'skin/hide', (73) 'to take', (77) 'thick', (78) 'long', (86) 'dog', (87) 'to cry/weep', (88) 'to tie', (94) 'salt', (95) 'small', (99) 'hard' and (100) 'to crush/grind'.

36.3.6 Multiple setting

Most examples of borrowing have a binary setting in common: they typically go from a model language into a recipient language. Especially for verbs and grammatical markers, examples of the same item progressing into a third or fourth language are relatively rare. The greater the number of languages or language families involved in the comparison, the more likely it becomes that the similarity observed is due to inheritance. One exception is prestige settings where one language serves as a lingua franca or dominates many others. The English verb *to film*, for instance, has been exported to many other languages such as German / Dutch *filmen*, French *filmer*, Polish *filmować*, Greek *filmaro*, Hungarian *filmez*, Finnish *filmata* etc.

However, the Transeurasian unity consists of 5 families and, based on the archeological records, these do not reflect unilateral prestige relationship. Besides, the correlations under discussion concern basic vocabulary and are thus prototypically culture-independent. Nevertheless, the majority of Transeurasian basic vocabulary is shared over more than two branches. Some comparative sets such as (12) 'breast', (25) 'to do/make', (32) 'big', (39) 'this', (50) 'what?', (56) 'not', (80) 'wood', (96) 'wide' and 99 'hard' are even represented in as much as four or all five branches. Interestingly, it is especially grammatical markers and verbal concepts that are well distributed and preserved. Since repetitive borrowing is particularly rare for verbs and grammatical items, this observation argues against borrowing.

36.3.7 Absence of prototypical loan-characteristics of form and meaning

Even if it is improbable that basic items get borrowed, it is not impossible. However,

in such cases, there are still a number of characteristics that can betray a set of borrowed items, notably unilateral morphological complexity and secondary semantics. Borrowing is a likely explanation in cases when the similarity concerns a morphologically complex word in one language that cannot be analyzed as such in the other language. This is, for instance, the case for the correspondence between Manchu *amila-* 'to anoint a Buddhist icon's eyes with blood and thereby impart life to it' and Written Mongolian *amila-* 'to give life, enliven, animate an image by making strokes on a sacred image, come to life' (Rozycki 1994: 5, 17–18). The Mongolian verb is a denominal derivation from WMo. *ami(n)* 'life, breath' with the manipulative suffix WMo. *-lA-* (Robbeets 2015: 221–222). The derivation holds for Mongolian but not for Manchu because the basic nominal form is absent there. The unparalleled morphological complexity of the Mongolic form in Manchu is indicative of borrowing. Moreover, the observation that Manchu semantics is restricted to a secondary developed meaning in a Buddhistic cultural context further confirms the borrowing scenario of a complex Mongolian verb into Manchu. Comparanda displaying prototypical loan-characteristics in form and meaning have been eliminated from the evidence. For instance, under (73) 'take', I eliminated a possible Mongolic cognate **ali* 'give! (imperative)' reflected in WMo. *ali*, Khal. *aĺ*, *aliv*, Bur. *aĺe:*, Kalm. *aĺ*, *aĺə*, Ordos *ali*, Dgx. *ali*, Bao. *an* and Mgr. *ali* because the verb only occurs in one inflected form, namely the imperative, but does not display a complete verbal paradigm.

36.3.8 Well-spread distribution

If the distribution of a certain basic item would be restricted to a single language or to

only few languages of a certain subgroup, this could serve as an indication of borrowing. Therefore, I have eliminated corresponding forms that are poorly distributed within the individual branches of Transeurasian. Based on his hypothesis that West Old Japanese or its immediate predecessor absorbed a large number of loanwords from Old Korean, Vovin (2010) proposed to reject all cognates that are only attested in Western Old Japanese—and thus are missing in Eastern Old Japanese and the Ryukyuan languages—as probable loanwords. Although this procedure is not entirely sound, I paid particular attention to the attestation of Ryukyuan cognates.[3] Out of 59 Proto-Japonic forms in Table 36.1, 49 are supported by reflexes in the Ryukyuan languages, corresponding to 83% of all Proto-Japanic reconstructions. Therefore, the solid distribution of Japonic cognates in the Ryukyuan languages reduces the probability of borrowing from Old Korean at a time when Mainland Japanese had already separated from the Ryukyuan languages.

For some Proto-Japonic forms such as PJ *na '2 SG pronoun' and PJ *tuti 'soil, ground', Vovin (2010) explicitly mentions the paucity of the distribution of Ryukyuan reflexes. However, contrary to Vovin's (2010: 65) claim for PJ *na that "the distribution in the Ryukyus (no attestations in Sakishima) and its mild honorific nature suggest that Ryukyuan naa is a loan from Japanese," there is an attestation of this form in the Sakishima languages, notably Taketomi (Yaeyama) naara 'you, self'. Moreover, Vovin (2010: 124) argues that OJ tuti 'earth, soil' is a loanword from Korean because "there is only one isolated attestation in Ryukyuan: Ishigaki tsïtsï 'earth'". However, the forms for 'ground, earth' are well distributed across the Ryukyuan languages, e.g. Kin (Okinawa) sicii (B), Hirara (Miyako) ^dzï: 'ground', dzï: (B), Hatoma (Yaeyama) sici (B), Ishigaki (Yaeyama) cïcï (B), Yonaguni di: (B), all

meaning 'ground'. In addition, these forms all have accentuations, which are consistent with the Mainland Japanese accentuation, so we can presume they are not borrowings from the Mainland.

36.4 Arguments against coincidence

In addition to the regular correspondences in the basic vocabulary discussed above, I previously established systematic correlations in the lexicon in general (Robbeets 2005) and in morphology (Robbeets 2015; this volume: Chapter 30) on the basis of the classical comparative method. However, I did not make explicit chance calculations until the research in Robbeets (submitted). In this paper, I calculated the number of comparative sets needed to exceed chance when five proto-languages are compared and tested to what extent the lexical evidence advanced in support of Transeurasian affinity meets these criteria. The method is based on a revision of Nichols' (2010) method for determining the number of comparative sets required in a binary comparison, which I expanded to a multiple setting, by focusing on the binary comparisons implied in the multiple sets. Arguing that the genealogical relatedness between Koreanic and Japonic, Japonic and Tungusic, Tungusic and Mongolic, and Mongolic and Turkic can be probabilistically confirmed, I logically concluded that the same is true for the Transeurasian family as a whole.

In order to confirm genealogical relatedness between two proto-languages we must show that the number of observed similarities significantly exceeds the number of random similarities expected to occur by chance. The similarities in Table 36.1 consist in sound matches between basic vocabulary items of roughly equivalent meaning. When comparing such items, we basically have 2 possible outcomes:

"success" or a sound match and "failure" or a sound mis-match. Therefore, we can approach this problem using a binomial distribution. The cumulative binomial distribution indicates the probability of observing at most x successes in n trials with the probability of success on a single trial denoted by p.

The p in our formula is the probability of finding a phonological match by pure chance. I estimated it by multiplying the probabilities of finding a possible match for the consonant, vowel and consonant in a CVC sequence in another proto-language and found a value of 0,0060. The n in our formula is the number of trials involved in comparing 2 100-word lists. I calculated it by estimating the number of trials involved in comparing two 100-word lists and reached a value of 699 trials. Starting from these values for n and p, the cumulative binomial distribution indicates that we need 9 binary corresponding basic vocabulary items to exclude chance with 99% probability. The more comparative sets we have, the more certain we are that chance similarity can be excluded.

In Table 36.4, I listed the Leipzig-Jakarta numbers of the comparative pairs that are marked in gray shading in Table 36.1 because they have a correspondence for at least each subsequent consonant, vowel and consonant of their root. In the third column, I counted the number of comparative pairs available for each pairwise set of proto-languages. Applying the formula of the cumulative binomial distribution, I calculated the probabilities that the observed correlations are not coincidental in the final column.

Table 36.4 Binary probabilities calculated on the basis of the number of comparative pairs within the basic vocabulary

	Basic vocabulary comparative pairs with CVC(V) or CVCC(V) match	Number of comparative pairs	Probability
Japonic and Koreanic	**5, 19b, 24, 31a, 31b, 32, 37, 45, 55, 63, 68, 73, 74, 77, 78, 80, 81, 84, 91, 96, 97, 98, 99**	**23**	**100% (1)**
Japonic and Tungusic	**8, 12, 32b, 36, 37, 53a, 88, 95, 96a, 100a**	**10**	**100% (0,9961)**
Japonic and Mongolic	12, 14, 32b, 46, 77, 96, 99	7	94% (0,9370)
Japonic and Turkic	**12, 32a, 32b, 46a, 51a, 55, 68, 94, 99, 100b**	**10**	**100% (0,9961)**
Koreanic and Tungusic	26, 37, 46b, 65, 67, 90, 96	7	94% (0,9370)
Koreanic and Mongolic	4, 65, 77, 96, 99, 100c	6	87% (0,8688)
Koreanic and Turkic	32a, 33, 55, 68, 99, 100c	6	87% (0,8688)
Tungusic and Mongolic	12, 32b, 51b, 65, 69, 96	6	87% (0,8688)
Tungusic and Turkic	12, 32b	2	21% 0,2102
Mongolic and Turkic	**12, 16, 19c, 22, 31c, 32b, 43, 46, 74b, 88b, 96b, 100c**	**12**	**100% (0,9996)**

As we reach nearly 100% certitude that the binary relationships between Japonic and Koreanic, Japonic and Tungusic, Japonic and Turkic and Mongolic and Turkic are not coincidental, the transitivity of family relationship leads to the conclusion that the correlations between the Transeurasian languages as a whole are probably not coincidental. Although the transition from binary to multiple probabilities is statistically more complex than the approach summarized here, the assumption that the Transeurasian correlations are not due to chance remains the most parsimonious one. Given the elimination of borrowings in Section 36.3, it is thus highly likely that the Transeurasian languages are genealogically related.

Even if it supports kinship for Japanese, Korean, and Tungusic, Brown's application (this volume: Chapter 42) of the Beck-Wichmann-Brown system leads to a different conclusion, notably that "binary comparisons sifted from Robbeets' 359 lexical sets provide no compelling support for the genetic unity of all five Transeurasian language groups." Although the Beck-Wichmann-Brown approach represents a laudable attempt at developing a more objective standard for the assessment of language classifications worldwide, it is not without problems. The approach proposed by Brown in this volume is ineffective because (1) it requires translational equivalence, (2) it surpresses evidence in calculating the mean standard against which the reliability of Transeurasian is measured and (3) it is based on an incomplete dataset of Transeurasian comparative pairs.

First, lexical change involves not only sound but also meaning. With a short elapse of time, we expect to find a large extent of translation equivalence between related languages, but at deeper time depths, meanings will have developed: they may still be similar, but not identical. Brown's requirement of semantic equivalence of the cognate sets is unrealistic and may even be taken as an indication of borrowing. The approach based on cross-linguistic colexifications proposed in Section 36.2 may provide a more sensible alternative to formalize the range of permissable semantic associations.

Second, Brown suggests that all language comparisons with COMi values larger than or equal to the COMi mean can be considered strong support for genealogical relationship, while comparisons with COMi values smaller than the mean should be considered less than strong support for genealogical association. This threshhold involves "cherry picking" as it is based on a selection of genealogical proposals

worldwide made by the author himself. Indeed, if we would pick only strongly supported or closely related families, the mean against which we measure Transeurasian will be high. In case we would select weakly supported or distantly related families, the mean against which we evaluate Transeurasian will be low. Moreover, the threshold does not distinguish between distant relatedness and weak relatability.

Finally, Brown's dataset consists of collections of paired comparative sets based on Robbeets' (2005) Transeurasian proposal. Since the key objective of this work was to find out whether Japanese is related to the other Transeurasian languages, the comparisons here are restricted to etymologies with a Japanese participant. Thus, comparisons between Koreanic and Tungusic, Tungusic and Mongolic or Mongolic and Turkic are not included, unless they display a Japanese cognate. This severely reduces the number of binary cognate sets. As a result, only for the comparison between Japonic and Koreanic, the number of valid cognates equals the actual number of cognates listed by Robbeets (2005), in the other cases there are more valid cognate pairs than the listed ones. It should thus not come as a surprise that Brown finds that the genealogical connection between Japanese and Korean is more strongly supported than the other relationships.

36.5 Conclusion

In contrast to the frequently uttered concern that the Transeurasian languages do not have enough basic vocabulary in common, it is possible to find 93 etymologies for 64 different concepts on the Leipzig-Jakarta basic vocabulary 100 list. The correlations

are phonologically and semantically regular: the correspondences in form answer to the requirements of a list of consonant and vowel correspondence, previously established on the basis of the entire lexicon, while the correspondences in meaning are checked against an empirically based standard of cross-linguistically observed polysemies and semantic associations.

On the basis of a number of observations, it is possible to argue that it is linguistically more sensible to attribute the correlations in the basic vocabulary to inheritance than to borrowing. In addition to the convincing power of the basic vocabulary *per se*, these observations relate to the relative high proportion of etymologies for verbs and grammatical items, the sharing of bare verb roots in contrast with the expected non-verbal strategy of verbal borrowing, the regularity and complexity of the sound correspondences, the absence of cognates in one or more intermediate contact branches, the five-fold setting of Transeurasian comparison, the absence of morphological and semantic characteristics prototypical in situations of borrowing and the well-spread distribution of the cognates.

In addition to borrowing, it is also possible to exclude coincidence as an explanation of the observed correlations in the basic vocabulary. Applying a cumulative binomial distribution, I calculated that we can be nearly 100% certain that the binary relationships between Japonic and Koreanic, Japonic and Tungusic, Japonic and Turkic and Mongolic and Turkic are not coincidental. The transitivity of family relationship then led to the conclusion that the correlations between the Transeurasian languages as a whole are probably not coincidental.

In sum, there are a considerable number of form-meaning matches within the basic vocabulary of the Transeurasian languages and it is possible to rule out non-

genealogical explanations such as borrowing and coincidence as an explanation of the similarities. Therefore, even if the evidence leaves much room for future elaboration and improvement, the Transeurasian languages appear to be genealogically relatable within the limits of the classical historical-comparative method.

Acknowledgement

---

[1] Abel-Rémusat (1820: 394): "Les mots de ces .. langages, particulièrement ceux qui désignent des objets de première nécessité,... sont radicalement différents..."

[2] Compare, for instance, (12) 'breast' and (69) 'suck', in which the Tungusic and Mongolic basic items for 'breast' are derived from the verbs 'to suck' by a deverbal noun suffix in *-n: PTg *xökö- 'to suck' → *xökö-n 'breast' and PMo *kökö- 'to suck' → *kökö-n 'breast'.

[3] See Unger (2010) for an overview of the weaknesses involved in Vovin's assumption. Especially the requirement that each Proto-Japonic reconstruction should

be supported by cognates in Eastern Old Japanese is not legitimate because there is extremely little textual evidence of this historical variety. As the aphorism goes, absence of evidence is not evidence of absence. Therefore, one cannot draw inferences from what is not found in the barely 400 Eastern Old Japanese poems that survive.