RANKO MATASOVIĆ
University of Zagreb
E-mail: rmatasov@ffzg.hr
Homepage: www.ffzg.hr/~rmatasov

AREAL AND TYPOLOGICAL AFFINITIES OF PROTO-INDO-EUROPEAN

Università di Macerata, April 8-9 2008

1. PHONOLOGY

Many linguists believe that PIE had a minimal vowel system with just two phonologically contrasting vowels.

SYSTEM I

*o [ə]

*e [a]

Minimal vowel systems are rare typologically, but common in NW Caucasian; 3-vowel system is attested in Classical Arabic (?and Proto-Semitic).

Languages of NE Caucasus and most of NE Eurasia have rich vowel systems.

If syllabic allophones of glides *y and *w are accepted as vowels, we arrive at the following vowel system:

SYSTEM II

i u

e o

In certain morphological categories, primary vowels can be lengthened; although predictable by morphological rules, the occurence of *ē and *ō is not predictable phonologically. Also, in a few rare instances the vowel *a can be reconstructed (e. g. PIE *laywo- "left" > Lat. *laevus*, OCS *lěvъ*). Although *ē, *ō, and *a probably did not exist in Early PIE (EPIE), many linguist posit them for Late PIE (LPIE):

SYSTEM III

i u

e o ē ō

a

ABLAUT

The PIE ablaut patterns can be illustrated by the following examples:

PIE *ph_2-tr-os (> gr. *patrós*)
PIE *ph_2-ter-m (> gr. *patéra*)
PIE *$h_1$su-ph_2-tor-m (> gr. *eu-patór-a*)
[illegible])*r*
PIE *$h_1$su-ph_2-tōr (> gr. *eu-pátōr*).

Ablaut (morphologically conditioned vowel alternations) exists in NW Caucasian, Kartvelian, and Afro-Asiatic, but is generally rare in Eurasia. Cp. Kabardian *than* "read" (intr.) vs. *then* "read (something)" (transitive), *-s-* "transitive 1 sg. prefix" vs. *-se-* "intransitive 1 sg. prefix"; Georgian 1. sg. aorist *da-v-dev* vs. 3 sg. *da-dv-a;* Common Kartvelian **šer-t-* "be extinguished" vs. **šr-et-* "extinguish".

CONSONANTS

a) "laryngeals": *h_1, *h_2, *h_3
b) fricatives: *s
c) resonants
*m *n *l *r *y *w

d) stops:

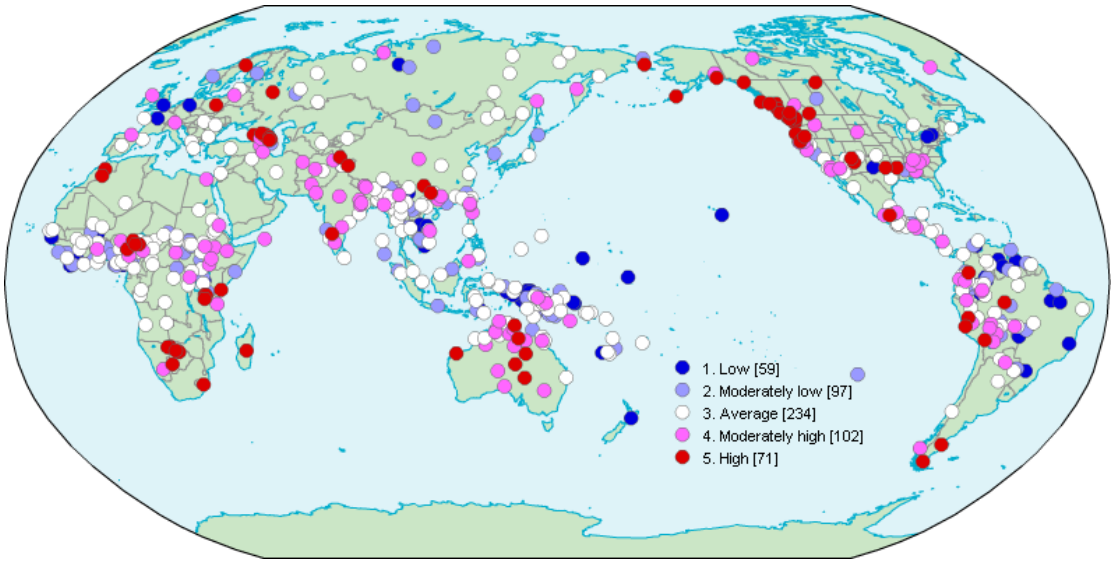
labials	*p	(*b)	*b^h
dentals	*t	*d	*d^h
velars	*k	*g	*g^h
palatalized velars	*k’	*g’	*g'^h
labiovelars	*k^w	*g^w	*g^{wh}

Three series of stops are significantly less common in Eurasia than two series. Such systems are found in the Caucasus, in Semitic, and in the Indian Subcontinent (where languages with four series of stops are also common). In NE Eurasia such systems are found in Sino-Tibetan and Korean, as well as in some isolates, but not in Uralic and Altaic.

Consonant to vowel ratio is high[1] for PIE (12,5) if one accepts that the proto-language had only two vowels in the earliest reconstructable stage. Such systems

[1] WALS classifies languages as having "high" consonant/vowel ratio if the number of consonants divided by the number of vowels is above 6,5. Only 10 out of 563 languages in WALS have the ratio of 12 or higher. If 4 vowels are reconstructed for PIE, then it should be classified as a language with "moderately high" c/v ratio (4,5-6,5).

are rare cross-linguistically, but common in the Caucasus (especially in NW Caucasian).



EXCURSUS 1: The Glottalic Theory

"Glottalic theory" (proposed by T. Gamkrelidze, V. Ivanov[2], and P. Hopper) presupposes that voiced stops should be reinterpreted as ejectives (*p', *t', *k', etc.). Arguments supporting the glottalic theory are disputed, and all would be valid for EPIE as well:

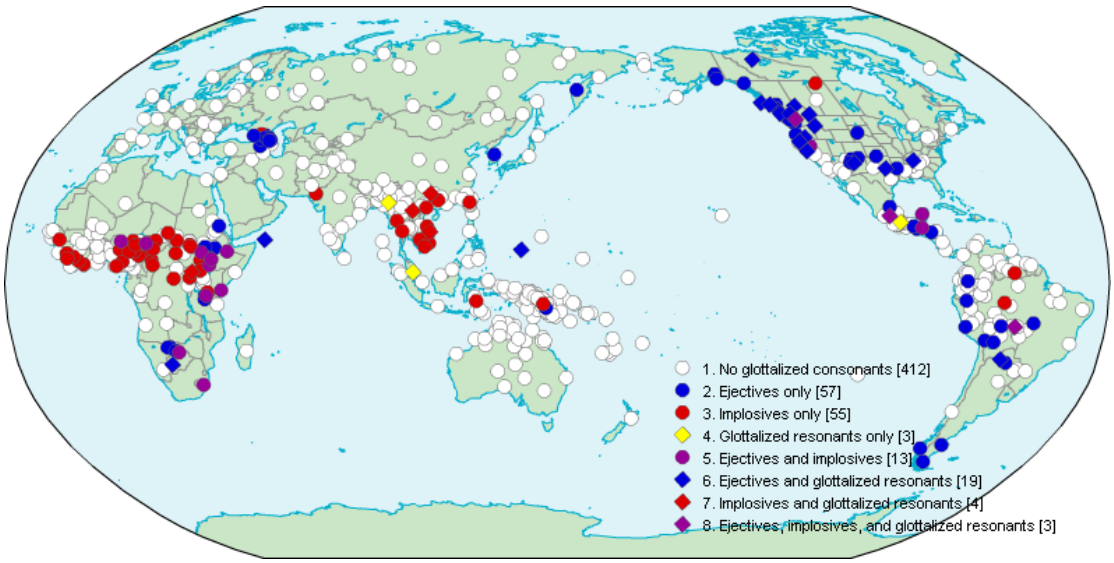
1. PIE had no *b, but it had *p; such languages are unattested[3], but there are languages in which glottalized *p'* is absent, but voiceless bilabial stop *p* (and/or its voiced and aspirated counterparts) exists (e. g. Avar, a NE Caucasian language). In languages with glottalized consonants, that series of stops is statistically the least common (e. g. in Kartvelian). Traditional voiced stops are the least common series in the reconstructed PIE.
2. There are no roots with two voiced occlusives in PIE (i.e. roots like *bed- *geb- *deg-, etc.), but other combinations of stops are possible (e. g. *pet- "fly", *ped- "foot", *b^{h}ewd^{h}- "be awake", etc.). In languages with glottalic consonants, prohibitions against roots with two glottalic stops are common, e. g. in Kartvelian, Shuswap (Salish), Yucatec (Mayan), Quechua, etc.
3. Traditional voiced stops are rare or non-existent in inflectional morphemes in PIE. Such restrictions on the occurrence of glottalics are common in languages with glottalic consonants, e. g. in Kartvelian (Gamkrelidze & Ivanov 1984).
4. The glottalic consonants are allegedly preserved in some IE languages (dialects of Eastern Armenian, Sindhi), or have left clear traces; according to F. Kortlandt, the development of the Balto-Slavic acute intonation on syllables lengthened before traditional "voiced stops" (the so-called "Winter's law") is most easily understood if one assumes that they were actually glottalized, cp. PIE *udreh$_2$

[2] See Gamkrelidze & Ivanov 1984.
[3] This universal also appears not to hold without exception; Dargin, an NE Caucasian language, has *p$^{(h)}$* and *p'*, but not *b*, at least according to the description by Abdullaev (in Vinogradov, ed. 1967: 598).

the theory of tonogenesis that high tone often develops before glottal stops (e. g. in Vietnamese and other languages of the Far East).

The "glottalic theory" had many adherents in the 1980s and 1990s, but now it seems to lose popularity (though some scholars, like F. Kortlandt, still accept it). I remain unconvinced, but find it possible that Early PIE had glottalic consonants.

Caucasus is the only area in the whole of Eurasia where glottalic consonants are common.



ROOT STRUCTURE

CVC (*b^{h}er- "carry"), CVCC (*melh$_2$- "grind"), CCVC (*h$_2$melg'- "milk"). There seems to have been a restriction against vowel-initial roots (nearly all such roots can be interpreted as having the initial laryngeal). Words could not begin with an *r- (a restriction that still holds in Proto-Greek and Armenian, cp. Matasović 1992).

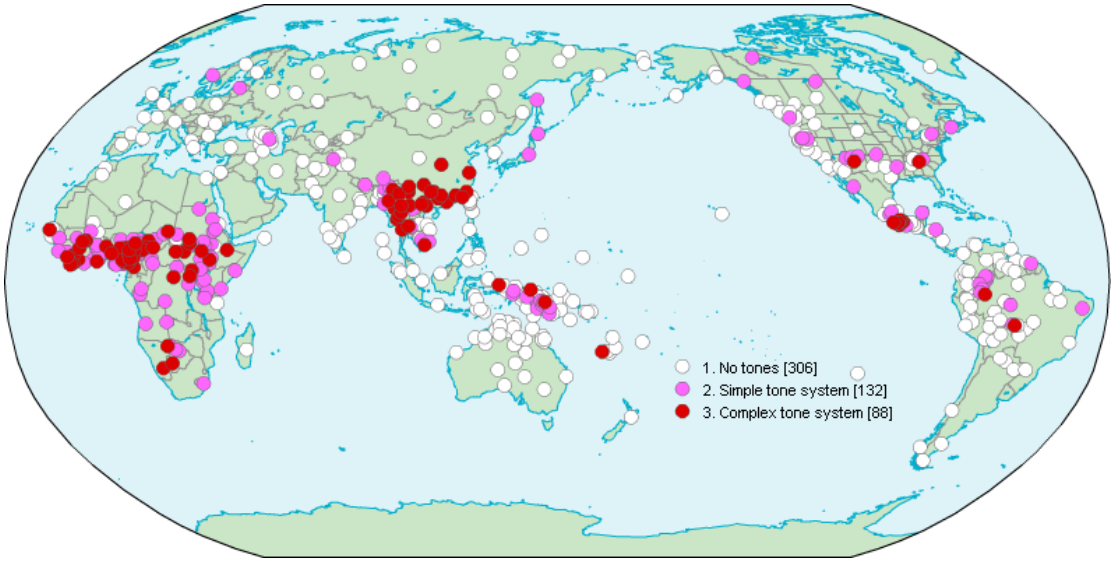
Root structures proposed for PIE are typologically very common. Restrictions against vowel-initial roots exist in NW Caucasian (e. g. Kabardian) and Afro-Asiatic. Restrictions against word-initial *r-* exists in all languages of the Caucasus, but in other languages as well (e. g. in Basque).

ACCENT

PIE had a free accent; in most ancient languages (Vedic, Greek), as well as in some younger ones (Balto-Slavic) it is tonal, rather than dynamic in nature.

It remains disputed whether PIE had complex (rising/falling) tones (as in Greek or Lithuanian). It appears most likely that only one accent in the word could be marked with a higher pitch.

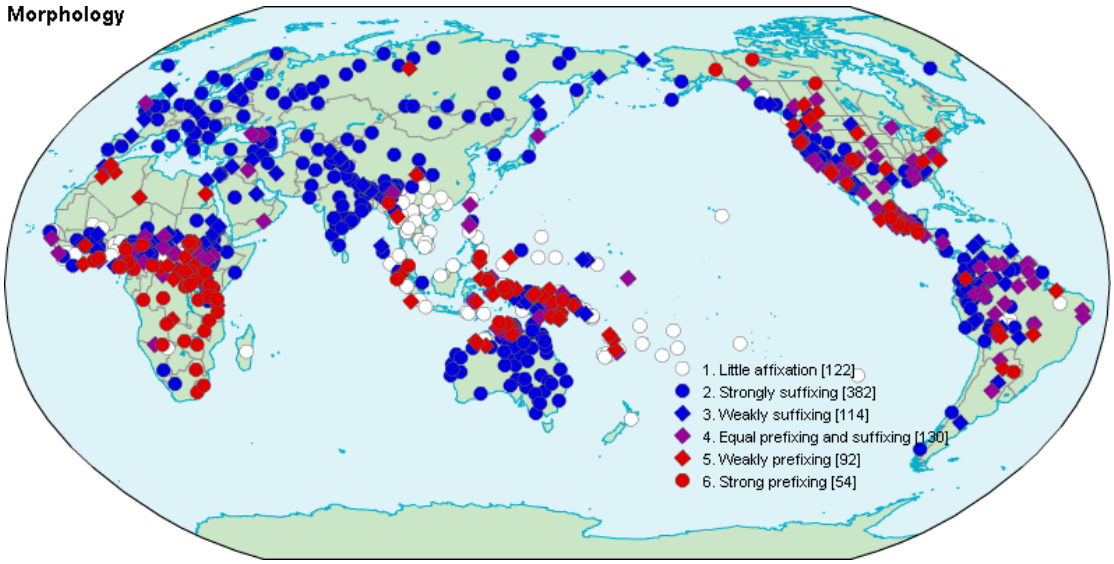
Tonal languages are rare in most of Eurasia, with the exception of E Asia; an isolated tonal language is, e. g., Ket (Siberia).



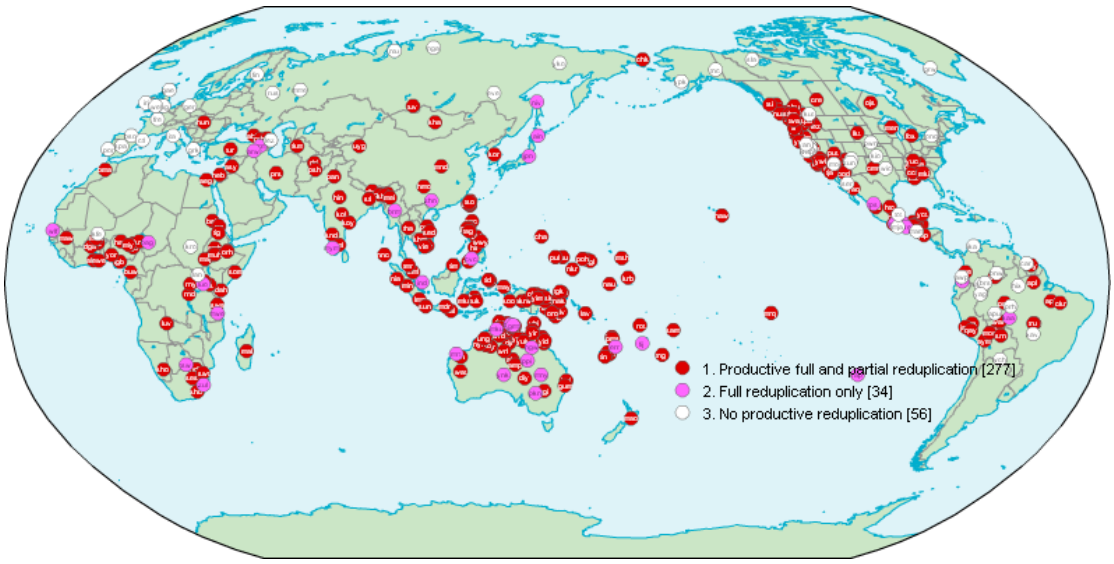
2. MORPHOLOGY

2.1. PIE was a consequently suffixing language. This type is rather common in Eurasia, where suffixing is the default morphological marking, with some isolated exceptions (e. g. Ket, Abkhaz).

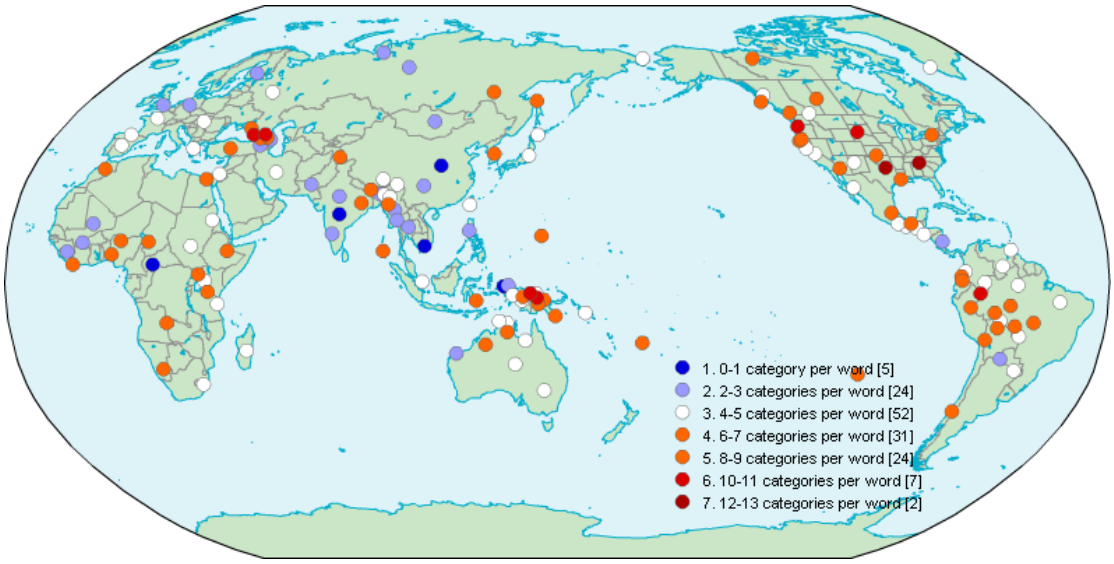
Typological parallels to PIE are many suffixing languages and families in N and NE Eurasia: Uralic, Tunguso-Manchurian, Turkic, Japanese, Mongolian. Languages of the Caucasus are classified as either "strongly suffixing" (NE Caucasian), weakly suffixing (Georgian), or "equally prefixing and suffixing" (NW Caucasian). There are no exclusively suffixing languages (such as PIE) on the Caucasus.



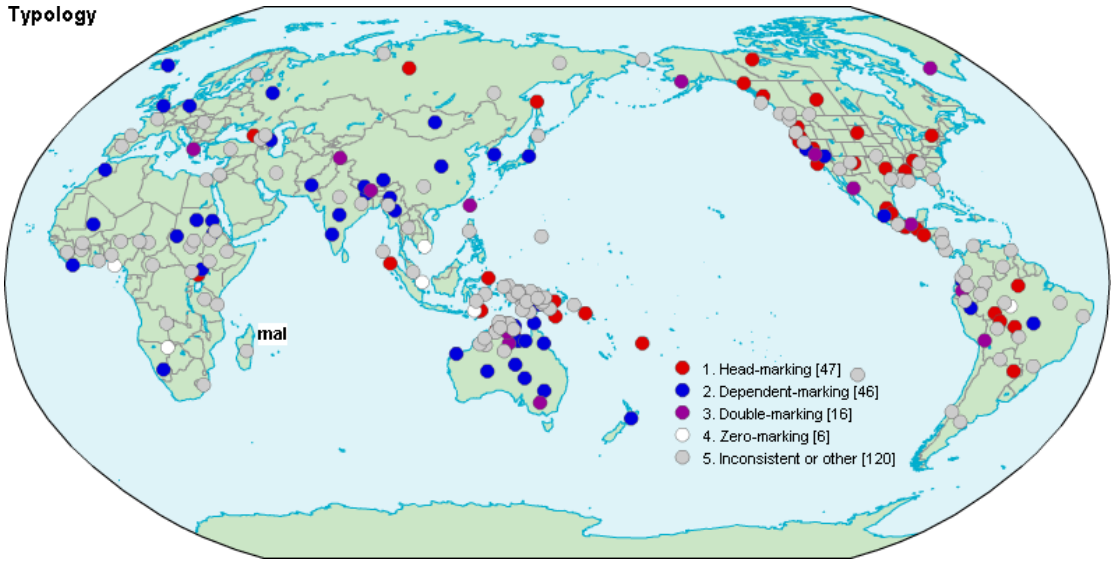
PIE has reduplication in verbal morphology (e. g. in the perfect *le-loyk^{w}-h$_2$e "I left" > Gr. *léloipa*), less clearly in some nominal derivatives (e. g. *b^{h}e-b^{h}ru- "beaver" > Lith. *bebrus*). Languages with reduplication are rather rare in N Eurasia, but common in S Eurasia.



PIE was clearly fusional. The number of categories per word as measured by the maximal number of categories expressed on a verbal form (in the WALS sense) is 5 (person, number, version, tense and mood), which makes PIE belong to the most common type cross-linguistically (52 out of 145 languages in WALS). However, Eurasia is dominated by low-synthesis languages, except for the Caucasus and the Himalayan region.



In terms of Johanna Nichols' "Locus of Marking" typology (Nichols 1992), PIE was clearly dependent marking: it expresses possession by a genitive case, marks nouns rather than adpositions in adpositional phrases, does not mark object on verbs, etc. Such languages are the default in Eurasia, except for some NW Caucasian languages (e. g. Abkhaz) and some isolates (e. g. Ket, Nivkh).

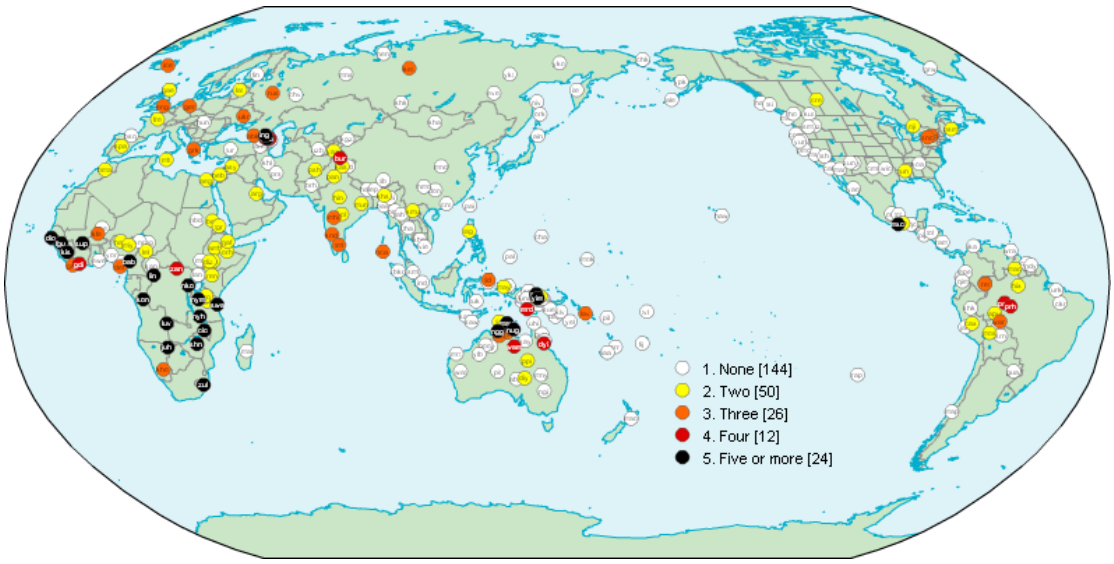


NOMINAL INFLECTION

Case: PIE had Nominative, Vocative, Accusative, Genitive, Ablative, Dative, Locative, and Instrumental. A simpler system presumably existed in Early PIE.

Number: Singual, Plural, Dual; the existence of the Dual in EPIE has been disputed, since it is lacking in Anatolian, but the fact that personal pronouns are inflected in the dual probably testifies to its antiquity. Anatolian languages presumably lost it.

Gender: Masculine, Feminine and Neuter. EPIE probably had just the opposition between the Common Gender (> Masculine/Feminine) and Neuter. In several early IE languages adjective distinguish only two forms, one for the m. and f. genders, and the other for the neuter gender, e. g. Latin *suavis* (m. and f.), *suave* (n.) "sweet". Also, feminines and neuters share a number of case forms which suggests that the feminines were originally only a sub-class of neuters. In terms of the number of Gender, PIE is within the average for Eurasia, where languages with more than three genders are found only in the NE Caucasian family (e. g. Bats, Ingush), and in the isolated Burushaski, with four genders.



We shall look at the gender system of PIE from the areal and typological point of view in the last part of these lectures.

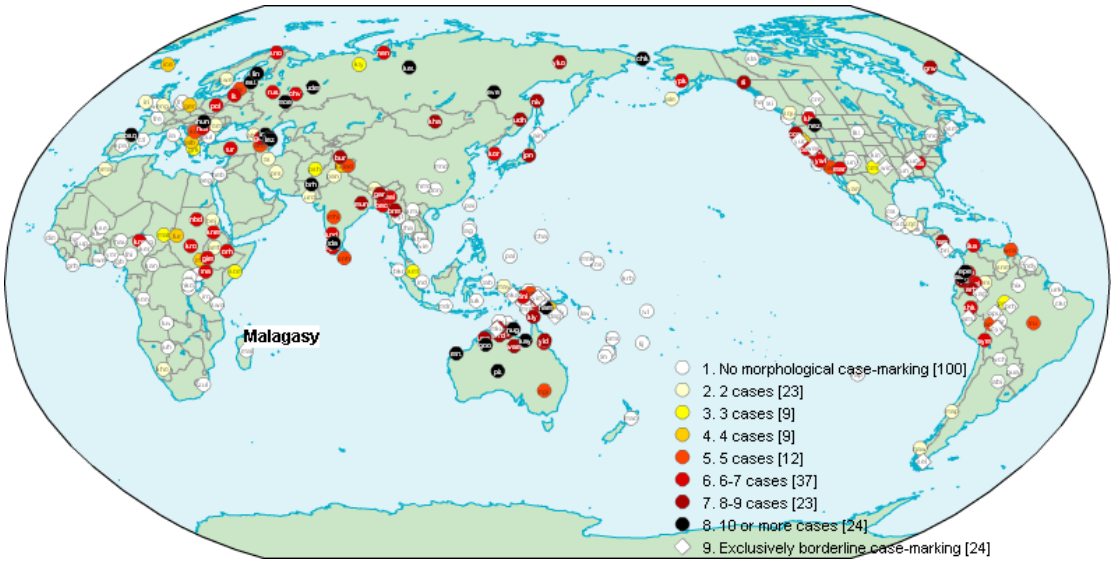
Examples of nominal inflection:

PIE *pōds "foot"

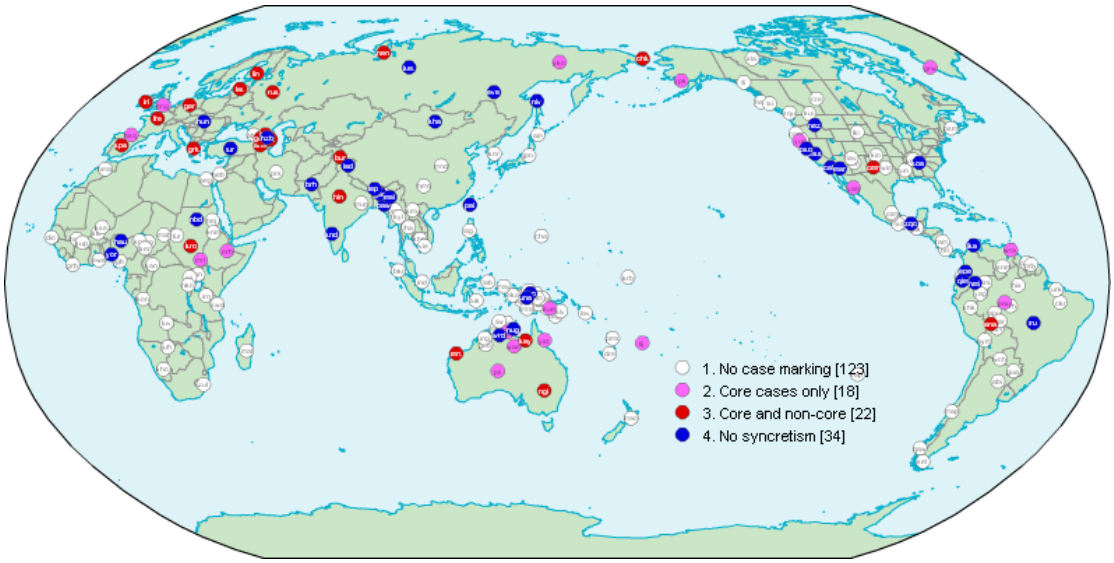
	PIE	Skr.	Gr.	Lat.
sg.				
N,V	*pōds	pāt	poús	pēs
A	*podm	pādam	póda	pedem
G,Abl.	*pod-os	padas	podós	pedis
D	*pod-ey	pade	podí	pedī
L	*pod-i	padi	(Abl.) pede	
I	*podeh_1	padā		
du.				
NVA	*pod-oh_1	padā(u)	póde	
GL	*pod-ows	padoš		
DIAbl.	?*pod-b^{h}yōm	padbhyām	(GD) podoiin	
pl.				
NV	*pod-es	pādas	pódes	pedēs
A	*pod-ns	padas	pódas	pedēs
G	*pod-om	padām	podõn	pedum
DAbl.	*podb^{h}os	padbhyas	possí	pedibus
L	*podsu	patsu		
I	*podb^{h}is	padbhiš	(Myc.) poppi	

Morphological case marking is quite common in Northern Eurasia; it is absent in part of NW Caucasian (Abkhaz, Abaza) and quite recent Adyghe-Kabardian. It is absent in the isolating languages of the Far East (e. g. Mandarin) and Indochina (e. g. Vietnamese). In terms of number of cases PIE belonged to a type that is well

represented in Eurasia (Uralic and NE Caucasian languages often have 8 or more cases).

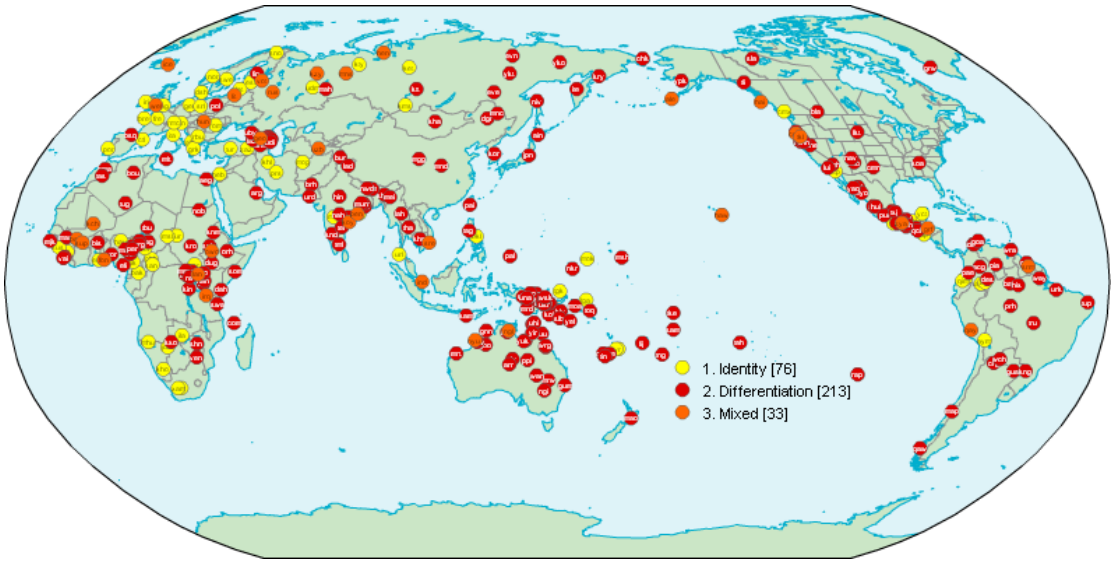


Case syncretism can be of two types: 1) involving core cases only (e. g. the PIE syncretism of N and A in the neuters) and 2) involving core and non-core cases (e. g. the Kabardian syncretism of Ergative and Oblique cases for definites). The second type, which is quite common in modern IE languages, seems to have been absent in PIE, but it is well represented in the languages of N Eurasia.



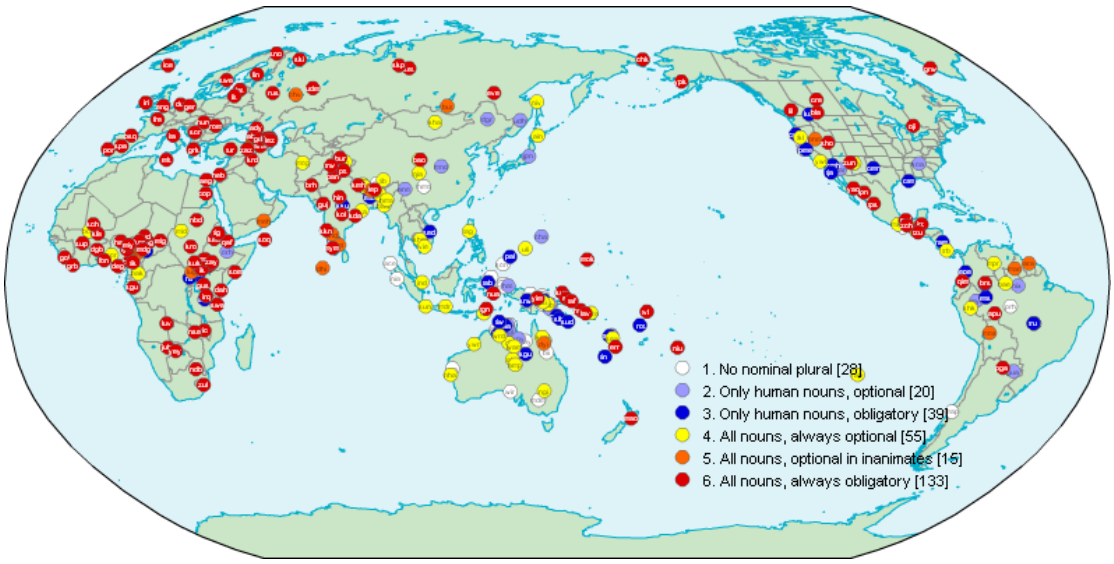
With respect to case function, it should be noted that PIE was most likely a language in which instrumentals were differentiated from comitatives, unlike most modern European languages (not only Indo-European!). Early IE languages clearly differentiate between the instrumental construction (e. g. Latin *pisces hamo capio* "I catch fish with an angle") from the comitative construction (Latin *eo tecum* "I go with you"). Today, languages conflating instrumental and

comitative function are virtually limited to W Eurasia and parts of sub-Saharan Africa.



The morphological expression of number is ubiquitous in Eurasia; it is lacking in the isolating languages of the Far East and Indochina. The dual number existed in Proto-Semitic (and it is preserved in Classical Arabic), otherwise it is rare in Eurasia (it occurs in some Uralic languages).

In PIE, the coding of nominal plurality was always obligatory, i. e. every noun had to be marked for number. A possible exception to this may have been the neuter nouns, which did not have a proper plural, but rather a collective noun which triggered singular agreement on the verb (in Greek, this is the so-called *tà zôa trékhei* rule, still valid in Homer). Languages in which plural marking on nouns is obligatory predominate in the whole of Eurasia, except in the eastern periphery. Languages where plural marking is optional on inanimates are found sporadically in the Caucasus (Laz, Kabardian) in India (Kannada, Tamil), and elsewhere (Buriat, Chuvash):



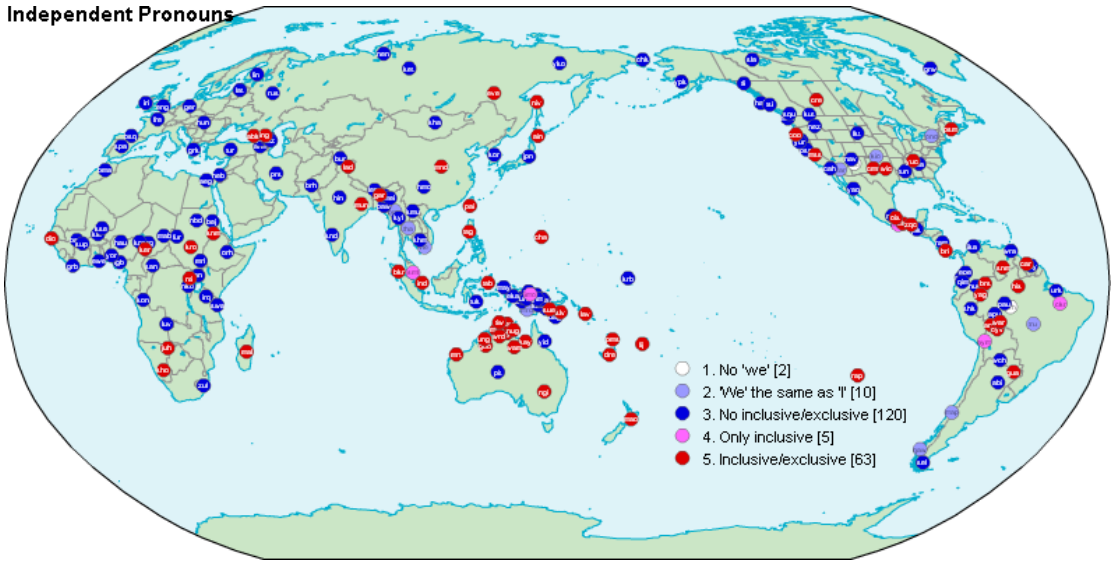
Personal pronouns:

	PIE	OCS.	Lith.	Lat.	Ved.
N	*eg'$h_2$om, *eg'oh_2, *eg'	azъ	àš	ego	ahám
A	*meh$_1$(m) *me	mene, mę	manè	mē(d)	mām, mā
G	*mene	mene	manęs	meī, mīs	máma, mē
Abl.	*med (?)			mēd	mad
D	*megthoy *moy	mъně, mъně, mi	mán	mihī	máhya(m), me
L	?*moy	mьně	manyjè		máyi
I		mъnoją	manimì		máyā

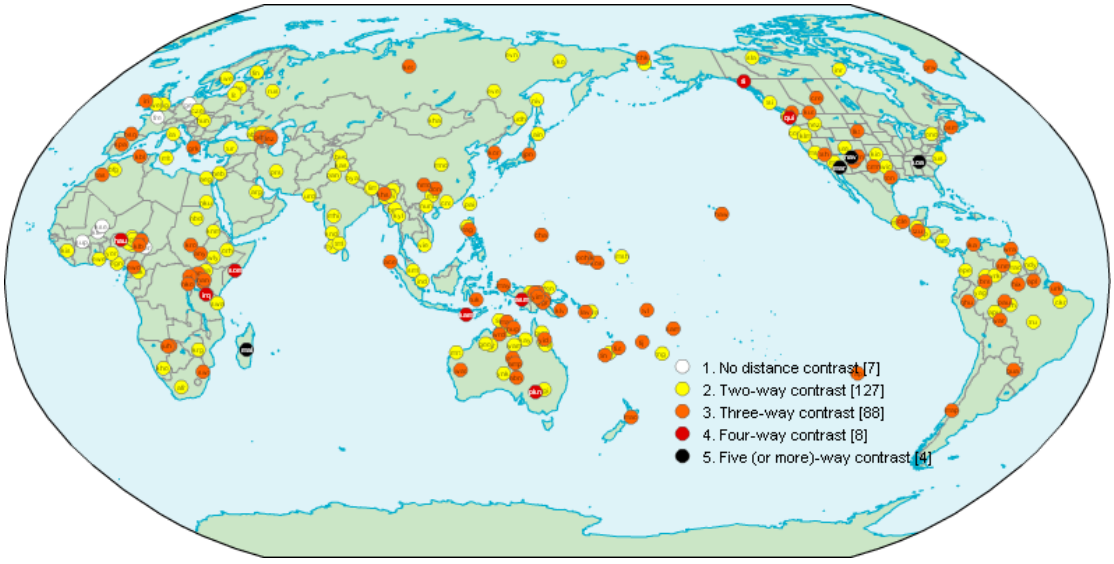
	PIE	OCS	Lith.	Lat.	Ved.
N	*wey-/*no-	my	mẽs	nōs	váyam
A	*ns-	nasъ, ny	mùs	nōs	′ām
G		nasъ	′ūn ų	nostrī, -rum	′ām
Abl.					asmád
D		namъ, ny	mùms	nōbīs	′ām
L		nasъ	mumysè		′ām
I		nami	mumìs		′ām

Personal pronouns show suppletion for number, which is quite common in Eurasia; the existence of several stems in the 1st and 2nd person plural (and dual) is unaccounted for. The stem *we- occurs as both 1st person plural (e. g. in Vedic) and as the 2nd person plural (e. g. in Latin), while the stem *no- occurs only in the 1st person plural and dual (e. g. in Latin and Slavic). This distribution is consistent with the assumption that the stem *we- was originally the *inclusive* pronoun (meaning "we" as "the speaker, the addressee and others", while *no- was the *exclusive* pronoun (meaning "the speaker and others, but not the addressee").

It is possible that the inclusive pronoun subsequently changed its meaning from "we and you" into just "you" in some languages. The inclusive/exclusive opposition is rare in Eurasia, but common in the Caucasus (e. g. in Ingush, Adyghe, etc.). It also occurs in some isolates (Ainu, Nivkh) and in some languages of the Himalayas (Garo, Ladakhi):



Several demonstrative stems can be reconstructed (e. g. *so-, *to-, *(h_1)ey-, *Hen-), but it is unclear how many distance contrasts existed. Languages with three-way constrast predominate among the earlier IE languages (e. g. Latin, OCS), but two-way contrasts also occur. Languages with no distance contrasts (e. g. German) are a recent development. In Eurasia, both two- and three-way contrasts are common.

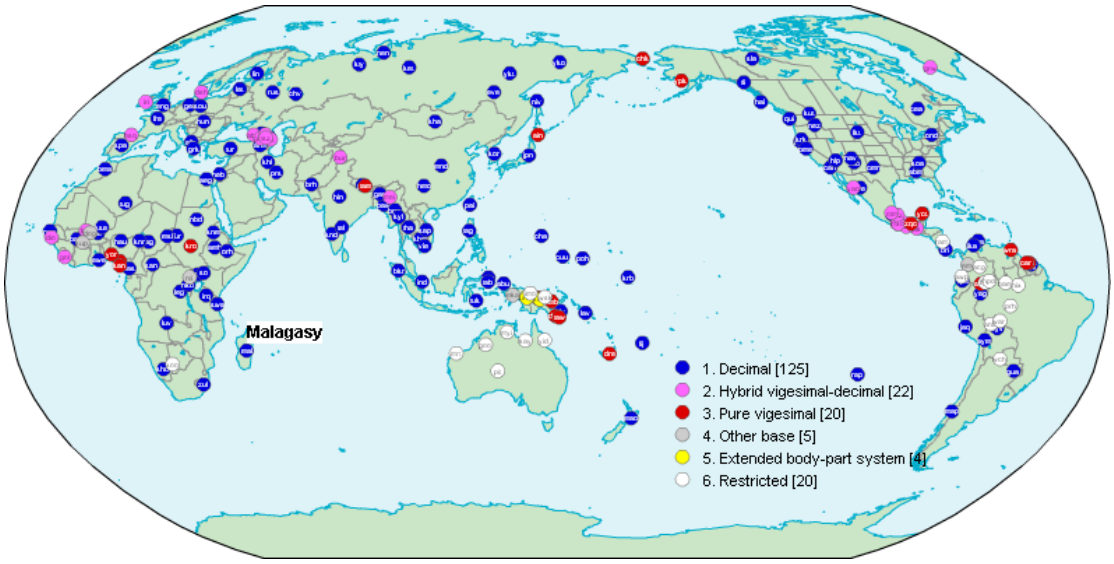


Numerals:

PIE	OCS	Lith.	Lat.	Skr.

*(H)oynos	*jedinъ*	*vienas*	*ūnus*	*ékas*
*d(u)woh$_1$	*dъva*	*dù*	*duo*	*dvá*
*treyes	*trьje*	*trỹs*	*trēs*	*tráyas*
*k^{w}etwores	*četyre*	*keturì*	*quattuor*	*catvā́ras*
*penk^{w}e	*pętь*	*penkì*	*quīnque*	*páñca*
*s(w)ek's	*šestь*	*šešì*	*sex*	*ṣáṭ*
*septm	*sedmь*	*septynì*	*septem*	*sápta*
*h$_3$ek'tō(w)	*osmь*	*aštuonì*	*octō*	*aṣṭá*
*newn	*devętь*	*devynì*	*novem*	*náva*
*dek'm(t)	*desętь*	*dẽšimt*	*decem*	*dáśa*

The numeral system is clearly of the decimal type, cp. OCS *jedinъ na desęte*, Lat. *undecim*. Traces of the vigesimal system occur in Insular Celtic, and are a recent development in some languages, perhaps under areal influence (e. g. French *quatre-vingts*). Languages with vigesimal system are otherwise rare in Eurasia, but do occur in some isolates (Ainu, Chukchi). Caucasian languages (e. g. Kabardian, Tsez) often have hybrid decimal-vigesimal systems.



VERBAL INFLECTION

The present tense (athematic):

PIE	OCS	OLith.	Skr.	Lat.	Hitt.
*h$_1$esmi	jesmь	esmi	ásmi	sum	ešmi
*h$_1$essi	jesi	esi	ási	es	ešši
*h$_1$esti	jestъ	esti	ásti	est	ešzi
*h$_1$smos	jesmъ	esame	smas	sumus	ešweni
*h$_1$st(H)e	jeste	esate	stha	estis	ešteni
*h$_1$senti	sątъ	(esti)	sánti	sunt	ašanzi

The dual (present indicative):

Skr.	Lith.	OCS
1. *svas*	*esva*	*jesvě*
2. *sthas*	*esta*	*jeste*
3. *stas*	-	*jesta, jeste*

PIE
1. *h_1s-wo-
2. *h_1s-t(H)o(s)
3. *h_1s-tos

The present tense (thematic):

PIE	Skr.	Lat.	Lith.	OCS.
sg.				
*b^heroh$_2$	bharāmi	ferō, legō	vedu	vedą, berą
*b^herey	bharasi	fers, legis	vedi	vedeši, bereši
*b^here	bharati	fert, legit	veda	vedetъ, beretъ
pl.				
*b^herome	bharāmas	ferimus, legimus	vedame	vedemъ, beremъ
*b^heret(H)e	bharatha	fertis, legitis	vedate	vedete, berete
*b^heronti	bharanti	ferunt, legunt	veda	vedątъ, berątъ
du.				
*b^herowo (?)	bharāvas		vedava	vedevě, berevě
*b^heretHo (?)	bharathas		vedata	vedeta, bereta
*b^hereto(s) (?)	bharatas			vedete, berete

The aorist:

PIE	skr.	gr.
sg		
1. *d^heh$_1$-m	a-dhā-m	é-thēk-a
2. *d^heh$_1$-s	a-dhā-s	é-thēk-as
3. *d^heh$_1$-t	a-dhā-t	é-thēk-e(n)
pl.		
1. *d^hh$_1$-me	a-dhā-ma	é-the-men
2. *d^hh$_1$-te	a-dhā-ta	é-the-te
3. *d^hh$_1$-(e)nt	a-dh-ur	é-the-san

The perfect:

PIE	Skr.	Gr.	Lat.
1. *woyd-h$_2$e	véda	oîda	vīdī
2. *woyd-th$_2$e	véttha	oîstha	vīdistī
3. *woyd-e	véda	oîde	vīdit

mn.

1. *wid-me	vidmá	ídmen	vīdimus
2. *wid-e (?)	vidá	íste	vīdistis
3. *wid-r	vidúr	ísāsi(n)	vīdērunt, vīdēre

The original difference in meaning between the aorist and perfect is unclear, but it may have been one of aspect rather than tense. It appears possible, also, that the perfect was originally just the present of stative verbs (personal endings of a large class of presents in Anatolian are easily reconcilable with the perfect endings in other IE languages).

There were two diatheses, or "versions", active and middle. The original function of the middle is unclear, but it is usually claimed that the middle expresses the action performed to the advantage of the subject.

The middle:

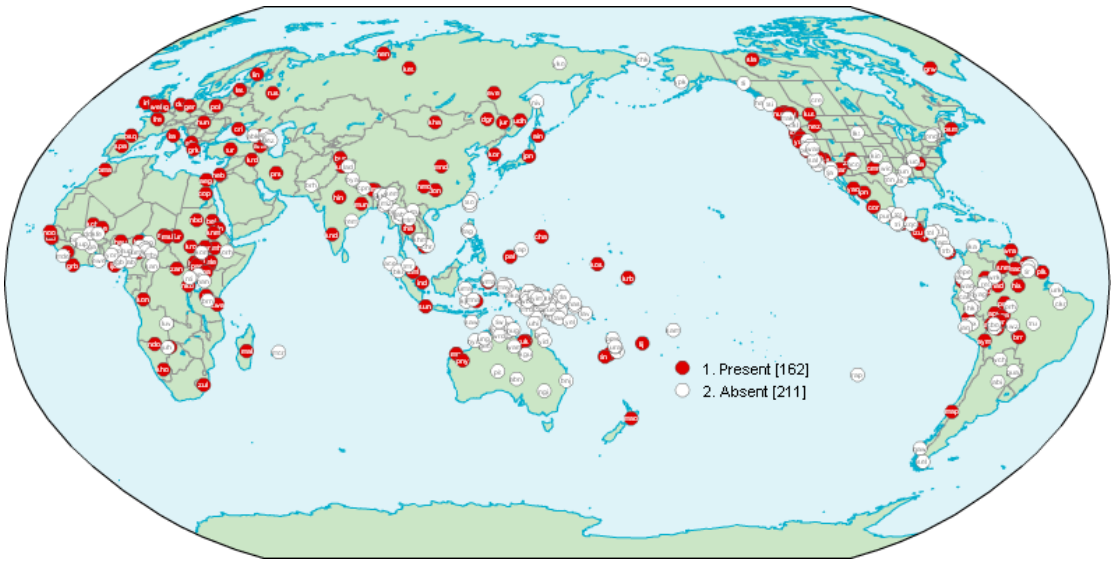
PIE	Skr.	Gr.	Lat.	OIr.
1. *sek^{w}-o-h$_2$e (?)	sace	hépomai	sequor	sechur
2. *sek^{w}-e-sor	sacase	hépēi	sequeris sequere	sechither
3. *sek^{w}-e-tor	sacate	hépetai	sequitur	sechithir

pl.

1. *sek^{w}-o-med^{h}h$_2$	sacāmahe	hepómetha	sequimur	sechimmir
2. *sek^{w}-e-d^{h}we	sacadhve	hépesthe	sequiminī	sechithe
3. *sek^{w}-o-ntor	sacante	hépontai	sequuntur	sechitir

Formations parallel to the PIE middle are common in the languages of the Caucasus, but appear to be rare elsewhere in Eurasia.

Notably absent from PIE is any morphological marking of the passive. There are also no indications that the passive was formed analytically in the protolanguage, so the most probable conclusion is that there was no passive in PIE. The geographical distribution of languages with and without passive clearly links PIE with the Caucasus region; the only other region of Eurasia where languages without passive are common is Indochina.



Passives are mostly new formations in early IE languages, as well as other valence-reducing operations such as reflexivization (the Romance type of French *se laver*, Croatian *prati se* "to wash oneself"). On the other hand, causative was a productive valence-increasing category in PIE; it was formed by ablaut of the root (which was in the o-grade) and by adding the suffix *-eye-, cp. OCS *u-mьrěti* "die" vs. causative *moriti* "kill" < *moreyeti (Skr. *mārayati*), etc. Languages with productive causative formation abound in Eurasia.

There were the following moods in PIE: indicative, imperative, subjunctive and optative. Traces of subjunctive and optative are absent in the Anatolian branch.

The imperative: *b^{h}ere "carry!" (Lat. *fer*, Gr. *phére*, Skr. *bhára*).

The subjunctive (2. sg.) *b^{h}erēti (Gr. *phérēis*, Skr. *bharāti*).

The optative:

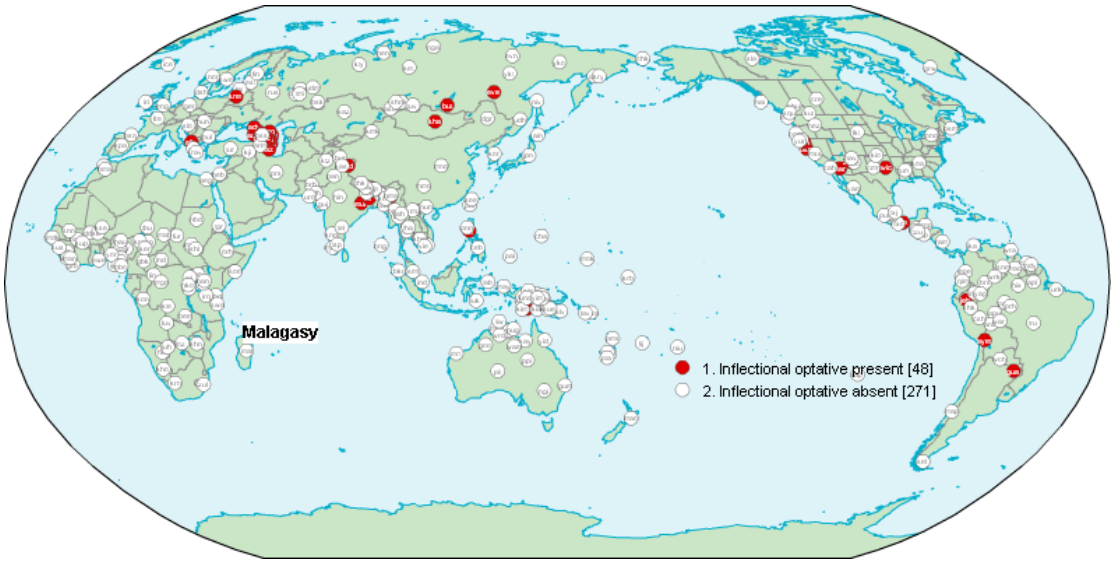
PIE	Skr.	OLat.
1. *h$_1$s-yeh$_1$-m	syām	siem
2. *h$_1$s-yeh$_1$-s	syās	siēs
3 *h$_1$s-yeh$_1$-t	syāt	siet

pl.

1. *h$_1$s-ih$_1$-me	syāma	sīmus
2. *h$_1$s-ih$_1$-te	syāta	sītis
3. *h$_1$s-ih$_1$-nt	syur	sīnt, sient

The function of the optative was the expression of wish or desire, at least in the main clauses (it dependent clauses it marks the subordinated status of the verb). Morphological optative is extremely rare in the languages of Eurasia, but it is quite common in the Caucasus, occurring in both the indigenous languages (NE

and NW Caucasian, Kartvelian), and in the "intruders" (Azerbaijani, Kumyk, Noghai).

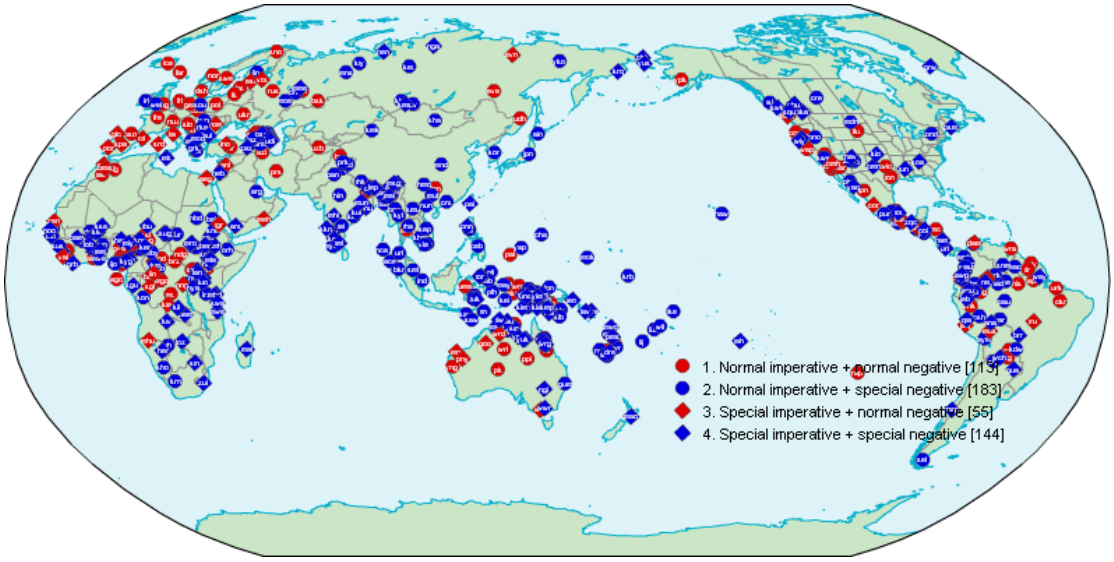


?The future/desiderative

Lith.	Gr.
sg.	
1. liksiu	leípsō
2. liksi	leípseis
3. liks	leípsei
pl.	
1. liksime	leípsomen
2. liksite	leípsete
3. liks	leípsousin

The existence of the future tense in PIE is uncertain. It appears that the parallel sigmatic formations in Greek, Lithuanian and Indo-Iranian developed from independently from some modal form in the Proto-Language.

Typologically interesting is the existence of a special prohibitive negation *meh$_1$ [illegible] language in which prohibition was marked on the negation, rather than on a special verbal form (e. g. Spanish prohibitive subjunctive *no cantes* vs. the imperative *canta*).

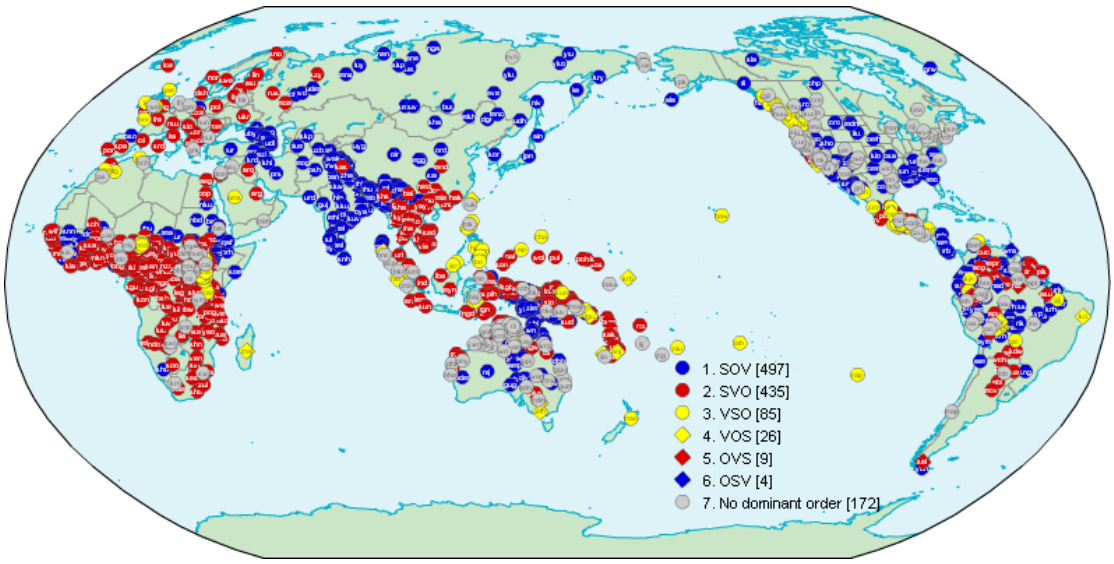

1. Normal imperative + normal negative [113]
2. Normal imperative + special negative [183]
3. Special imperative + normal negative [55]
4. Special imperative + special negative [144]

3. SYNTAX

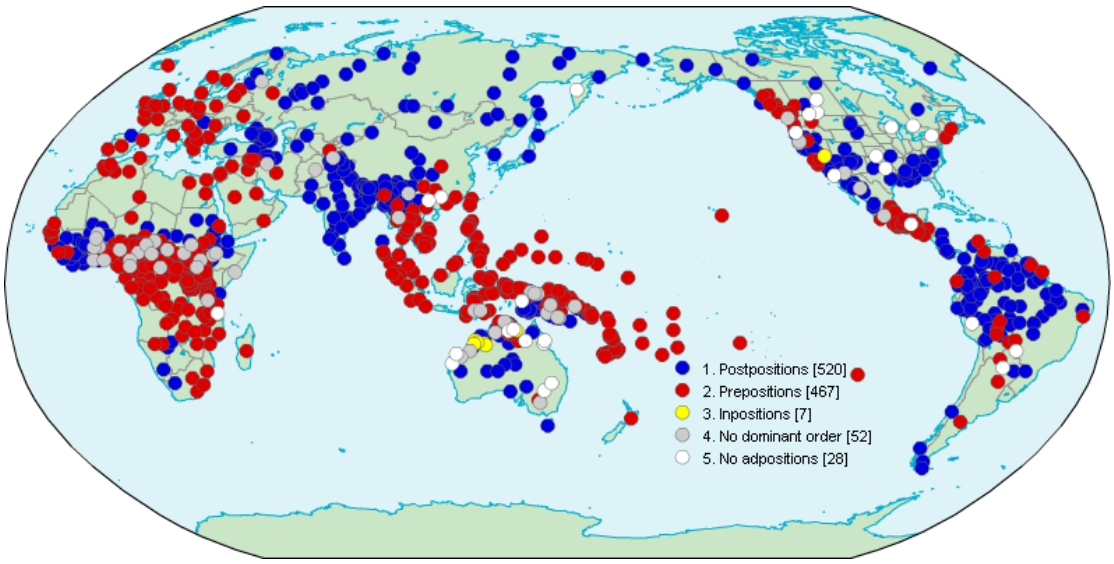
WORD ORDER

PIE is usually reconstructed as SOV; of the early languages, Hittite, Vedic, and Early Latin are SOV (with a rather rigin SOV order in Hittite). Insular Celtic languages are VSO, and Slavic and Romance languages are predominantly SVO:

Hitt. NINDA-an *ēzzašši wātarra ekušši* "you will eat bread (NINDA) and drink water (*wātar*)"
Ved. *mántram...atakšan* "they made an oath (*mantram*)"



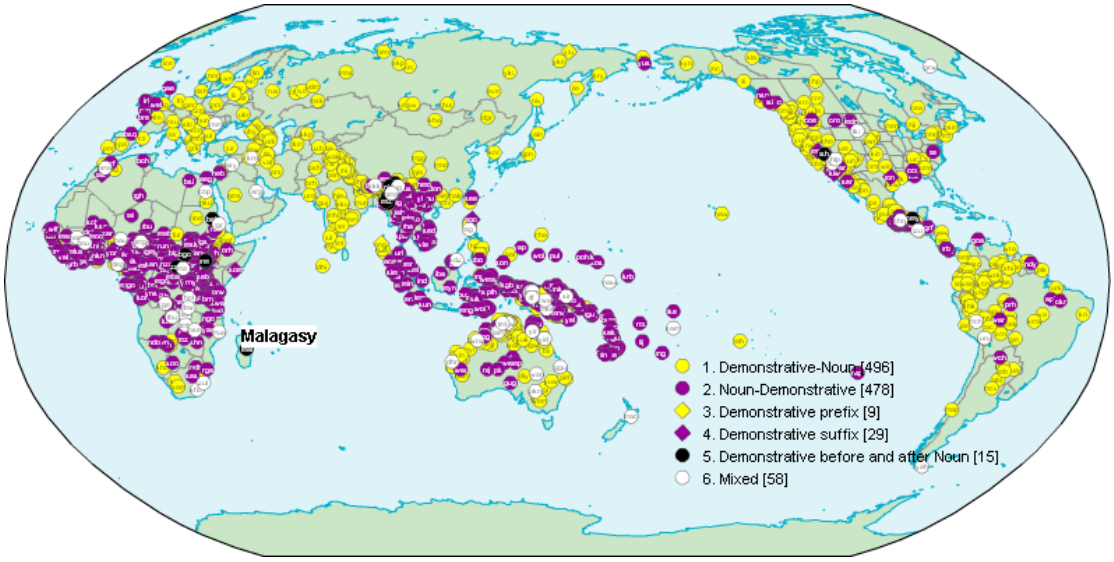
Anatolian languages and Vedic have postpositions, otherwise prepositions predominate in early IE languages. There are traces of postpositions in Latin (e. g. *te-cum* "with you") and Slavic (e. g. Croatian *radi* "because of" can be a postposition as well as preposition). However, as noted already by Greenberg (1978), postpositions are expected in a SOV language (this is one of the firmest implicational universals in word order typology). Languages with postpositions predominate in most of Eurasia, except in Europe and the Far East.



Both N-Gen (e. g. Slavic) and Gen-N (e. g. Hittite, Lithuanian) orders are found, as well as both Adj.-N (e. g. Hittite, Slavic) and N-Adj. (e. g. Insular Celtic) orders. Many languages (including Vedic, Latin, and Greek) have a rather free order of these constituents, but it would appear that N-Adj. and Gen-N are more common.

Hittite *šuppi wātar* (clear-water) "clear water"
Gr. *hieròn ménos* (holy.NAcc.sg.neut. power-NAcc.sg.) "holy power"
Ved. *vásu śravas* (good-NAcc.sg.neut. fame-NAcc.sg.) "good fame"
Hitt. *parn-aš išha-š* (house-Gsg. lord-Nsg.) "lord of the house"
Gr. *hēlḗ-ou kýklo-s* (sun-Gsg. wheel-Nsg.) "wheel of the sun"
Ved. *sūrya-sya cakrá-m* (sun-Gsg. wheel-Nsg.) "wheel of the sun"

As in the majority of languages of Eurasia (except, e. g., in Basque), the demonstrative pronouns precede the NPs, but in some languages N-Dem is optional (e. g. Lat. *homo ille* besides *ille homo*). Among the languages of Europe today only Basque has the dominant order N-Dem (*etxe hau* "this house"), otherwise the opposite order is the default:

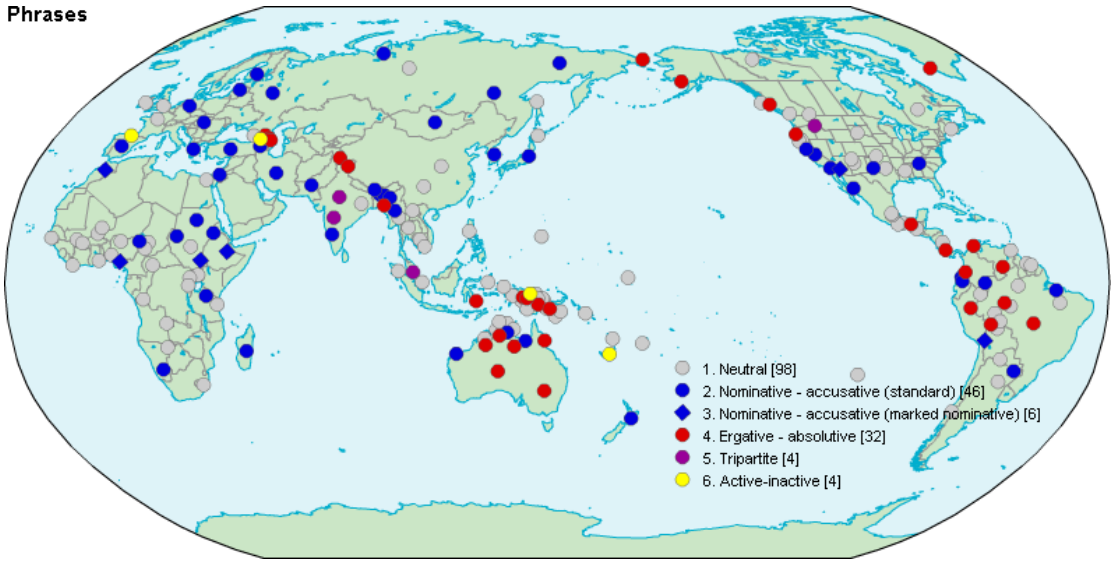


CLAUSE ALIGNMENT

Late PIE was almost certainly an accusative language. The hypothesis that it was ergative is not based on sound arguments. Of all the early IE dialects only Hittite shows a split-ergative pattern, and this seems like a recent development.

Languages of the NE Eurasia are mostly accusative. This includes all Uralic, Altaic, Sinitic, Mon-Khmer, In SW Eurasia the accusative type also predominates, e. g. Afro-Asiatic, Dravidian and Tai-Kadai languages are all accusative.

Ergative languages are found in the Caucasus (NW and NE Caucasian, Kartvelian as "split ergative" / active), in the Pamir (Burushaski) and the Himalayas, as well as in the Indian Subcontinent (Hindi is "split ergative"). There are ergative "residual zones" in the Siberia (Ket, Chukchi, Eskimo-Aleut), and the languages of the Ancient Near East were mostly ergative (Sumerian, Hattic, Hurrian).



This concludes our survey of typologically relevant features of PIE from an areal perspective. I have chosen not to talk about some features that are conspicuously absent in PIE and rare or non-existent in its daughter languages, but might be common elsewhere, e. g. "associative plural", honorific systems, alienable vs. inalienable possession, evidential marking, switch-reference, noun incorporation, case-stacking, etc. Looking at the distribution of such features among the languages of Eurasia might shed some new light on the areal affinities of PIE.

Our survey of the areal-typological features of PIE so far was rather superficial. Now we shall look at one particular grammatical category (gender) and the syntactic mechanism by which it is expressed (agreement) in some more detail.

DIACHRONIC AND AREAL TYPOLOGY OF GENDER AND OTHER AGREEMENT SYSTEMS
Università di Macerata, April 8-9 2008

1. Gender in PIE

PIE had three genders: masculine, feminine, and neuter. It is probable that this system developed from an earlier one with just two genders, which can be called common and neuter, but this hypothesis has been disputed, and it need not concern us here. In any case, the existence of a three-gender system in Late PIE, excluding the Anatolian branch, is well established.

The loci of gender agreement included adjectives, most pronouns, and numerals 1-4:

*so $h_1$esu-s wi$h_1$ro-s vs. *se-h_2 wesw-ih_2 meh_2tēr vs. *to-d wesu-0 $h_3$nomn
"that good man" "that good mother" "that good name"

Gender was not marked on nouns (i. e. the system was covert, rather than overt), but some suffixes were used to derive feminine nouns from masculine/neuter stems. It is probable that some adjectives did not distinguish between masculine and feminine forms. Neuter adjectives and pronouns were distinguished from the masculine ones only in the nominative and accusative singular, dual, and the plural.

The principles of gender assignment were partly semantic, partly formal. The gender of nouns near the two ends of the animacy hiearchy could be predicted from their meaning only. For example, names of substances and fluids, such as *$h_2$eyos "copper" and *wodr "water" are consistently neuter, but nouns denoting persons are masculine and feminine, e. g. *h_2nēr "man" is masculine, while *$h_1$wid^{h}eweh$_2$ "widow" is feminine. However, for nouns in the middle of the hierarchy formal principles applied: their gender was largely predictable from their characteristic suffixes and/or declensional patterns (Matasović 2004).

The morphemes expressing gender were always fused with morphemes expressing other categories, such as number and case.

2. Development of gender in IE languages:

The Anatolian branch will be excluded, because it is controversial whether the two-gender system found in that branch is an innovation, or an archaism (as argued in Matasović 2004).

3. 1. Indo-Iranian

In Sanskrit, attested since late 2nd millennium BC (the Vedas) the inherited gender system is preserved, with very few alterations. Although it is not completely correct to say that Modern Indo-Aryan languages descend directly from Sanskrit, most specialists agree that the proto-language was very close to it, almost as close as Latin is to the ancestor of the Romance languages.

Modern Indo-Aryan languages can be divided into three groups:

A) languages preserving all three IE genders: Konkani, Gujarati, Bhili, Khandeshi, Bhadarwahi, mostly spoken in the southwest of the Indo-Aryan part of the Indian Subcontinent.

B) languages with two genders, generally having merged the masculine and the neuter. They are spoken in the central part of India: Hindi, Punjabi, Sindhi, Lahnda, Central Pahari, and Rajasthani.

C) languages that lost gender altogether, spoken mostly to the east of Allahabad[4] in northern India, e. g. Bengali, Nepali, Assamese, Oriya, etc. In some of these languages (e. g. in Bengali), under the influence of Sanskrit, there is limited gender agreement with adjectives, but only in learned, higher varieties, never in the colloquial style.

A special case is represented by Sinhalese, where the gender system has been largely restructured; Sinhalese has animate and inanimate genders, and distinguishes masculine and feminine subgenders in the animate class, but the scope of gender agreement has been drastically reduced to just some verbal forms.

The development in the Iranian languages are by far the most complex ones in any branch of Indo-European. More different developments of the gender system are found here than in all other branches of Indo-European taken together. Whereas the gender system of Avestan, attested in the first half of the lst millennium BC, is rather similar to the one in Sanskrit, gender systems of the Middle Iranian languages have changed considerably. In Western Iranian languages, gender was lost by the Middle Iranian period in Pehlevi and Parthian. There, final syllables were lost in late lst millennium BC. Most Modern Western Iranian languages also lack gender, e. g. Modern Persian (including Tajik and Dari), Baluchi, Luristani (Luri), Gilyani, Mazanderani, Tati and Talysh. However, gender appears in some Northerh Kurdish dialects (Kurmanji).

Gender was generally better preserved in Eastern Iranian during the Middle Iranian period, though the neuter was generally lost (except in Sogdian). In Modern Eastern Iranian, the opposition of masculine and feminine is sometimes preserved, e. g. in Pashto, but all gender distinctions were lost in Ossetic, spoken on the Caucasus, in Parachi and Ormuri, in Southern Iran, and in Yaghnobi (in Tajikistan) and Wakhi (on the border of Chinese Turkestan).

[4] Zograf 1976: 51.

In the (Eastern Iranian) languages of the Pamir region, such as Shughni, Roshani, Munji, Yidgha and Yazgulami, an interesting development took place. The inherited gender systems (masculine and feminine) were reorganized, and the principles of gender assignment became strictly semantic. In languages such as Roshani[5], it is in principle possible to predict the gender of a noun from its meaning: for example, besides male humans, all abstract nouns and names of large animals are masculine, while feminines include names of concrete objects and small birds and animals.

The third group of Indo-Iranian is represented by Nuristani (Kafir) languages (attested only in the 19th century). They have the opposition of masculine and feminine gender, usually expressed on adjectives and some verbal forms, but not on pronouns. In Kati, there is agreement on adjectives in both attributive and predicative position, while in Waigali and Ashkun, there is agreement on adjectives and some verbal forms, but not on pronouns.

Zemyaki, a small language with just around 500 speakers, has lost gender; it is an enclave spoken in the region where Pashto is the dominant language (in Afghanistan); originally it is a language of some Waigali settlers. In that language there is also no agreement in number and case in adjectives.

3. 2. In Greek, attested (as Mycenaean) since late 2nd millennium BC, the three PIE genders are preserved. The distinction of the three genders perists in Modern Greek, except in the now extinct dialect of Capadocia, where gender oppositions were lost (presumably under Turkish influence). The rules of gender assignment for persons have remained mostly as they were in Classical Greek, but the formal rules of assignment were much simplified, so that the gender of the large majority of nouns can be predicted from their Nom. sg. ending[6]. Nouns receiving the Nom. sg. ending *-s* are masculine, nouns with the ending *-a* are feminine, and those ending in *-o* are neuter; the ending *-i* (written *-i* and *-ē*) is ambiguous, since it is shared by the feminine and neuter genders.

3. 3. Italic (and Romance)

Latin, substantially attested only since the 3rd century BC, is the only Italic language on which we have sufficient data for a close examination of its gender system. It is the direct ancestor of all the Romance languages. Its gender system more or less agrees with the one found in Greek and Sanskrit, but:

- agreement spread to some verbal forms formed analytically (with gendered participles and auxiliary verbs), e. g. *amatus est* (m.) "I was loved" vs. *amata est* (f.)

- there is a rather large class of adjectives distinguishing only two forms, one agreeing with m. and. f. nouns, and the other with neuters, e. g. *facilis* m. and f., *facile* n. "easy".

[5] Cp. Edel'man 1990: 161ff.
[6] Trijandafilidis 1995: 107ff.

The neuter merged with the masculine gender in Romance languages, partly as a consequence of the drop of the final consonants. The beginnings of this process can be observed already in Vulgar Latin. However, in Rumanian, the neuter remains as the class of nouns taking masculine forms of adjectives and pronouns in the singular, and feminine forms in the plural:

un prieten bun 'a good (m.) friend' vs. *prieteni buni* 'good (m.) friends'

un stilou bun 'a good (m.) pencil' vs. *stilouri bune* 'good (f.) pencils'

ună prietenă bună 'a good (f.) friend' vs. *prietene bune* 'good (f.) friends'

This class of nouns should probably be counted as a separate, third gender in Rumanian, since it is not residual (new borrowings can be assigned to this class).

3. 4. Celtic

In Continental Celtic, as far as we can tell, the gender system was rather symilar to the one in the classical languages; in Insular Celtic languages, attested from the late 6th century AD, with the loss of the final syllables, inflexional suffixes were largely replaced by a system of consonant mutations affecting following words in a phrase. Thus, in Old Irish, lenition of the word-initial consonant of the word immediately following a feminine noun in the N sg. is the indication that the noun is feminine, whereas masculine nouns in the N sg. do not cause any consonant mutations. Similarly, although definite articles are superficially identical in N sg. of the m. and f. genders, the feminine one lenites the following noun, while the masculine one does not:

in ben threbar "the clever woman"
/ín ven þrevar/

vs.

in fer trebar "the clever man"
/in fer trevar/

Similar rules of consonant mutations apply in the Brittonic languages Welsh, Cornish, and Breton. In those languages case inflection disappeared early, and gender was all but lost on adjectives during the Middle Welsh and Middle Breton periods. What was left was gender agreement on articles, reflected as consonant mutations on nouns following the feminine article.

The neuter gender is lost in Old Irish at the beginning of the Middle Irish period (10th century), while it disappeared with very little traces before Welsh, Breton, and Cornish were attested.

3. 5. Germanic

Germanic languages are substantially attested from the 4th century AD, with the corpus of Gothic texts. Other early attested Germanic idioms include Old High German, Old English, and Old Norse, and all these languages have preserved the PIE three-gender system, with extensive agreement on pronouns and adjectives. This system has been changed in different ways in the modern languages. German, Icelandic, Faroese, Nynorsk, and some varieties of Bokmal preserve the three gender system, sometimes with reduced agreement on adjectives, but with the added locus of agreement on articles (as in German). In English, gender distinctions were lost during the late Old English and early Middle English periods (11-12th century), so that only a pronominal gender system remains in the modern language. In Dutch (at least in the standard language), the masculine and the feminine genders merged, so the new common gender is opposed to the old neuter. Similar development took place in Frisian, (standard) Danish and in Swedish. In all these languages, as in Modern German, the main locus of gender agreement is the article, but there is agreement on adjectives as well, at least in some syntactic environments. Gender agreement was lost in Afrikaans, and in some Swedish dialects spoken in Finland.

3. 6. Armenian

In Armenian, attested since 5th century AD, gender was lost before the first monuments of that language. Very few traces of former distinctions of typically masculine and feminine suffixes exist, but all adjectives and pronouns have ony one form for referring to males, females and inanimates. That is, unlike Modern English, Classical Armenian does not distinguish between "he", "she", and "it".

3. 7. Tocharian

The two now extinct Tocharian languages, referred to as A and B, were spoken in Chinese Turkestan. Their documents are attested in the second half of the 1st millennium AD. In both languages the masculine and the neuter merged, but a large class of nouns take masculine forms of adjectives and pronouns in the singular, and feminine forms in the plural. They are traditionally called *alternantia*, and they can, if one so choses, be called the third gender in Tocharian. This development is quite parallel to the one we saw in Rumanian. Agreeing parts of speech include adjectives, many pronouns (but, interestingly, not relative and indefinite pronouns), and some numerals. Tocharian A is also noteworthy as the only IE language which has a separate feminine form of the 1st person sg. personal pronoun.

Besides the inherited opposition of masculine and feminine genders, Tocharian also developed a parallel, strictly semantic gender opposition of animate and inanimate genders, which intersects with the m./f. system. All animate nouns form a special oblique case, and adjectives also have this case form only when agreeing with animate nouns, irrespective of whether they are masculine or feminine[7]. Tocharian thus has what Corbett (1991: 184) calls a "combined" gender system, which appears to be rather rare typologically.

[7] Matasović 2004: 64.

3. 8. Slavic

The earliest attested Slavic language is Old Church Slavonic, with texts from the 9th century. It has three genders, and this system is preserved in all Slavic languages until today, except in a very few dialects, where the neuter gender has been lost, e. g. in the Molise dialect of Croatian, and in the Sele Fara dialect of Slovene[8]. Gender assignment in all Slavic languages is esentially preserved on adjectives, pronouns, and at least some numerals (depending on language). In Bulgarian and Macedonian, agreement spread to postponed articles, which developed from demonstrative pronouns, and in many Slavic languages the preterite tense of the verb, which is originally a past participle, also shows gender agreement.

An interesting development in Slavic is the creation of subgenders within the masculine gender, which occurred after the period of the earliest Slavic texts, and had different end-results in various Slavic languages. In Croatian, for example, in masculine animate nouns the accusative singular form is the same as the genitive form, while in the inanimate nouns it is the same as the nominative. Since the modifying pronouns and adjectives also have different forms in the two cases, there are clearly two different agreement patterns for two subclasses of masculine nouns, i. e., two subgenders:

ovoga velikog konja	*ovaj veliki stol*
this big horse	this big table

Similar, but not identical subgenders exist in other Slavic languages as well.

3. 9. Baltic

Baltic languages are attested only since the 14th century (Old Prussian) and 16th century (Latvian and Lithuanian). Old Prussian preserved the three PIE genders, while the masculine and the neuter merged in Latvian and Lithuanian before their first attestations. All Baltic languages preserve agreement on adjectives, pronouns, and some numerals. Latvian dialects spoken in Estonia have lost gender altogether, presumably under influence of Estonian.

3. 10. Albanian

Albanian is attested since the 15th century. Old Albanian still has three genders, but the class of neuters is small in number, and only slightly distinguished from the masculines. It appears that masculines and neuters are in the process of being merged in the modern language. Albanian has gender agreement on some adjectives and pronouns, but its main locus of gender agreement is on the rather complex set of preposed and postposed articles.

[8] Cp. Priestly 1983.

3. Shared developments

3. 1. A very common development in several branches of Indo-European is the merger of the masculine and neuter genders. This merger occurred in many Indo-Aryan languages, in Nuristani and Dardic languages, in several Iranian languages, in all Romance languages (with the possible exception of Rumanian), in the Goidelic and Brittonic branches of Celtic, in Tocharian, in Eastern Baltic, and, occassionally, in Slavic dialects (e. g. in the Slovene dialect of Sele Fara in Carinthia and in Molisan in Italy). The merger is all but complete in Albanian. The different times at which this merger occurred in various branches shows that this was an independent development, but it is easy to see how it was triggered. We saw that the distinction of masculine and neuter forms of most adjectives and pronouns had been rather weak, and limited to just a few case forms. The elimination of this distinction was therefore a natural step taken by many languages, especially those in which final syllables were lost or weakened. The merger of the masculine and feminine genders in Eastern Scandinavian (Danish and Swedish) and in Dutch is, on the other hand, an areally isolated phenomenon within Germanic.

Interestingly, in no branch of Indo-European has the number of genders actually been increased. The development of animate/inanimate and rational/irrational subgenders within the masculine gender in Slavic, and the development of animate/inanimate gender system, parallel to the inherited opposition between masculine and feminine genders in Tocharian, are not real exceptions to this claim.

Although the tendency toward reduction of the number of genders i can be observed in several other language families, it is certainly not universal. For example, it is almost generally accepted that the languages of the Nakh[9] group of the Eastern Caucasian family (Chechen, Ingush, and Batsbi), which have up to eight genders (in Batsbi[10]), actually innovated, and that the number of genders in Proto-East Caucasian did not exceed four (still distinguished in many languages, such as Archi). The number of genders in Proto-Niger-Congo languages is probably impossible to reconstruct, but it seems likely that it was less than the number of genders found in some West Atlantic languages, such as Fula, some dialects of which have 25 genders[11]. In this respect, then, Fula has innovated, which shows us that the number of gender distinctions can be increased in languages. It is also probable that the distinction between masculines and feminines that exists in some Dravidian languages (e. g. Tamil) does not go back to Proto-Dravidian, for which only the opposition between human and non-human genders can be reconstructed[12].

[9] Cp. Dešeriev 1963.
[10] Only five of these are represented by large noun classes, while the remaining three can be treated as exceptions.
[11] Koval' and Zubko 1986.
[12] Steever (ed.) 1998.

3. 2. In several languages, the old semantic and morphological motivation of IE genders was lost, so that gender assignment has become largely arbitrary, e. g., in Modern Irish, German, and French. In those languages the semantic core of the system is wery thin, and there are just some phonological cues indicating, but not predicting, the gender of particular nouns. On the other hand, a reverse trend can be observed in many languages, where gender has actually become *less* arbitrary than in PIE. In some languages, the rules of gender assignment have become largely morphological, e. g. in Russian, where the gender of a noun can quite regularly be predicted from its inflectional class. In the Pamir branch of Iranian languages, as we have seen, new semantic rules of case assignment have been introduced.

3. 3. The reduction of the scope of agreement can be observed in several languages: in some Germanic languages (e. g. in German), and in many Indo-Aryan languages adjectives have lost agreement in the predicative position, and it could be argued that adjectives in Modern Breton do not show any gender agreement (depending on how one represents consonant mutations). It has been claimed[13] that a language must first lose gender agreement in adjectives, and only subsequently in pronouns, but there are clear counter-examples to this claim. Hindi, as well as some other Indo-Aryan languages, has only one demonstrative form for males, females and inanimates (*vah*), but distinguishes between masculine and feminine forms of adjectives, and it also has agreement on verbs[14].

However, an opposite trend towards the broadening of the scope of agreement can also be observed. In many languages we find agreement on articles, which diachronically developed from demonstrative pronouns, e. g. in Greek, Romance, Insular Celtic, Balkan Slavic languages, and in many Germanic languages. Moreover, in many languages agreement has spread to verbal forms, usually those that had originally been inflected participles, e. g. in Hindi, Russian and Polish. In some languages, typologically more "exotic" patterns of gender agreement developed. For example, some West Flemish dialects developed gender agreement on complementizers[15], while the Italian dialect of Ripatransone[16], as described by Helmut Lüdtke (1974), has gender agreement of nominal objects with the subject of the clause, i. e. nominal objects receive different affixes depending on the gender of the subject:

l-u *frə'ki* *ča 'fam-u* "the boy is hungry" (lit. "has hunger")
art.-m. boy(m.) has hunger-m.

l-e *frə'kine ča 'fam-e* "the girl is hungry"
art.-f. girl(f.) has hunger-f.

[13] Cp. Priestly 1983. Cp. also Corbett 1991: 259, where it is stated that the loss of agreement on adjectives *usually* precedes the loss of agreement on pronouns, which may be true.
[14] McGregor 1986.
[15] Corbett 1991.
[16] Incidentally, the cases of West Flemish dialects and the dialect of Ripatransone shows how dangerous it is to limit oneself just to standard languages, because many typologically interesting developments of gender systems are found only in non-standard varieties.

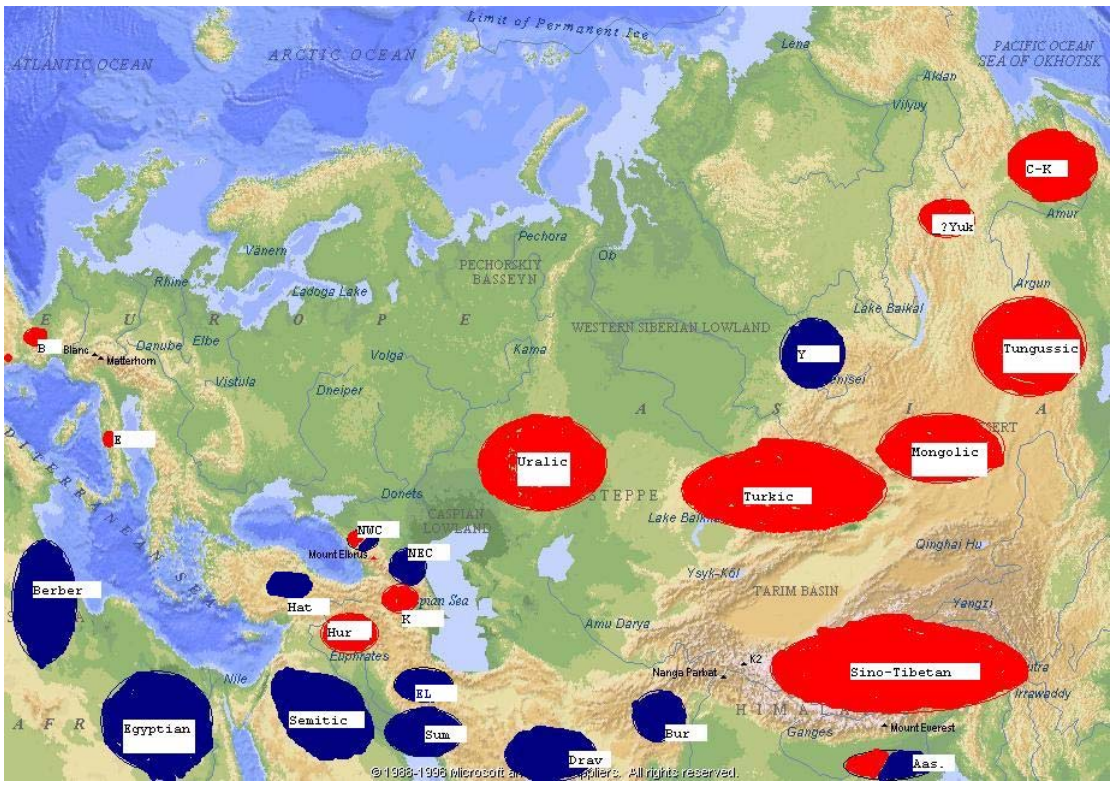
l-u *frə'ki* *ɛ 'it-u* *a 'ro:m-a* "the boy went to Rome"
art.-m. boy(m.) went-m. to Rome-m.

l-e *frə'kine* *ɛ 'it-e* *a 'ro:m-e* "the girl went to Rome"
art.-f. girl(f.) went-f. to Rome-f.

4. 4. A rather uncommon development is the complete loss of gender, which occurred in rather few languages. Actually, the most common tendency in Indo-European languages was the stubborn preservation of at least some gender distinctions. This is the subject of the last part of this paper.

4. The diachronic stability of gender

The remarkable cross-linguistic stability of gender has been noted by Johanna Nichols (1992, 1995, 2003). In Eurasia, there are several language families that are consistently genderless (e. g. the Sino-Tibetan[17], Miao-Yao, Tai-Kadai languages in SE Asia, the Turkic, Tungusic, and Mongolian languages, Kartvelian, Uralic, and Chukchi-Kamchadal languages, as well as the isolates such as Ainu, Japanese, Nivkh and Basque). On the other hand, there are language families in which the overwhelming majority of languages has gender, such as Afroasiatic, Dravidian, NE Caucasian, as well as the nearly extinct Yeniseian family in Siberia and the isolate Burushaski in Pakistan. It is fair to say that gender, as a category, characterizes whole language families rather than individual languages.



[17] Some languages, e. g. Newari, have agreement in animacy between nouns and adjectives (and/or demonstratives). As far as I can tell, this comes very close to what is in other families called gender agreement (with animate/inanimate genders), see e. g. Korolev 1989.

This conclusion is fully confirmed by our examination of gender in IE languages (Matasović 2004): of 144 members of the IE family listed by Ruhlen (1991), gender is preserved in at least[18] 109 (75,69 %). Of around thirty languages that lost gender, the large majority belong to the Indo-Iranian branch.

Similar statistics apply to other language families of Eurasia; of 28 Dravidian languages, gender is lacking in three (Toda, Kurruba[19] and Brahui), which gives us 89, 29 % of gender retention rate. In NE Caucasian, 26 out of 29 languages have retained gender (it is absent in Udi, Lezgian and Agul)[20], so the retention rate is 89, 65 %. Similar statistics would doubtlessly apply to Afroasiatic and other gendered families.

5. Gender or Agreement?

Although it is perfectly clear that gender is stable both genetically and areally, a more detailed cross-linguistic look at this category is in order. Gender is defined as grammatical agreement based on nominal classification (Corbett 1991, 2006)[21]. These are clearly logically independent parameters, since there is nominal classification without agreement, and vice versa, there are other systems of agreement that do not involve nominal classification. The question to be answered is: is the particular co-occurrence of agreement and nominal classification stable, or is it rather the more general principle of agreement?

Agreement is a rather complex phenomenon; in order for two words to agree, several conditions must be met:

A bound morpheme A (agreement) has to appear in a particular word T (target), if that word is in a syntactic relation R to a word belonging to a class C (controller), which belongs to, or expresses the grammatical category X.

Thus, agreement involves the following conditions:

a) the obligatory presence of a bound morpheme[22] A, triggered by C, on T
b) the specification of the syntactic relation R between T and C
c) the obligatory expression of the grammatical category X on C

The complexity of the conditions a-c show why rules of agreement can be difficult to acquire. Of course, children are perfectly capable to acquire even most complex rules during the critical period of language acquisition, but acquisition of grammatical complexities at a later stage is seriously hampered. On

[18] For five languages, all of them Iranian, I simply did not have available data to ascertain whether they have gender or not.
[19] Also known as Kurrumba, according to some linguists just a southern dialect of Kannada.
[20] Gender is also apparently absent in some southern dialects of Tabasaran.
[21] The classification in question does not need to be entirely based on semantic criteria, but all gender systems have a semantic core (Corbett 1991).
[22] As is well known, the distinction between bound and free morphemes is not sharp; this is why it is difficult to dismiss some instances of "clittic agreement" adduced by Corbett 2006.

the other hand, nominal classification *per se* is not difficult to acquire, involving only the learner's ability to memorize to which noun class a particular word belongs. Moreover, since there is a semantic principle at the basis of all nominal classification, the learner does not have to memorize the gender of each item individually: he or she can predict it from the general principles. Making mistakes in cases where gender classification appears to be largely arbitrary (as in German or French) will not affect the *existence* of gender, as a category, for as long as the easily learnable general semantic principle of gender assignment remains intact. However, if the learner is unable to learn the complex conditions that apply to nominal agreement, as a grammatical rule, then the whole system of categories based on agreement, including gender, will collapse.

Thus, *a priori* considerations seem to lead us to assume that gender will be a diachronically very stable category in languages that have agreement in their inventory of basic grammatical rules that any learner acquires in the critical, early period of language acquisition. On the other hand, in situations where normal acquisition and transmission of language is interrupted (i. e. in cases of pidginization and language shift), it will not be specifically gender that is lost, but agreement in general. The rest of this paper is dedicated to testing these hypotheses.

This present investigation is limited to:

(1) languages of the Old World (Eurasia and Africa), because here the genetic relationships between language families are well explored and undisputed[23].

(2) agreement in the nominal categories of Gender, Case, and Number, i. e. nominal categories that have the noun as the controller and the NP as the primary domain. The category of person is excluded for theoretical reasons (which I cannot discuss here), but also because person agreement is nearly omnipresent in non-isolating languages, so its presence is clearly not a good predictor of genetic or areal relationships. I have also excluded some typologically rare and/or theoretically suspect patterns of nominal agreement (e. g. agreement in definiteness in Hebrew, or "honorific agreement" in Korean[24]).

Thus, our investigation is focused on the typological contrast between two NP structures, exemplified by Latin and Basque:

h-aec	*puell-a*	*pulchr-a*
dem.-Nsg.F	girl-Nsg.F	pretty-Nsg.F

h-arum	*puell-arum*	*pulchr-arum*
dem.-Gpl.F	girl-Gpl.F.	pretty-Gpl.F

etxe handi-a

[23] Some would even argue that the traditional notion of genetic relatedness cannot be applied to some macroareas, such as Australia (see Dixon 1997).
[24] See Corbett 2006.

house big-ABS.def.

etxe handi-a-k
house big-ABS.def.-pl.

etxe hau
house this

etxe hau-ek
house this-ABSpl.

6. Evidence for genetic and areal stability of nominal agreement

Clearly, if nominal agreement is a better predictor of genetic and areal relatedness, then language families and/or areas will present a greater degree of variation with respect to the presence vs. absence of gender, than with respect to the presence vs. absence of nominal agreement. Of course, since there is no gender without agreement, it is to be expected that gender will appear genetically and/or areally stable even if nominal agreement is the fundamental feature, i. e. if the stability of gender is just a consequence of the more general stability of agreement. It is just that we expect nominal agreement to appear *even more* stable than gender.

Our database contains 120 languages of Eurasia and Africa; each entry contains information about the area in which the language is spoken, its affiliation, and whether it has nominal agreement with respect to gender, number, and case. Moreover, the database contains language codes (according to the last edition of *Ethnologue*), notes about peculiarities of agreement in each language, and whether categories which do not show agreement are expressed morphologically, or not. Existing agreement patterns are illustrated by examples, contained in a separate file. The database also contains references from which data were collected, as well as the information whether the language in question belongs to a family that originally had gender, or not.

The database is still being developed, so it has some weak points:

-some language families are still underrepresented (especially the Tibeto-Burman, Khoisan, and Nilo-Saharan languages)
-the sample is biased towards Indo-European languages, especially those that have lost gender;
-the reliability of our sources is uneven; it is especially difficult to prove that something does not exist in a language, and so it is with agreement. A minor (or "reduced") agreement pattern can exist without being noted by a descriptive grammar.

DISTRIBUTION OF LANGUAGES WITH AND WITHOUT GENDER

Although the exact statistics do not tell us much (because of the biases in the sample), some patterns can still be discerned by analysing the data:

(1) Nominal agreement is a good predictor of areal distribution of languages. There appears to be less variation in agreement within language areas than in gender, so nominal agreement seems to be a better predictor of areal relatedness than gender.

(2) Nominal agreement is not a better predictor of genetic relatedness than gender; within language families there is actually more variation in the presence vs. absence of nominal agreement than in the presence vs. absence of gender.

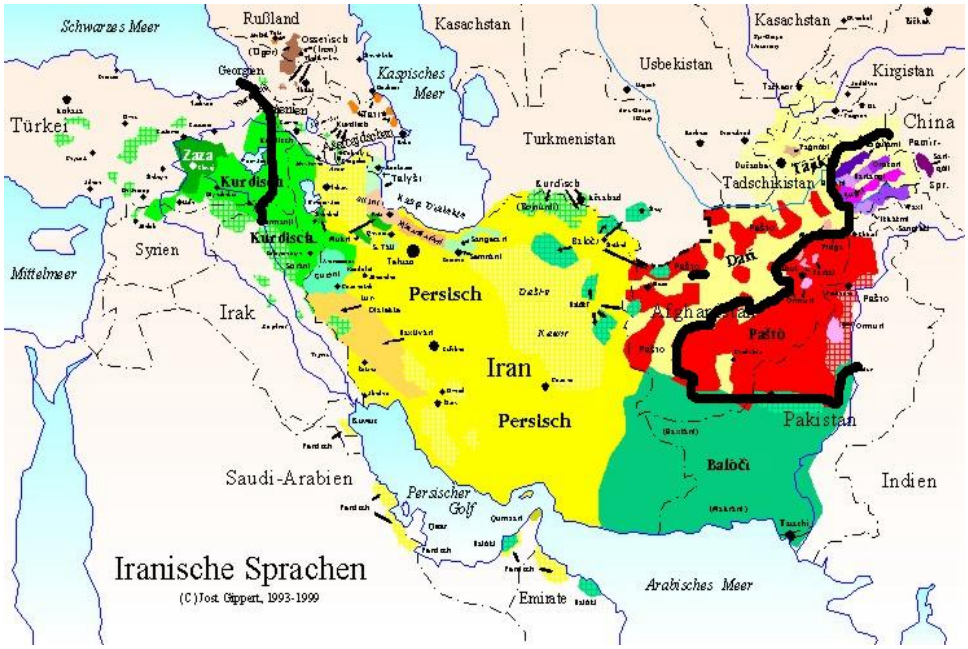
Of the 22 language families in our sample, 19 of which are represented with more than a single language, 5 are consistent with respect to presence vs. absence of gender, while 4 are consistent with respect to presence vs. absence of nominal agreement. However, these figures do not tell us much, because large families such as Afroasiatic and Indo-European are treated the same as small ones, such as Kartvelian and Chukchi-Kamchadal. Also, all families consistent with respect to the absence of agreement are also consistent with respect to the absence of gender, and all families consistent with respect to the presence of gender are also consistent with respect to the presence of nominal agreement. So the indicative cases are families consistent with respect to the absence of gender (4) and those consistent with respect to the presence of nominal agreement (1). These figures give us a better indication that gender is much more genetically stable than nominal agreement.

In the case of languages that differ, with respect to nominal agreement, from the default of the family to which they belong, there is in every single instance an

areal explanation: such languages are spoken in areas, in which languages that differ from the default of their family, predominate. Languages without agreement belonging to families with agreement all tend to cluster areally, as do languages with agreement belonging to families without agreement.

This is the case, e. g., with:

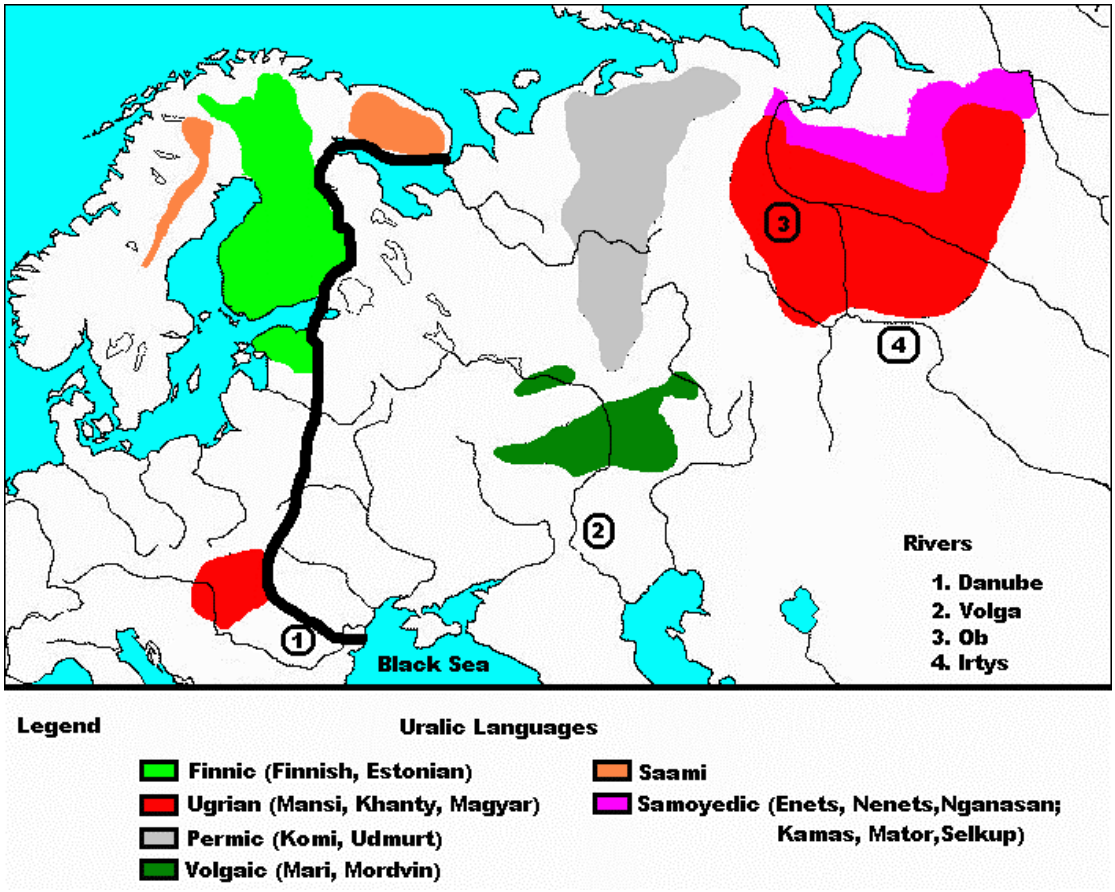
-Armenian, spoken on the border of languages with and without agreement (Turkic, and those Iranian languages that lost agreement)
-Indo-Iranian languages without nominal agreement (e. g. Persian, Oriya, the Pamir languages)



-NE Caucasian languages that lost gender, all spoken on the southern periphery of the area of the and bordering Turkic and Iranian languages without agreement.

Genderless NE Caucasian languages

-Uralic languages that have agreement (e. g. Finnish, Saami, and Hungarian), all spoken on the western periphery of the Uralic family and surrounded by Indo-European languages with agreement.



-Dravidian languages that lost agreement, such as Brahui (surrounded by Baluchi which has no agreement), as well as Kurruba and Toda, which form a compact group on the western fringe of the Dravidian area.

-Mande languages in West Africa, which form a compact group neighboring the Kru languages, which also lack agreement, and with Songhai and Dogon, from the Nilo-Saharan family (also without agreement).

6. Loss of gender and loss of nominal agreement

Loss of gender in a language can be the result of a wholesale loss of nominal morphology, as in English or Afrikaans; such cases are not very instructive: if all noun morphology is lost, gender will be lost as well as all other nominal agreeing categories. This is, of course, trivial. Intersting cases are those languages that lost gender, but preserved rich nominal morphology, including case and number, e. g. Armenian, many Iranian languages, as well as NE Caucasian and Dravidian

languages that lost gender. Ossetic, for example, has lost gender; it still has a rich case and number inflection, but does not have any agreement (Abaev 1964):

mae zaerond fyd "my old father"
mae zaerond fydaen "to my old father"
nae zaerond fydaltae "our old fathers"
nae zaerond fydaeltaen "to our old fathers"

As will be shown below, such wholesale loss of all nominal agreement is the default development when gender is lost. Similar evidence for the causal relationship can be found in Pidgins and Creoles. Pidginization does not have to involve the loss of gender, or other forms of agreement. Michif, which is the result of contact between two gendered languages (French and Cree) still has a rich system of gender agreement. But all pidgins and creoles known to me to have lost gender, such as Haitian Creole French, Negerhollands, and Sango, also lost all nominal agreement.

Conversely, the rise of agreement in languages belonging to families that do not have it also seems to be motivated by language contact. In Uralic, we find case and number agreement only in westernmost languages such as Hungarian, Finnish, and Saami (e. g. Swedish Saami):

Finnish (Szinnyei 1922: 71):

matala-ssa möki-ssä "in the small cottage"

piene-t poia-t "small boys"

In Hungarian only some demonstratives agree, while adjectives in the attributive position remain uninflected (they agree in number in the predicative position). Only one demonstrative pronoun shows full case and number agreement in the attributive position (this is the "reduced agreement" we mentioned earlier):

ír-om *az-t* *a* *level-et*
write-1sg.pres. DEM-acc.sg. ART letter-acc.sg.
"I am writing this letter"

Similarly, some genderless Mande languages have developed number agreement through contact with neighbouring Gur languages, which have extensive gender and number agreement (Dienst 2004: 61). Cp. the examples from Bobo:

sù *kabè* *sia* *kobá*
cup.sg. empty.sg. cup.pl. empty.pl.

In the present investigation I have examined 29 languages in which gender was lost. These languages all belong to families that have gender, as a rule, and for which gender can be reconstructed in the proto-language. The large majority of

the languages in my database have at least some nominal morphology, and some of them actually have rich nominal inflection.

I have excluded from investigation languages from families for which it is at present uncertain whether they have originally had gender, or not. There are two such families in Eurasia (NW Caucasian and Austro-Asiatic) and two in Africa (Nilo-Saharan and Khoisan[25]). In any case, some languages belonging to these families have gender while others do not; in NW Caucasian, Abkhaz and Abaza have gender, while Adyghe, Kabardian and the extinct Ubykh do not have it. Within Austro-Asiatic, the languages spoken in India (Munda languages, Khasi and others) have gender as a rule, while languages spoken in Indo-China (e. g. Vietnamese, Khmer, etc.) do not have it, without exception. As a general rule, in those families, languages that do not have gender, also do not have case and number agreement.

LANGUAGES WITH GENDER LOSS

It is clear from our database that case and number agreement are typologically much rarer than gender agreement. However, morphological case and number are not rare at all, being present in the majority of languages in our database. Number marking of some kind is all but universal in languages with any nominal morphology. There are languages in which number does not exist as a category (e. g. Chickasaw), but such languages are not included in our database.

The examination of the areal distribution of languages with gender loss shows that they always cluster at the perifery of areas where gendered languages predominate. This is the case, e. g., with Mande languages in Africa, with the Dravidian languages that lost gender, as well as with Armenian and Indo-Aryan

[25] The genetic validity of Nilo-Saharan and, especially, Khoisan, is disputed.

languages without gender. In the Caucasus, too, the NE Caucasian languages that lost gender are situated on the southern border of the family, in direct contact with the genderless Turkic and Iranian languages.

Our database shows that of the 29 languages that lost gender, other forms of nominal agreement were preserved in just five, and even in those languages (except in Bobo, on which see below) we have only what we can call reduced agreement. This is the agreement pattern usually involving only a handful of targets and/or domains, as in the English examples below:

this book these book-s
that book those book-s

The same phenomenon is found, e. g. in Brahui, where just a few pronominal adjectives agree, e. g. bāz *"many" (Andronov 1980: 55):*

rupaīt-eāy bāzāy "for many rupees"
rupee-lative many-lative

Other adjectives and pronouns do not agree (Andronov 1980: 96):

juānō inām-as
"good present"

juānō 'ullī
"good horses"

In Sembla (Mande), adjectives agree only with nouns that have irregular plurals:

nìgì bare *nìgì baré*
cow fat cow fat.PL
„fat cow“ „fat cows“

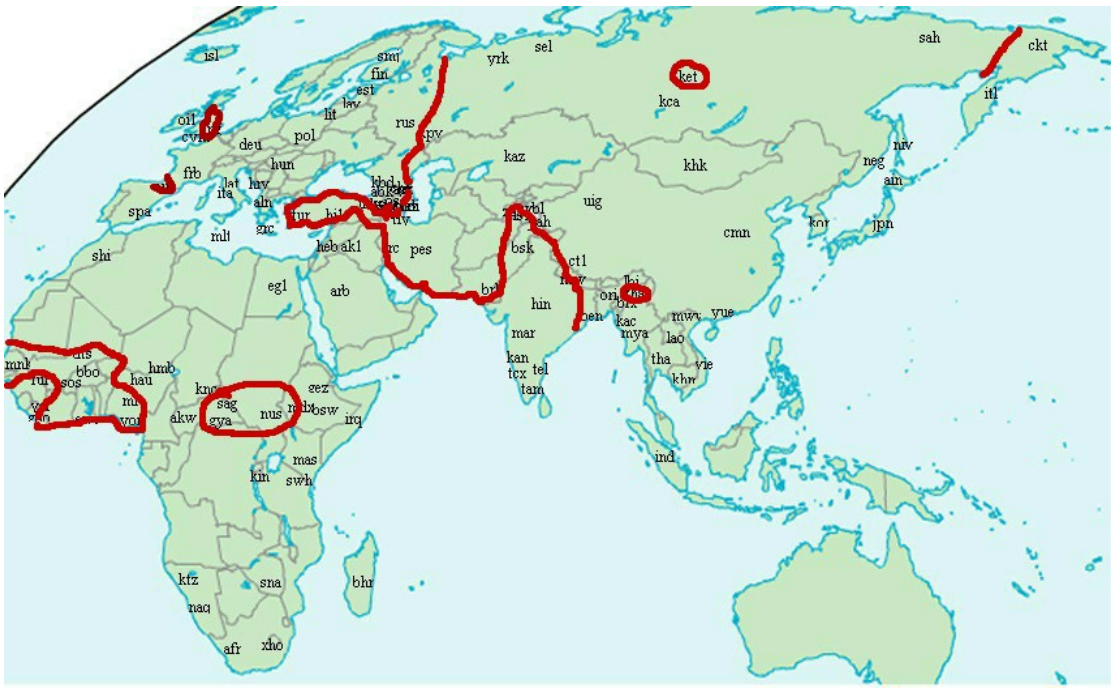
à don be *à dzuró bé*
DEM child DEM.SG DEM child.PL DEM.PL
„this child“ „these children“

(*don* "child" has a suppletive plural *dzuró*)

In one case it is difficult to decide whether the loss of agreement was total because of the nature of evidence. Classical Armenian still has limited number and case agreement[26]; however, in Modern Armenian, all agreement is lost[27]. One has to bear in mind that almost all texts in that language represent translations from Biblical and Byzantine Greek, and that the syntax of Classical Armenian

[26] See Meillet 1936: 137; basically, adjectives agree with the head nouns if they follow them, but not if they precede them, cp. *bazum awurkᶜ* "many days" but *awurkᶜ bazumkᶜ* "id.".
[27] Abeghian 1936: 138.

was certainly heavily influenced by Greek syntax. It is at present unknown to what extent was Classical Armenian ever a spoken language, but Meillet noted (1962: 40) that: "Cet ordre de mots [sc. agreeing adjectives following nouns], fréquent dans les traductions des textes sacrés où l'original grec est, autant que possible, rendu en arménien sans aucun déplacement de mots, est sensiblement plus rare dans les textes originaux." Meillet (*ibid.*) also notes many instances of non-agreeing postposed adjectives. Therefore, the case of Armenian is at best inconclusive.



This means that, even when agreement is no longer a productive grammatical process, it can remain as an archaism. But there are no cases, to my knowledge, where other forms of nominal agreement survived the loss of gender as a fully productive grammatical process. In my opinion this confirms that it is the presence of agreement, as a type of grammatical rule (rather than gender, as a category), that is actually subject to change, and that this change usually occurs under areal pressure (more often than not in cases of interruption of normal language transmission).

Family	*no. of languages*	*with gender*	*with nominal agreement*	*with reduced agreement*
Afroasiatic	13	11	11	0
Austro-Asiatic	3	1	1	0
Austronesian	2	0	0	0
Chukchi-Kamchatkan	2	0	1 (?)	0
Dravidian	5	3	3	1

Indo-European	32	20	20	2
Isolates	6	1	1	0
Kartvelian	2	0	2	0
Khoisan	2	1	1	0
Miao-Yao	1	0	0	0
Mongolian	1	0	0	0
NE Caucasian	4	1	1	0
Niger-Kordofanian	16	6	6	2
Nilo-Saharan	5	1	2	0
NW Caucasian	3	1	1	0
Sino-Tibetan	8	1	1	2 (?)
Tai-Kadai	2	0	0	0
Turkic	4	0	0	0
Uralic	8	0	4	2 (?)
Yeniseian	1	1	1	0

Since gender markers are so often fused with number and/or case markers, the loss of gender agreement nearly always (through phonological erosion) involves the loss of other forms of agreement (in number and case). Case and, especially, number are cognitively and communicatively more important categories, so they are usually quickly re-introduced into the morphological system of the language in question (through the common grammaticalization patterns, such as cliticization of adpositions). So it comes about that languages that lose gender often have case and number as categories, although they do not have any agreement: case and number are renewed as morphological categories, while gender, once lost, is almost never renewed. This explanation is perfectly possible, but in desperate need of a proof. It is far from certain that phonological erosion of agreement markers is the universal, or even the most common, cause of loss of agreement. It is equally possible that loss of agreement precedes the phonological erosion of affixes which are no longer needed for distinguishing different agreeing categories.

7. Conclusion and some questions

1. Not only gender, but all kinds of nominal agreement are diachronically stable, though gender agreement is more common than number and case agreement. It appears that gender is more genetically stable than other kinds of nominal agreement. This means that it will be preserved, as a rule, in normal conditions of language transmission (unless language shift / pidginization occurs).

2. Gender appears to be less areally stable than nominal agreement in general. There are a number of areas where both gendered and genderless languages are found (W Africa, Caucasus, NE Europe), but those areas are much more homogenous with respect to agreement (languages without gender often have other types of nominal agreement). There are documented cases of languages

acquiring case and number agreement through language contact (Bobo, Baltic Fennic languages, perhaps Classical Armenian), but no instances of languages acquiring gender through contact. This means that the presence of the type of rule involved in nominal agreement can be acquired or lost if (but not necessarily only if) the normal conditions of language transmission are interrupted (through language shift or pidginization).

3. Loss of gender as a rule involves the loss, or radical reduction, of other kinds of agreement.

Agreement loss appears to be a strongly areal phenomenon; languages that lose gender usually cluster together and are spoken in contiguous areas where languages without agreement predominate.

Finally, all our conclusions so far are supported only by data from languages of the Old World (Eurasia and Africa). It is quite possible that examining languages from Australia, Oceania and the Americas would yield a different picture. However, in order for this picture to eventually appear, a lot of work has to be done, especially in the field of genetic classification of the languages of the New World.

BIBLIOGRAPHY

Abaev, V. I. *A Grammatical Sketch of Ossetic*, Mouton, The Hague 1964.

Abaev, V. I. (ed.) *Osnovy iranskogo jazykoznania*, Nauka, Moscow 1981.

Abeghian, A. *Neuarmenische Grammatik*, De Gruyter, Berlin 1936.

Abondolo, D (ed.). *The Uralic Languages*, Routledge, London 1997.

Aksenova, I. A. & Toporova, I. N. *Grammatika jazyka akva*, URSS, Moscow 2002.

Allan, E. J. "Dizi" in Bender, M. Lionel, *The Non-Semitic Languages of Ethiopia*. Michigan State University, East Lansing 1976: 377-392.

Ambros, A. A. *Bongornu, kif int? Einführung in die maltesische Sprache*, Reichert, Wiesbaden 1998.

Andronov, M. S. *The Kannada Language*, Nauka, Moscow 1969.

Andronov, M. S. *Dravidian Languages*, Nauka, Moscow 1970.

Andronov, M. S. *The Brahui Language*, Nauka, Moscow 1980.

Andronov, M. S. *Grammatika tamil'skogo jazyka,* Nauka, Moscow 1987.

Aristava, Š. K. *et alii*, *Grammatika abxazskogo jazyka*, Suhumi 1968.

Barreteau, D. *Aspects de la morphologie nominale du mofu-gudur*, (MS) 1977, http://www.bondy.ird.fr/pleins_textes/pleins_textes_5/b_fdi_14-15/20553.pdf

Bickel, B. "Grammatical relations, agreement, and genetic stability", MS (available from http://berkeley.socrates.edu/~bickel), 1999.

Brauner, S. *A Grammatical Sketch of Shona including Historical Notes*, Koppe, Köln 1995.

Bykova, E. M. *The Bengali Language*, Moscow, Nauka 1981.

Cincius, V. I. *Negidal'skij jazyk*, Nauka, Leningrad 1982

Claudi, U. *Zur Entstehung von Genussystemen. Überlegungen zu einigen theoretischen Aspekten, verbunden mit einer Fallstudie des Zande*, Buske, Hamburg 1985.

Comrie, B. *The Languages of the Soviet Union*, CUP, Cambridge 1981.

Corbett, G. *Gender*, CUP, Cambridge 1991.

Corbett, G. "Agreement",in: Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann (eds.) *Syntax: An International Handbook of Contemporary Research*, vol. 2, Berlin: de Gruyter, 1995: 1235-1244.

Corbett, G. *Number*, CUP, Cambridge 2000.

Corbett, G. *Agreement*, CUP, Cambridge 2006.

Corbett, G. & Hayward, R. J. "Gender and Number in Bayso", *Lingua* 73/1987: 1-28.

Creissels, D. "Typology", u: B. Heine & D. Nurse (eds.), *African Languages. An Introduction*, CUP, Cambridge 2000: 231-258.

Čaušević, E. *Gramatika turskoga jezika*, HSN, Zagreb 1996.

Dekeyser, X. "The Diachrony of the gender systems in English and Dutch", in: *Historical Morphology,* ed. by J. Fisiak, The Hague: Mouton 1980: 97-111.

Dešeriev, Ju. D. *Sravnitel'no-istoričeskaja grammatika naxskix jazykov*, URSS, Moscow 1963 (reprinted 2006).

Diakonov, I. M. *Afrasian Languages*, Nauka, Moscow 1988.

Dienst, S. *Nominalmorphologie und Numerus im Bobo (Mande)*, Working Papers 34, Johannes Gutenberg-Universität, Mainz 2004.

Dirr, A. *Theoretisch-praktische Grammatik der modernen georgischen (grusinischen) Sprache*, Vienna and Leipzig 1904.
Dixon, R. M. W. *The Languages of Australia*, CUP, Cambridge 1980.
Dixon, R. M. W. *The Rise and Fall of Languages*, CUP, Cambridge 1997.
Ėdel'man, D. I. *Sravnitel'naja grammatika vostočnoiranskix jazykov. Morfologija. Ėlementy sintaksisa*, Moscow: Nauka 1990.
D. I. Édel'man (ed.) *Dardskie i nuristanskie jazyki*, Indrik, Moscow 1999.
Efimov, V. A. "Obščie čerty ėvoljucii kategorii roda v iranskix jazykax", in: *Iranskoe jazykoznanie*, Nauka, Moscow: 23-34.
Englund, R. & Wolf, W. *Finnische Sprachlehre*, Groos, Heidelberg 1960.
Frank, W. J. *Nuer Noun Morphology*, M.A. Thesis submitted to SUNY Buffalo, Buffalo 1999.
Gamkrelidze, T. & Ivanov, V. *Indoevropejskij jazyk i indoevropejcy, I-II,* Izdatel'stvo tbilisskogo universiteta, Tbilissi 1984.
Greenberg, J. (ed*.) .Universals of Human Language: IV: Syntax.* Stanford University Press, Stanford, California, 1978
Grjunberg, A. L. *Jazyk kati*, Moscow 1980.
Güldemann, T. "Noun Categorization in Non-Khoe lineages of khoisan", *Afrikanistische Arbeitspapiere Köln* 63: 5-33.
Hagman, Roy S. *Nama Hottentot Grammar.* Bloomington: Indiana University 1977
Handsah, R. C. & N. C. Murmu, *A Framework for Learning and Understanding Santali in Ol Chiki Script*, Wesanthals 2005.
Hannah, H. B. *A Grammar of the Tibetan Language*, Motilal Banarsidas, Delhi 1972.
Heine, B. & Nurse, D. (eds.) *African Languages. An Introduction*, CUP, Cambridge 2000.
Heine, B. & Reh, E. *Grammaticalization and Reanalysis in African Languages*, Hamburg: Buske 1984.
Hodge, S. *An Introduction to Classical Tibetan*, Airs & Phillips, Warminster 1990
Hoffmann, C. *A Grammar of the Margi Language*, OUP, Oxford 1963.
Hudson, D. F. *Bengali*, The English Universities Press, London 1965.
Huffman, F. E. *Modern Spoken Cambodian*, Cornell University, Ithaca 1970.
Hurskainen, A. "Noun Classification in African Languages", in: Unterbeck & Rissanen (eds.) 2000: 665-687.
Hutchinson, J. P. *The Kanuri Language: A Reference Grammar*, University of Wisconsin, Madison 1981.
Innes, G. *An Introduction to Grebo*, School of Oriental and African Studies, London 1966.
V. K. Jakovleva, *Jazyk joruba*, Nauka, Moscow 1963.
Janhunen J. "Grammatical Gender from East to West", in: Unterbeck & Rissanen (eds.) 2000: 689-708.
Johanson, L. & Csató, E. Á. (eds.) *The Turkic Languages.* Routledge: London, 1998.
Karamšoev, D. "Problema kategorii roda v pamirskix jazykax", *Voprosy jazykoznanija* 6/1979: 15-33.
Katenina, T. E. *Očerk grammatiki jazyka marathi*, Izdatel'stvo literaturi na inostrannyx jazykax, Moscow 1963.
Kimenyi, A. *A Relational Grammar of Kinyarwanda*, University of California Press, Berkeley 1980.

Kirsch, B. *et alii*, *Teach Yourself Xhosa*, Hodder Arnold, London 2004.
Klimov, G. A. *Vvedenie v kavkazskoe jazykoznanie*, Nauka, Moscow 1987.
Klimov, G. A. "Megrel'skij jazyk" in: N. A. Alekseev, *Kavkazskie jazyki*, Academia, Moscow 1999.
Klimov, G. A. & Edel'man, D. I. *Jazyk burušaski*, Nauka, Moscow 1970.
Korolev, N. I. *Nevarskij jazyk*, Nauka, Moscow 1989.
Koshal, S. *Adakhi Grammar*, Motilal Banarsidas, Delhi 1979.
Koval', A. I. & Zubko, G. B. *Jazyk fula*, Moscow: Nauka 1986.
Krejnovič, E. A., "Immenye klassy i grammatičeskie sredstva ix vyraženij v ketskom jazyke", *Voprosy jazykoznanija* 2(1961): 106-116.
Kuno, S. *The Structure of the Japanese Language*, The MIT Press, Cambridge 1973.
Kwee, J. B. *Indonesian*, Hodder & Stoughton, London 1965
Lehmann, C. "Universal and typological aspects of agreement", in: Hansjakob Seiler and Franz J. Stachowiak (eds.) *Apprehension: Das sprachliche Erfassen von Gegenständen.* Tübingen: Narr, II, 1987: 201-267.
Lehmann, C. "On the Function of Agreement", in: Michael Barlow and Charles A. Ferguson (eds.) *Agreement in natural language: approaches, theories, descriptions.* Stanford: CSLI Publications, 1988: 55-65.
Lüdtke, H. "Die Mundart von Ripatransone - ein sprachtypologisches Kuriosum", *Acta Universitatis Carolinae, Philologica 5*/1974: 173-177.
Magometov, A. A. *Agul'skij jazyk*, Mecniereba, Tbilissi 1970.
Marchese, L. "Noun classes and agreement systems in Kru: a historical approach", in: M. Barlow & C. A. Ferguson (eds.) *Agreement in Natural Language: Approaches, Theories, Descriptions*, CSLI, Stanford 1988: 323-341.
Matasović, R. "Distribucija indoeuropskoga fonema *r: kavkasko-indoeuropska tipološka podudarnost?", *Suvremena lingvistika* 34/1992: 201-207.
Matasović, R. "Proto-Indo-European *b and the Glottalic Theory", *JIES* 22/1994: 133-149.
Matasović, R. *Gender in Indo-European*, Winter, Heidelberg 2004.
Matasović, R. *A Short Grammar of Kabardian*, http://deenes.ffzg.hr/~rmatasov, 2007.
Matasović, R. *Jezična raznolikost svijeta* ("Language Diversity of the World"), Algoritam, Zagreb 2005b.
McGregor, R. S.*Outline of Hindi Grammar*, Oxford University Press, Oxford 1986.
Meillet, A. *Esquisse d'une grammaire de l'arménien classique*, Vienna, Mekhitaristes 1936.
Meillet, A. *Recherches sur la syntaxe comparée de l'arménien*, Imprensa nacional de Lisboa, Lisabon 1962.
Menu, B. *Petite grammaire de l'égyptien hiérogyphique*, Geuthner, Paris 2002.
Meyer, R. *Hebräische Grammatik* I-IV, De Gruyter, Berlin 1966.
Miao Language Team "A brief description of the Miao language", in: H. C. Purnell Jr. ed., *Miao and Yao Linguistic Studies: Selected Articles in Chinese*, Cornell University, Ithaca 1972.
Mithun, M. *The Languages of Native North America*, CUP, Cambridge 2001.
Moravcsik, Edith A,. "Agreement", in: Joseph H Greenberg (ed.) *Universals of Human Language: IV: Syntax.* Stanford University Press, Stanford, California, 1978: 331-374.

Morev, A. N. *Laosskij jazyk*, Nauka, Moscow 1972.
Moseley, C. *Colloquial Estonian*, Routledge, London 1994.
P. Muysken, "The Grammatical Elements in Negerhollands: Loss, Retention, Reconstitution", MS, 2006.
Nagaraja, K.S., *Khasi, a Descriptive Analysis,* Poona, Deccan College, 1985.
Nichols, J. *Linguistic Diversity in Space and Time*, The University of Chicago Press, Chicago 1992.
Nichols, J. "Diachronically stable structural features", in: H. Andersen (ed.) *Historical Linguistics 1993,* Benjamins, Amsterdam: 337-355.
Nichols, J. "Diversity and Stability in Language", in: B. D. Joseph & R. D. Janda (eds.), *The Handbook of Historical Linguistics*, Blackwell, Oxford 2003: 283-310.
Nichols, J. "Chechen Morphology (with Notes on Ingush", MS, 2007.
Omel'janovič, N. V. *Samoučitel' birmanskogo jazyka*, Izdatel'stvo "Meždunarodnye otnošenija", Moscow 1971.
Oranskij, I. M. *Iranskie jazyki*, Nauka, Moscow 1963.
Paxalina, T. N. *Sravnitel'no-istoričeskaja morfologija pamirskix jazykov*, Nauka, Moscow 1989.
Payne, D. "Maasai gender in typological perspective" *Studies in African Linguistics* 27/1998: 159-175.
Puzickij, E. V. *Kačinskij jazyk*, Nauka, Moscow 1968.
Renault, R. "Genre grammatical et typologie linguistique", *BSL* 82/1987: 339-363.
Rounds, C. *Hungarian. An Essential Grammar*, Routledge, London 2001.
Ruhlen, M. *A Guide to the World's Languages. Volume 1: Classification*, Stanford University Press, Stanford 1991.
Sadiqi, F. *Grammaire du berbère*, Harmattan, Paris 1997.
Samarin, W. J. *The Gbeya Language*, UCLA Press, Los Angeles 1966.
Samarin, W. J. *A Grammar of Sango*, Mouton, The Hague 1967.
Schmitt, R. *Grammatik des Klassisch-armenischen,* IBS, Innsbruck, 1981.
Skorik, P. Ja. *Grammatika čukotskogo jazyka*, I, AN SSSR, Moscow 1961.
Smirnova, M. A. *The Hausa Language*, Routledge, London 1982.
Smyth, D. *Thai. An Essential Grammar*, Routledge, London 2002
Sohn, Ho-Min, *The Korean Language*, CUP, Cambridge 1999.
Starinin, V. P. *Èfiopskij jazyk*, Nauka, Moscow 1967.
Steever, S. B. (ed.) *The Dravidian Languages*, Routledge, London 1998.
Szinnyei, J. *Finnisch-ugrische Sprachwissenschaft*, Leipzig 1922.
Tokarskaja, V. P. *Jazyk malinke (mandingo)*, Nauka, Moscow 1964.
Topuria, G. V. "Zur Geschichte der grammatischen Klassen in den ostkaukasischen Sprachen", in: *Studia Indogermanica et Caucasica. Festschrift für K. H. Schmidt zum 65. Geburtstag*, ghg. von R. Bielmeier, De Gruyter, Berlin & New York 1994: 534-534.
Trijandafilidis, M. A., *Mala novogrčka gramatika*, Aristotelov Univerzitet u Tesaloniki, Thessalonike 1995.
Tuite, K. *Svan*, Lincom Europa, Munich 1997.
Unterbeck, B. & Rissanen, M. (eds.) *Gender in Grammar and Cognition*, Mouton de Gruyter, Berlin & New York 2000.
Vinogradov, V. V. (ed.) *Jazyki narodov SSSR*, IV: *Iberijsko-kavkazskie jazyki*, AN SSSR, Moscow 1967.

Vinogradov, V. A. (ed.) *Osnovy afrikanskogo jazykoznanija. Imennye kategorii*, Aspekt press, Moscow 1996.
WALS = *The World Atlas of Language Structures*, ed. by M. Haspelmath *et alii*, OUP, Oxford 2006.
Welmers, W. *A Grammar of Vai*, UCLA Press, Los Angeles 1976.
Werner, H. *Zur Typologie der Jenissej-Sprachen*, Harrassowitz, Wiesbaden 1995.
Xajdakov, S. M. "K osobennostjam funkcionirovanija klassnoj sistemy v dagestanskix jazykax i jazyke fula", *Voprosy jazykoznanija* 6/1977: 90-100.
Yakovleva, V. K. *Yazyk Yoruba*, Nauka, Moscow 1963.
Zograf, G. A. *Morfologičeskij stroj novyx indoarijskix jazykov*, Moscow: Nauka 1976.

ABBREVIATIONS:

Gr. = Greek
Hitt. = Hittite
Lat. = Latin
Lith. = Lithuanian
OCS = Old Church Slavic
OIr. = Old Irish
PIE = Proto-Indo-European
Skr. = Sanskrit
Ved. = Vedic