INTRODUCTION:

THE NOSTRATIC MACROFAMILY

§ 1. The Nostratic macrofamily. This is a hypothetic macrofamily of languages, including Indo-European (**IE**), Hamito-Semitic (**HS**) (= Afroasiatic) (comprising Semitic [S], Egyptian [Eg], Berber [B], Cushitic [C], Omotic [Om] and Chadic [Ch]), Kartvelian (**K**), Uralic (**U**) (= Finno-Ugric [FU], Samoyed [Sm] and Yukaghir [Y]), Altaic (**A**) (= Turkic [T], Mongolic [M], Tungusic [Tg], Korean [Ko], and Japanese [J]), and Dravidian (**D**). The hypothesis is based on a large amount of common roots (more than 2,800) and many common grammatical morphemes, in which regular sound correspondences have been established (cf. IS MS, IS SS, IS I-III, AD LRC, AD SShS, AD LZL, AD PP, AD NGIE, AD NVIE, AD NM). Among the most important resemblances is that of personal pronouns and inflectional person-markers of the 1st and 2nd persons (***mi** for 'I' in IE, U, A and K, ***ṭʻuˀ** (> *ṭi) ~ ***śʻuˀ** (> **śi) for 'thou' in IE, HS, U and M, etc.), that of interrogative pronouns (originally ***Ḳo** for 'who?' and ***mi** for 'what?', surviving entirely or partially in IE, HS, K, U and A), basic lexical words (roots in descenfant languages) such as ***ʔäˀśˀo** 'stay' (> 'be') preserved in IE (*es-), HS, U and K, ***ʔitê** 'to eat' (IE, HS, M), ***baˀʔˀeri** 'to hold, take' (all branches except U), ***ʻwetê** 'water' (all branches except K), ***nimʔ∇** 'name, word' (IE, HS, U, A), as well as words connected with culture of the final paleolythic age (cf. AD NM), such as ***kälû** 'woman of another moiety' > words for 'daughter-in-law', 'sister-in-law' and 'bride' in IE (Latin glōs, Greek γάλως, Slavic *zъl̸ / *zъlъve), S, U, A and D. The original Nostratic phonology (as reconstructed by V. Illich-Svitych and A. Dolgopolsky) had a rich consonant system (see below) and 7 vowels. The grammatical structure was, most probably, analytic with a rigid word order (a sentence-final verb, attributive precedes its head, pronominal subject follows its verb) and with grammatical meanings expressed by word order, postpositions (***nu** for genitive, ***mA** for marked accusative, and others) and grammatical pronouns.

It is very plausible that there are other members of the Nostratic macrofamily: Chukchee-Kamchadal, Eskimo-Aleut, Gilyak, Elamic (possibly connected with Dravidian) and possibly also Etruscan. But the comparativistic and etymological investigation of these languages is still at its very beginning, therefore at the present stage of Nostratic research they have not yet been included in the framework of comparison.

§ 2. Phonology.

§ 2.1. Consonants. According to the extant comparative evidence, proto-Nostratic had a rich consonant system and 7 vowels.

Nostratic consonant chart

	Stops and affricates			Fricatives		Central approximants	Nasals	Lateral sonants	Vibrants
	Voiced	**Voice-less**	**Emph.**	**Voiced**	**Voice-less**				
	b	p	p̣			w	m		
	d	t	ṭ				n	ḷ	
	ʒ	c	c̣	z	s				
	ǯ	č	č̣	ž	š		ṅ = ŋ	ḷ	r
	ʒ́	ć	ć̣	ź	ś	y	ń	ĺ	ŕ
	ʒ̂	ĉ	ĉ̣	ẑ	ŝ				
	g̱	k	ḳ				ŋ		
	g	q	q̣	ɣ	χ				
					ħ (= ḥ)	ʕ			
		ʔ			h				

Symbols in the chart: affricates: ʒ = d͡z, c = t͡s, ǯ = d͡ž, č = t͡š; lateralobstruents: ʒ̂, ĉ, ĉ̣, ẑ, ŝ - lateralized ʒ, c, c̣, z, s; palatalized consonants: ʒ́, ć, ć̣, ź, ś, ń, ĺ, ŕ = palatalized ʒ, c, c̣, z, s, n, l, r; ḷ and ṅ (= ŋ) = cacuminal or retroflex n and l; uvular stops: g (voiced), q (voiceless), q̣ ("emphatic"); uvular fricatives: χ = Spanish j, ɣ = Arabic ġ; epiglottal (pharyngeal) consonant: voiceless ħ (= ḥ = Arabic ح), voiced ʕ (= Arabic ع).

In proto-Nostratic, as it is reconstructed on the basis of exstant data, there are three series of stops and affricates: voiced (*d, *g̱, *ʒ, etc.), voiceless (*t, *k, *c, etc.), and "emphatic" (*ṭ, *ḳ, *c̣, etc.). The exact phonetic realization of the "emphatic" consonants is not yet clear. Illich-Svitych and myself (up to the recent years) interpreted them as glottalized ejectives. But to-day I do not insist on this particular interpretation. In fact, the emphatic stops are represented in K as glottalized, in HS as glottalized or plain voiceless (the distribution being probably due to prosodic factors), in U (in the intervocalic position) as geminated voiceless stops, in A as fortes, in IE (in its traditional interpretation) as voiceless. The common denominator of their K, HS, U and A reflexes is an additional effort (if compared to the reflexes of N plain voiceless stops). One cannot determine the original phonetic realization of this additional effort (glottalization, aspiration, fortis articulation?). I prefer to denote them as "emphatic" and to use the traditional Orientlistic underdot as their symbol.

Recently Starostin proposed to interpret the emphatic stops as voiceless fortes (out *ṭ = his *t‘), see S NSR 306.

In the following table of sound correspondences the symbol "–" denotes zero. The sign ":" symbolizes the lengthening of the preceding vowel, "⊥:" denoted lengthening of the consonant. The sign "." denotes glottalization (emphaization) of an adjacent consonant, "�‿" is its uvularization, "ʿ" is its tensification (transformation of a lax consonant into a tense one [fortis]), "⊥‗" is its devoicing, ` is its retroflexivization, ʸ is its palatalization. The symbol ° denotes here labialization of the adjacent vowel, the sign ¨ denotes its palatalization. Within conditioning formulas, "_Ụ" means "before a labialized vowel", "_Ẹ" means "before a palatal vowel". IE +*(s)- denotes the addition of the initial IE *s mobile (as a reflex of N word-middle palatal elements). The symbol "**" is used for working hypotheses: in cases when we have sufficient factual confirmation for a class of N phonemes only rather than for each individual N phoneme, e.g. in the case of *n and *ń, where a distinction is possible only if the phoneme is represented in Ostyak, so that in daughter languages where there are no *n|ń-words common with Ostyak we cannot find formal froof of representation of N *n and N *ń separately, but only representation of unspecified *n|ń. In such cases we suppose (as a working hypotheses) that both phonemes (in the case described *n and *ń) are reflected in the same way, which is symbolized by "**". The letter "N" symbolizes an unspecified non-labial nasal consonant. IE *Gʰ = *gʰ|gʰʷ|ǵʰ, *K = *k|ḱ|kʷ; M *G = *g|*ġ, *K = *k|q; ⊥_/ means "after a cns.", _⊥/ is to be read "before a cns.". The query ? denotes our doubts (because the reflex in question is represented in very few roots). The cover symbol X for IE means *x, *x̂, or *xʷ (depending on the adjacent N vw.). The cover symbol Ḥ (in IE) means *ħ, *ħ̂, or ħʷ (here also the choice depends on the adjacent N vw.). IE H is a cover symbol for all laryngeals (except for *ʔ).

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*b-	*b	b	*b	*b	*bʰ	*p	*b	*b	*b	*p
*-b-	*b	b	*b, *β	*b	*bʰ	*w, ⊥_/*p	*b	*b	*b	*V
*p-	*p	f	*f	*p	*p, *b	*p	*b, *pʼ	*φ, ?*b	*p	*p

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*-p-	*p	f	*f	*p, ?*b	*p, *b	*p, ?*w	*p, *b, *∅	*b, *β >*ɣ	*p, *b	*√
*p̣-	*p	p	*f	*p, *p̣	*p	*p	*h >*∅	*ɸ	*p	*p
*-p̣-	*p	p	*f	*p, *p̣	*p	*pp	*p, *b, *∅	*b, *β >*ɣ	*p	*pp
*d-	*d	d	*d	*d	*dʰ	*t	*j	*d, _i/ *ʒ́	*d	*t
*-d-	*d	d	*d	*d	*dʰ	*δ	*δ	*d	*d	*ṭ/ṭṭ
*t-	*t	t	*t	*t	*d	*t	*t'	*d, _i/*ʒ́	*d	*t
*-t-	*t	t	*t	*t	*d	*t	*t	*d	*d	*ṭ /ṭṭ, *⊥t
*ṭ-	*ṭ,*t	d	*ḍ	*ṭ	*t	*t	*t'	*t, _i/ *č́	*t	*t
*-ṭ-	*ṭ,*t	d,t	*ḍ,*t	*ṭ	*t	*tt	*t'	*t	*t	*tt /t
*g-	*g	g,ȝ	*g	*g	*gʰ, *ĝʰ, *gʷʰ	*k	*k'	*g,*g	*g	*k
*-g-	*g	g,ȝ	*g	*g	*gʰ, *ĝʰ, *gʷʰ	*ɣ	*g	*g,*g, *ɣ,*ɣ	*g	*:
*k-	*k	k,c	*k,*g?	*k	*g,*ĝ, *gʷ	*k	*k'	*k,*q	*k	*k
*-k-	*k	k,c		*k	*g,*ĝ, *gʷ	*k	*g,*k	*g,*g, *ɣ,*ɣ _i/? *y	*g	*k
*ḳ-	*ḳ,*k	q,k	*ɣ	*ḳ	*k,*k̂, *kʷ	*k	*k̒, *k'	*k,*q	*x	*k
*-ḳ-	*ḳ	q	*ɣ,*k	*ḳ	*k,*k̂, *kʷ	*kk	*k	*k,*q	*k	*kk
*g-	*ɣ	ꜥ?		*ɣ	*x, *xʷ, [*x̂?]	*∅	*∅	*∅	*∅, ?*g	*∅

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*-ʕ-	*ʕ	ꜥ		*ʕ	*X, ?*H̲	*∅, ?*ɣ	*∅	*∅	*∅, ?*g	*∅
*ḥ-	*χ	χ, χ́ (ḥ)	*H	*ḥ	*x, *xʷ, [*x̂?]	*∅	*∅	*∅	*∅	*∅
*-ḥ-	*χ	χ, ?χ́ (ḥ)	?*H	*ḥ	*H	*∅	*∅	*∅, *g, ?*g	*∅, ?*g	*∅
*g̣-	*ḳ, *χ	q, χ	*ɣ	*q	*k, *k̂, *kʷ	*k	*kʿ, *kʼ	*k, *q	*x	*k
*-g̣-	*ḳ	q	?*ɣ	*q	*k, *k̂, *kʷ	*k, *kk	*k	*k, *q	*k	*k, *kk
*ʒ-	*z	ʒ?		*ʒ $=*ʒ_1$	*s	*s	*j	$*ǯ_2$?	*j	$*c_2$
*-ʒ-	*z	ʒ?		*ʒ̀ $*ʒ_1$	*s	?*ć	*s??	$*ǯ_2$?	*j?	*ṭ?
*c-	*s (= *c)	s?		*c̀ $=*c_1$	?*(s)K ?_⊥ /*s	*ć	*ç	$*č̣_2$?	*ç	$*c_2$
*-c-	*s	?c	?*s	*c̀ $=*c_1$	*s	*ć	*ç?	$*č̣_2$?		
*ç̣-	*ç̣	???ʒ		*c̣̀ $=*c̣_1$	*(s)K̂, ??*Kʰ	*ć	*ç	$*č̣_2$	*ç	$*c_2$
*-ç̣-	*ç̣, *s	?ʒ		*c̣̀ $=*c̣_1$	*s	?*ć	*ç	$*č̣_2$??	*ç?	$*c_2$
*s-	*š	s	*s	*s̀ $=*s_1$	*s	*s	*s	*s	*s	$*c_2$
*-s-	*š	s	?*s	*s̀ $=*s_1$	*s	*s	*s	*s	*S	$*c_2$
*z-	*z			*z̀ $=*z_1$	*H	*s	*j	?*s	*s?	$*c_2$
¿*-z-?		z?	?*z	*z̀ $=*z_1$	?*H	*s		?*s, *ḷ		
*ź-	*z	z?		*ʒ	*s	*ć, *ś	*j	$*ǯ_2$?	*j?	
*-ź-	*z			*ʒ, *z	*s	*ć		$*ǯ_2$		$*c_2$
*ć-	*s	s?	*s	*c	*(s)K̂, _⊥ /*s	*ć	*ç		*ç	$*c_2$

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*-ć-	*s			*c	*s	*ć	*ç ?	$*ǯ_2?$		$*c_2$
*ć-	*ṣ	ȝ		*c̣	*(s)K, ??*Kʰ	*ć	*ç	$*ǯ_2$	*ç	$*c_2$
*-ć-	*ṣ, *s	?ȝ	?*s	*c̣	?*s	*ć(ć)	*ç	$*ǯ_2?$		?$*c_2$ (c_2)
*ś-	*š	s	*s	*s	*s	*ś	*s	*s	*s	$*c_2$
*-ś-	*š	s	*s	*s	*s	*ś	?*s	*s	*s	$*c_2$
*ź-	*z			*z	*H, *ʕ	*ś	*j	*s	*s	$*c_2$
*-ź-	*z	z?, s?	*z?	*z	*H	*ś		$*ǯ_2?$		$*c_2$
*ǯ-	*δ	z?		*ǯ, ?*ž	*s	*č, ?*ĺ	*j	$*ǯ_2$	*j	$*c_2$, ?*ṭ
*-ǯ-	*δ	ȝ, d		*ǯ	*s, *d, *sd?	*č, *δ	*δ, ?*y	$*ǯ_2$	*j	
*č-	*θ	s		*č	*(s)t-	*č	*ç ?*ť	$*č_2$??*d	*ç ?*j	$*c_2$
*-č-	*θ	s, ??c	*s	*č	*s	*č	*ç	$*č_2$	*ç?	$*c_2$
*č̣-	*θ̣	ȝ		*č̣	*(s)t	*č	*ç	$*č_2$	*ç	$*c_2$
*-č̣-	*θ̣	ȝ		*č̣	*tʰ, *t *st, ??*sd	*č	*ç	$*č_2$	*ç	$*c_2$
*š-	*š	s	?*s	*š	*s	*š	*s	*s	*s	$*c_2$
*-š-	*š	s	*s	*š	*s	*š	*s	*s	*s	$*c_2$
*ž-	*š	? z		*z	*H	*š	?*j	*s	*j, *s ?	
*-ž-	*š, ?*z	?? ȝ	*z	*ž, *z	*H	*š		?$*ǯ_2$		
*ʒ̂-	*ŝ	? s, ś		*ʒ̂	*l, _Vl/*	*λ	? *j	$*ǯ_2$	*j ?	$*c_2$
*-ʒ̂-	*ŝ ?		*s	*ʒ̂	*l	*ʒ̂	*l		*j ? *ṭ,*ṭṭ	
*ĉ-	*ŝ	ꟗ	?*s	*ĉ	*s	*ć	*ç	$*č_2$	*ç	$*c_2$, ? *Ḱ-
*-ĉ-	*ŝ	ꟗ		?*ĉ	*s	???*c	*ç	$*č_2$	?*ç	$*c_2$

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*ĉ-	*ŝ	ʔ ꜣ	?*ẑ	*ĉ		*ŝ	*ç	*c̬	*ç	*c̬, *ḱ-
*-ĉ-	*ŝ	ꜣ		*ĉ	*s	*ć	*ç	*c̬	*ç	*c̬, ?*ṭ /ṭṭ
*ŝ-	*ŝ	š	*š ?	*ŝ=*s and ?? *s	*s, *ḱs	*ŝ	*s	*s	*s	*c̬, ? *ḱ-
*-ŝ-	*ŝ	š		?*š	*s	*ŝ	*s	*s	*s, ?*l	*c̬
*ẑ-	*ŝ	š	?*z	?*l	*l, ?*s-	*l, *ŝ	*j	*s	*s	*n
*-ẑ-	*ŝ	??n	?*s	*l	*l, _⊥ /*s?	*ẑ	*l, ?*ĺ	*l	*l	*l, *l̥
*ɣ-	*ʕ	ꜥ		*ɣ	*X	*∅	*∅	*∅	*∅	*∅
*-ɣ-	*ʕ	?ꜥ	*H	?*ɣ, ⊥_/*χ	?*X, *∅	*∅, ?*ɣ	*∅	*∅, *g	*∅, *g	*∅, *..
*χ-	*ḥ	ḥ, ẋ (ẖ)	*H	*χ	*X	*∅	*∅ (< *hʔ	*∅	*∅	*∅
*-χ-	*ḥ	ḥ, ? ẋ (ẖ)	?*H‖∅	*χ	*X	*∅	*∅,*..	*∅	*∅, *..	*∅
*ʕ-	*ʕ	ꜥ	?*H‖ʔ	*∅	*H	*∅	*∅	*∅	*∅	*∅
*-ʕ-	*ʕ	ꜥ	?*H‖∅	*∅	*H	*∅,*..	*∅,*..	*∅	*∅, *..	*∅
*ħ-	*ḥ	ḥ, ẋ (ẖ)	*H	*∅	*H	?*∅	*∅	*∅		*∅
*-ħ-	*ḥ	ḥ, ẋ (ẖ)	*H‖∅	*∅	*H	*∅	*∅,*..	*∅		*∅
*h-	*h	?h		*∅	*X	*∅	*∅	*∅	*∅	*∅
*-h-	*h, *∅	h, ẋ (ẖ), ∅	?*∅	*∅	*X	*∅	*∅,*..	*∅	*∅	*∅, *..
*ʔ-	*ʔ	ı̓, ꜣ	*ʔ-, *H	*∅	*ʔ =*∅	*∅	*∅	*∅	*∅	*∅
*-ʔ-	*ʔ	ꜣ,y,∅	*∅,*ʔ	*∅, *⊥	*∅,*..	*ʔ =*∅	*∅,*..	*∅	*∅, *..	*∅

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*m-	*m		*m	*m	*m	*m	*b|m	*m, _#/* b	*m, _# / *b	*m
*-m	*m	m	*m	*m	*m	*m	*m	*m	*m	*m
*n-	*n	n	**n	**n	*n	*n	**j	*n	**n	*n
*-n-	*n	n	**n	*n	*n	*n	**n	*n	*n	*-ṇ-, _t/*n
*ǹ-	*n	n	**n	**n	*n	*ǹ	*j	*n	**n, ?? *ŋ	*n
*-ǹ-	*n	n	*n	**n	*n	*ǹ	*n	*n	**n	*-ǹ-, _t/*n, _#/*ṇ
*ń-	*n	n		*n	*y	*ń	*j(<*ɪ	*n, *ʒ̑	*ń	*n, *ń
*-ń-	*n			*n	*i̯, ?*n	*ń	*ń >*ḷ	?*n	*N	*-ǹ-
*ŋ-	*ʕ,*ɣ *n,*ʕn			*kn, *n, *'	*n,*ǹ, ?*∅-	*∅-,*, ?*∅-	*∅-,*r	*ŋ, ?*∅	?*n, ?*∅-
*-ŋ-	*n,*m	n	*n	*n	*n, *nGʰ	*ŋ, *-ŋ ~ ?*-n	*ŋ	*ŋ, *ŋG, *ŋK, *ɣ,*ɤ	*ŋ	*m(m), *ṇ, ??*-ǹ-
*l-	*l	?i̯	*l		*l	*l	*j	*n	?*l	??*t
*-l-	*l	r ?,3	*l	*l	*l	*l	*l	*l	*l	*l
*ḷ-	*l	?n	*l	*l	*l	*ḷ	*j	*n	*l	*n
*-ḷ-	*l		*l	*l	*l	*ḷ	*l	*l	*l	*ḷ
*ĺ-	*l		? *l	*l	*l	*ĺ	? *j	? *l, ? *n	? *l	*ń, **n
*-ĺ-	*l	r,3	*l	*l	*l	*ĺ	*ĺ	*l	*l	*ḷ
*r-	*r	r	?*r	*r	*r	*r	*j	? *n	*l, *n	*n
*-r-	*r	r,3	*r	*r	*r	*r	*r	*r	*r	*ṛ *r (< *r⊥)
*-ŕ-	*r	r,3	*r	*r	*r	*r	*ŕ; _⊥/ *ɪ	*r	*r	*ṛ

Nostratic	Semitic	Egyptian	Berber	Kartvelian	Indo-European	Uralic	Turkic	Mongolic	Tungusic	Dravidian
*w-	*w	ꭒ	*w	*w	*w (*u̯)	*w	*b,*°	*b	*b, *v, ?*°	_U/ *∅
*-w- /V_V	*w	ꭒ,ꜣ	?*w	*w, *∅	*u̯	*w	*b,*∅	*β,*b	*b	*v
*-w- /⊥_V	*w, *∅	?? ∅, ɥ	??*w	*w, *∅	*u̯,*∅	*w, *∅		*b,*∅	*b, *∅	*∅
*-w- /a,E _⊥	*w, *∅			*w	*u̯	*w,*ᶜ	?*°	*∅, *ᶜ	*∅	*∅, ? *v
*-w- /u_⊥	*:		?*:	*∅, *w	*:,*∅	*∅	*:	*∅		*:
*y-	*y	ꭒ̣	?*y, *	*∅, ?*y	*y,*ej̦	*y	*j	*y	?*y	*y, *∅, *··
*-y- /V_V	*y	y,∅	*y	*∅	*j̦/*i	*y	*y	*y	*y, *∅	*y,*∅
*-y- /⊥_V	*y,*∅	?∅	*∅	*∅	*∅,*j̦ +*(s)·	*y,*⊥!	*⊥ʸ, *∅,*··	*∅,*y	*∅, *∅, *⊥̂ *⊥ʸ, *j̦	
*-y- /V_⊥	*y,*∅	y,∅	?*y	*∅	*j̦,*∅	*y,*⊥!	*∅, *⊥ʸ	*∅,*y	*∅, *∅,*y *j̦	

Remarks:

1. The formula ***-b- > U *w, ⊥_/*p** is to be read: in the word-medial position N ***b** yields U ***-w-**, but after a cns. it is reflected as ***p**.

2. ***-p- > M *b, *β >*ɣ** means that N ***-p-** yields M ***-b-** or early pM ***-β- > pM *-ɣ-**.

3. The formulae ***d- > M *d, _i/ *ʒ̂**, ***t- > M *d, _i/ *ʒ̂** and ***ṭ- > M *t, _i/ *č̂** reflect the pM affricatization ***di > *ʒ̂i** and ***ti > *č̂i**.

4. The formula ***ʒ̂- > IE *l, _Vl/*s-** means: in the presence of ***l** in the IE root the N affricate ***ʒ̂-** yields IE ***s-**, otherwise ***ʒ̂-** yields IE ***l-**.

5. The formula ***-y- > IE *∅,* j̦, +*(s)-** is to be read: N ***-y-** yields IE zero or ***-j̦-** and causes the appearance of ***s-** mobile in the word-initial position.

6. The formulae ***-y- / ⊥_V > U *y,*⊥ʸ** and ***-y- / V_⊥ > U *y,*⊥ʸ** mean: in the positions ⊥_V (after a cns. and before a vw.) and V_⊥ (after a vw, and before a cns.) N ***-y-** remains in U as ***y** or palatalizes

the adjacent U cns. N *ya- yields D *ya(:)- and *e-, N *yä- is represented by D *e- ~ *a-, and N *ye- yields D *ī- and *e-. N *yi- yields M *i-.

7. N *ž̌ followed by a voiceless cns. yields T *ç.

8. N *ž̌ adjacent to a voiceless cns. islikely to yield Tg *s.

9. The controversial IE cns. cluster *ĝʰd̯ is still to be investigated. In one root (*ĝʰd̯em-/*ĝʰd̯om- [{EI} *dʰĝʰom-] 'earth') it may result from metathesis (IE {EI} *dʰĝʰom- > *ĝʰd̯em-/*ĝʰd̯om-). In another case (*ĝʰd̯yes / ? *ĝʰd̯is 'yesterday') **ĝʰd̯- is likely to go back to N *geˈdˈ- (cf. item 6O3a *geˈdˈüŜ∇ 'late, evening').

10. According to AD LZL 364-5, one of the sources of U *ẑ is the N clusters *-lʕ-, *-ļʕ-, and probably *-ļɣ- (cf. entries 131, 871, 1O42, 1O44, 1698b, 2197, 2475, and 2725 [in the latter U *ẑ < N *-ļɣ-]; in the entry 1698b the post-N cluster *-lʕ- goes back to N *-ļiʕ|ɣ-).

11. Sometimes (under some still unknown conditions) a K preconsonantal *-r- goes back to a N lateral cns. (*ĺ or what is denoted as *ĺ|l) (cf. K *c̦q̦r̦ta 'elbow', GZ *p̦r̦țq̦-el- 'flat', G perpl-i 'ashes' and χ∨irtli 'Winterhaar, Winterpelz [der Tiere]', and GZ *țq̦irp̦- 'spleen' - in the entries 355, 1719, 1723, 1866, and 2355).

12. As indicated in the chart, in IE in the word-initial preconsonantal position the N voiceless affricate *ć yields *s. The same may be (but not necessarily is) true about N *c.

13. The difference between N *p- and *p̦- is likely to be reflected in Chadic. In this paper I tentatively suggest that N *p- yields Ch {Stl.} *f- (> Hs f-, Su f-, ∨-, Bl, Wrj, Ron lgs., Tr, Bu, Mtk f-, Gzg f-, ∨-, Lgn f-, p-. Msg, Ms f-, Ke f-, w-, Mu f-), while N *p̦- is represented in Ch {Stl.} *p- (Hs f-, h-, Su, Bl p-, Sha p-, other Ron lgs. f-, Ngz p-, Tr & Bu (mostly) p-, Gzg p-, Lgn ∨-, Ms, Ke, Mu p-), as reconstructed in Stolbova's thesis (Stl. IF 23-34). Further research in the historical phonology of Ch and HS will correct my highly tentative reconstructions in this field.

14. N *ŋ- yields S *n- (rather than *ʕ or *ɣ) in the presence of a laryngeal or *m in the same N word.

15. The sign * ¨ denotes here palaralizing effect (e.g. N *ya- > D *e-).

On the alphabetical order of letters in the present dictionary see below § 9.

§ 2.2. Comment and additional explanations. This chart needs comment and additional explanations. But in this short introduction I cannot afford entering into details. Some of the problems have been

discussed by V. Dybo in his "Editor's Introduction" ("Ot redaktora") of IS I, in IS SS, AD NGIE and Ad NVIE.

§ 2.2.1. I may add here a short remark about the origin of the Dravidian intervocalic ɼ-consonants. According to the extant data, D *-ř- (= *-ṛ- of the traditional notation, retroflex vibrant/fricative) goes back to N *-ŕ- or *-ry-, D *-ṛ- (= *-r̥- or *-ṭ- of the traditional notation, a trill, becoming a stop if geminated) goes back to intervocalic N *-r-, while D *-r- goes back to N consonant clusters with *r.

§ 2.2.2. In HS the N emphatic consonants yield both emphatic and plain voiceless cnss. The distribution seems to be conditioned by prosody. This is suggested by the fact that in grammatical words and pronouns the HS de-emphatization is a rule. Therefore we do not find emphatic cnss. in HS affixes, pronouns and aux. words.

NB: In the following notes I use not only the above mentioned transcription signs, but also cover signs for unspecified N phonemes (such as P̣ for N p|p̣, û for u|ü, etc. For the meaning of such cover signs cf. below the chapter "Transcription".

§ 2.2.3. On the basis of very scarce extant data I have proposed a highly tentative working hypothesis about the origin of the oposition s- ↔ š- in Manchu:

N *ś-, *š- and *s- yield pTg *s- > LMc s-, as well as s- in other Tg languages (but h- in Lm). Cf. N *śilKᐁ 'to let out' > LMc silgi- ~ silki- 'get through (a narrow opeming), crawl (kriechen)', N *śüĺ|ẑü(-kê) ¬ *śilẑi(-kê) 'mucus, slime, saliva' > LMc silengi 'saliva, drivel', N *ś|ŝEHmᐁ 'to swallow' > LMc simi- 'to suck, swallow', N *śämₗᐁⱼgᐁ 'hair, fine hair' > Tg *seńńe(n) 'beard, fin, gill' > LMc seŋele 'cock's comb, gill', N *šubyᐁ 'spike, spear, to pierce' > LMc suyfun 'awl', N *šünigo or *šüŋo 'snow' > LMc su(ŋ)- v. 'become covered with hoar-frost', N *šiŋerᐁ 'mouse' > LMc singeri id., N *šehŕE 'be awake, watch (over), feel, notice' > LMc sere- 'be awake, feel, guess, understand', N *saP̣ʳüʳ-śᐁ ', pointed stake' > ClMc sabsi- 'stitch', N *sonʳᐁqʳüʳ 'sinew, tendon' > Mc suna ~ sūna 'rein of draught-dogs', N *säRʳUʳ 'sinew, fibre' > ClMc siren id., 'bow-string'.

But sometimes we find pTg *š- > LMc š-, as well as probably ç-, č-, and ć- in other Tg languages. It is likely to go back to *si- (that appears

due to contraction of syllables: N *Si|ü|E...∇ > *si̱∇. For instance, N
*sʳûʷwol∇ 'liquid, moisture' > LMc šula 'juice', N *ʳśʹEXar∇ 'bright,
daybreak' > *si̱aʟ;ʲr∇ > LMc šari 'white, clean', N *s̄iʳhʹw∇ 'sun' >
LMc šun 'sun' (÷ Pre-Classical Manchu siɣun), Ul siu(n-), Nn siũ ᵈ
siu(n-), Neg siɣun ~ siwun.

The Manchu reflexes of N *ŝ- are not yet clear. There are cases of
both Mc s- and š-, but the rule of distribution is still unknown.

§ **2.2.4.** Another tentative hypothesis concerns a curious pD
phoneme, reconstructed by G. Starostin as *ḱ- (GS RVP) and later as
*kɥ- (GS 203-7). This is a phoneme that yields *č²- (and its regular
reflexes) in most Dravidian languages (SD, SCD adn CD), but k- in
Kurux, Malto and Brahui. Earlier Emeneau (Em. NDV) tried to explain
this sound correspondence as a special treatment of pD *č²- precding
*u, but later gave up this hypothesis (due to counter-examples). Pfeifer
(Pf. 66) supposed that this is a cluster *kɥ- resulting from reduction of
earlier *keɥ- or *kiɥ-. In the light of external comparison I am inclined
to suppose that pD *ḱ- goes back to N lateral obstruents (see N
*ŝeɕʟaʲr∇ 'reach, approach, enter' and N *ĉ;Aɕ∇ 'perish, die').

§ **2.2.5.** Vladislav Illich-Svitych advanced a hypothesis about the
origin of IE *s- mobile ("movable *s-'). In his opinion it developed only
in N words with an inherited internal palatal element (*ɥ, *ń, *ĺ, and
the like). My interpretation of his hyp. is as follows. The N emphatic
consonants in the word-initial position were pronounced in IE with an
initial preaspiration. Usually this preaspiration left no traces. But in
roots with an inherited internal palatal element the whole root was
palatalized (non-phonemic supersegmental palatalization). The
preaspiration was transformed into a kind of *ḣ that later yielded a
movable *s-.

It means that the presence of a "movable *s- in an IE root suggests
that the N initial cns. was emphatic (glottalized?). This is important for
roots with an initial labial cns., suggesting that this N cns. was *ṗ-.

This apparently strange phenomenon has a typological parallel (that
was unknown to Illich-Svitych). In Salar (as described by Tenishev) the
vowel i before voiceless (preaspirated?) p, t and k (fortes) was
transformed into i + a preconsonantic sibilant (transcribed by Tenishev
in IPA as š, ʃ, ç: [iški // iʃki // içki] (i.e. iški ~ iɰki) 'two' (< iʰti),
[iʃpax] 'silk thread', [iʃt // içt] (i.e. išt ~ iɰt) 'dog', [pišti-//piʃte-

//piçti-] (i.e. pišti ← piɯ̩ti) 'write' (cf. Tn. SJ 77). In his formulation, the phoneme i had an allophone iš (Tn. SJ 8).

§ **2.2.6.** According to a preliminary working hyp. of mine, in the earliest pIE there was a phenomenon that may be interpreted as a word-initial prelaryngealization (preaspiration?). It existed in some roots and brought about a prothetic vowel before sonorants in Greek (sometimes in Arm and other lgs.) and an initial aspiration (h-) in Greek (in some words with etymological *w- and y-). I suppose that this prelaryngealization may go back to a prosodic phenomenon in pN, e.g. partial devoicing of word-initial sonorants or a kind of breathed voice (to use Ladefoged's terminology). In the present dictionary I denote this prosodic phenomenon by an initial *ʿ- in pIE and pN reconstructions. We shall call this *ʿ- "a laryngoid" (like "vocoid"). Its distribution proves that it was a phonologically relevant phenomenon.

It is also possible, that the NaIE *s- mobile before root-initial sonorants belongs here as well: in roots with an inherited internal palatal element and an initial *ʿ- this laryngoid underwent palatalization (*ʿ- -> *ś'-) and became a movable *s- (e.g. NaIE *(s)nēbʰ- 'narrow, thin' < *śnēi̯b∇ < N *ʿn̄∇Xib∇).

§ **2.2.7.** In several N words (U roots) one can observe what may be considered Uralic prosthetic *w- (preceding a N initial rounded vw.) or a kind of "labial fission" (an initial labialized vw. > U *w + non-labialized vw.):

N *goḲ∇ŝ|Ĉ∇(-Ɽ∇) 'є canine' > FU (in ObU only) *°⌈w̌⌉okŝ∇r∇ 'fox'

N *ʕuḲa 'see', 'eye' > FU (in Ugr only) *wok∇ (or *woɤ∇) v. 'see, look'

N *q⌈u⌉r∇ 'pierce, make a hole; hole, pit' > FU (att. in ObU) *°wur∇y∇ > ObU *wūrзy 'river-bed'

N *ʕulʕê 'recent' > U {UEW} *wuẑe 'new'.

I wonder if there is any regularity in this phenomenon. The matter needs investigation.

§ **2.2.8.** The pN consonantism is rich (50 phonemes), but rather economic (19 distinctive features), most phonemes being well integrated (in Martinet's terms, see Mart. EChPh 79-106), and none being isolated. The distinctive features include 10 orders (places of articulation), 3 glottal series (voiced ↔ voiceless ↔ emphatic), and 6 manners of articulation (stop\affricate ↔ fricative ↔ central approximant ↔ nasal ↔ lateral ↔ trill). Cp. Kartvelian: 18 phonemic

features (for 32 consonants), Semitic: 18 features (for 29 consonants), or Finno-Ugric: 17 (or 16) features (for 26 consonants).

§ **2.2.9.** In the overwhelming majority of registered cases N *g̣– yields T *k̄⁻-. But there are three apparent exceptions: [1] **621.** ? ₂ *gü|u˻ʔ˼ī▽ '≈ roe deer, antelope, (?) goat' > NaT *k̔ülmiz 'roe deer'; [2] **638.** *g˹o˺ʔiñ▽ 'beat, strike' > T *k̔ɪŋna- 'beat (so.), cause suffering', [3] **704.** *g˹a˺ž▽ 'to go; way, path' > ? NaT *k̔ay- > Tv xay- v. 'make the round'. In the first two cases it mat be tentatively supposed a contraction *g̣...ʔ > **ḳ- > T *k̔-, and the N etymology of Tv xay- remains semantically doubtful.

§ **2.2.10.** In the IE clusters *s + stops the opposition between the N vd., vl., and emphatic cns. was neutralized, so that all N stops yielded IE vl. consonants: N *S▽d- > IE *st-, N *S▽t- > IE *st-, N *S▽b- > IE *sp-, N *S▽g- > IE *sk- (or *sk̂-, *skʷ-) (cf. items ## 2014, 2020, 2132, 2186).

§ **2.3. Structure of N words.** The words have the structure CV (auxiliary words and pronouns only), CVCV, CVCCV, CV(C)CVCV, and CVCVCCV.

§ **2.4. Vowels.** The original system of vowels, as reconstructed by Illich-Svitych and accepted by the present author, is as follows:

*i	*u *ü
*e	*o
*ä *a	

The original vowels of the first syllable survive in proto-Uralic, partially in proto-Dravidian (where both *a and *ä yield *a) and partially in the Altaic languages (with mutual assimilation of the vowels within a word). The vowels in those languages are stable, i.e. do not undergo alternation (except for quantitative alternation of short and long vowels in Dravidian). In Indo-European, Hamito-Semitic and Kartvelian there is apophony, i.e. a morphologized alternation of vowels (as well as of simple and geminated consonants) that diminishes the importance of vowels for lexical distinction. This apophony is based on phonologization of former allophones (of accentual origin) and subsequent morphologization of the phonemic alternation. Another source of apophony (especially in Hamito-Semitic) is the incorporation of affixes (prefixes → infixes) into word stems, e.g. the prefix *w (of

passive and non-active verbs) turned into Semitic *u ~ *ū as marker of the passive voice within (or before) the stem, the prefix *-an- (< auxiliary verb used in periphrastic constructions of imperfect) turned into the infixes *-n-, *-a- and into gemination of the stem-internal consonant in Semitic, Berber, branches of Cushitic and possibly Chadic. Due to the apophony the vocalic distinction between roots has been partially lost in IE, HS and K, but indirectly preserved in the prevocalic velar and laryngeal consonants. Thus, the N consonants *g̣, *k, *ḳ and *q̇, when followed by N *o, yield IE *gʷʰ, *gʷ and *kʷ; if followed by N *e and *ä, they yield IE palatalized consonants *ĝʰ, *ĝ and *k̂; if followed by *a or a consonant, they yield plain velar *gʰ, *g and *k. But the N vowels *i, *u and *ü have been preserved better - as IE "sonants" (i.e. high vowels and glides) *i/*i̯ and *u /*u̯ (see examples in AD NVIE, AD NGIE, AD NM, IS I-III and in the present dictionary).

The basic representation of the pN vowels of the initial syllable in the descended languages may be tentatively formulated as follows:

pN *a yields U, T, M, Tg and D *a, K *a/zero and *e(/zero/*a), IE *e/o/zero (with preservation of the preceding velar conconants as *gʰ, *g, and *k) and possibly *a;

pN *ä yields U and T *ä, M and Tg *e, D *a (and sometimes *ē), IE *e/*o/zero (with representation of the preceding velar consonants as *ĝʰ, *ĝ, and *k̂), K *a/zero, *e, and (seldom) *i; in open syllables pN *ä sometimes yields D *ē (the exact rules have not yet been sidcovered);

pN *e yields U, T, M, Tg, and D *e, IE *e/o/zero (with representation of the preceding velar consonants as *ĝʰ, *ĝ, and *k̂), K *e/a/zero and *i/*e;

pN *i yields U, M, and D *i, T *i and *ı, Tg *i̯ and *i̯, IE *ei̯/*oi̯/*i (and, with loss of the glide, *e/*o), K *i and *e, as well as *i/*y in the HS languages. Sometimes (under still unknown consitions) N *i yields D *e and *ē. It is not yet clear if in the word-fonal position pN *i yields IE *i/i̯ or disappears. In the latter case the word-final N *æ in our pN reconstructions (based on IE evidence) should be replaced by a less specific *E.

pN *o yields mainly U, T, M, and D *o, Tg *o, *u̯, and *u̯, IE *e/o/zero (with representation of the preceding velar and lr. consonants as *gʷʰ, *gʷ, *kʷ, and *ꭓʷ), and K *w∇/*w/*u, *o, *a, zero (and sometimes *i and *e under still unknown conditions); in several N words pN *o yields D and M *u (probably due to some special developement under still

unknown conditions); in HS it is sometimes preserved as a C rounded vowel, as labialization of preceding velar consonants in some C languages, especially Bj and Ag (resulting in g ͨ, k ͨ, and ḳ ͨ), and (mainly in primary nouns and in biconsonantic verbs) as S *u and *ū; in the prehistory of IE there is as. *o...u|ü > pre-IE *u...u and later reg. development of this new *u (e.g. > *we/*wo) (cf. AD NVIE and AD NGIE); in N words with a front vw. of the second or third syllable N *o of the initial syll. often (but not always) yields T and M *ö.

pN *u yields U, M, and D *u, T *u (and *ɪ), Tg *u̯ and *u (as well as sometimes *o and *u̯e), IE *eu̯/*ou̯/*u (and, with loss of the glide, *e/*o, as well as forming with preceding velar and laryngeal consonants groups *gʰ ͨe/*gʰ ͨo, *g ͨe/*g ͨo, *k ͨe/*k ͨo, *x ͨe/*x ͨo, and a cluster *gw) and *w∇, K *u, *w-diphthongs, and sometimes *o, as well as often S *u and *-ū- (in primary nouns and biconsonantic verbs) and other reflexes of HS *u, as well as labialization of preceding velar consonants in some C languages. The delabialization *u > T *ɪ is a rather frequent phenomenon, but its rules and conditions have not yet been investigated. A similar phenomenon of delabialization (N *u > Tg *i, *i̯) is found in a few Tg roots (N *bʾuʾxŕa 'watercourse, river' > Tg *bi̱ra 'river, brook', N *Ḳum∇ 'sand' > Tg *xi̱mana- v. 'snow', N *ḳu|odʾaʾħ∇ 'pierce' > ?σ Tg *xidar- v. 'hurt by pricking', N *puḷu [or *puḷü?] 'to spring forth' > Tg *bi̱lku- vt. 'moisten, wet' [side by side with Tg *bulku- v. 'wash; spout, jet'], N *Suŕi '≈ squeeze out, filter, strain' > Tg *si̱ri- 'squeeze out [a liquid]'). In T, M, and U there is optional regr. as. (leading to "harmony of vowels"): in N words with a front vw. of the second or third syllable N *u yields T, M, and U *ü (or sometimes [under still unknown conditions] T and M *ö). N *uy yields NTg *ī (items ##332, 592).

pN *ü yields U, T, and M *ü, Tg *u̯ ,*ü ({Ci.} *y ͩ > Ewk, Lm, Ngd, Sln, Orc, Ud i, Ork, Ul, Nn, Mc u) and *u, D *u, IE *eu̯/*ou̯/*u (and, with loss of the glide, *e/*o) and *w∇ (in both vases the preceding velar consonants was palatalized, which brought about palatalized consonants *g̑ʰ, *g̑, and *k̑, and clusters *g̑ʰw, *g̑w, and *k̑w), K *u, *w-diphthongs, zero (and sometimes *i under still unknown conditions), as well as HS *u (> S *-ū-, etc.) and labialization of preceding velar consonants in some C and Ch languages; *ü is sometimes delabialized to *i in T, M, Tg, D, and possibly pre-IE (whence

IE *e̯i̯/*o̯i̯/*i̯), the rules and conditions of this delabialization still need investigating; sometimes (under still unknown conditions) *ü is represented by M *ö.

In non-initial syllables in late PU and pFU the N phonemes *e, *i, and *ü (unless changed to *u by vw. harmony) yielded *e. If the word-final vw. is attested in U only (by U *-e), the final vw. of the N word will be denoted as *ê (a cover sign for *e∥i).

In N and Tg the initial N *wa- yields, beside the expected *ba-, also (under still unknown [prosodic?] conditions) Tg *u- and *o-, M *bu-, bo-, and *o-.

N *we- yields T *ö- (cf. items Nos. 2457, 2489, 255O) and Tg *u (*u̯ and *ʉ) (items Nos. 2489, 2544, 255O), as well as possibly M *ö- (item No. 255O).

N *wi- (unless influenced by regr. as.) yields T *ö- and M *ö- ~ *bö- (cf. items Nos. 2467, 2479, 2539)

The N initial sequence *ya- is likely to yield D *e- and probably (under still unknown conditions) M *i- (item No. 262O). D *i- from N *y∇- (as in items Nos. 2622 and 2646) needs investigating.

Several items (Nos. 173 and 182) suggest that probably N *bo- could yield (under still unknown conditions) IE *bʰeu̯-/*bʰou̯-/*bʰu-.

N *co-, *ćo-, *c̣o-, *č̣o-, *Ḳo- and *Ḳu- + word-internal *y yield IE *sk- rather than *skʷ-, Therefore IE *skʷ- is extremely rare (Pokorny's dictionary mentions only one root: *(s)kʷalo-s, which is also doubtful: "*(s)kʷalo-s oder vielmehr *kʷalo-s 'eine größere Fischart'".

Several N etymons (Nos. 84Oa, 21O4, 2151, 2235a) suggests that N *-EHi-(*-äHi-, -eHi-, and possibly -üHi-) may yield M *ī (and *i̯) and Tg *ī.

According to AD AVD, the Tg ascending diphthongs go back to three sources: [1] a contraction *-aya- > Tg *-i̯a- ~ *ay- ~ *-a-, [2] a reversal *-ay- > Tg *-i̯a-, and [3] a vocal breaking (Vokalbrechung) N *V₁C(C)V₂ > Tg *V₁V₂C(C)V₂. It may be added to this hypothesis that Tg *-i̯a- may go back to N *-æya- as well.

On the alphabetical order of items in the present dictionary see below § 9.

§ 2.4.1. I can propose a tentative hypothesis about the prosodic origin of pT ascending diphthongs (in many vases, but not in all of them). These diphthongs are reconstructed on the ev. of the Chv reflexes of initial cnss.: Chv j- ÷ NaT *k'- & *k'- < pT *Ḵi̯- & *k'i̯-; Chv č-

÷ NaT *t̄- & *t‘- < pT *t̄i̯- & *t‘i̯-; Chv š- ÷ NaT *s- < pT *si̯-. In my hypothesis, the diphthongs go back to pre-T long vowels with a special (broken?) tone (that I denote with the symbol ˇ). These long vowels may result from compensatory lengthening before simplified cns. clusters, e.g. N *ṭüʕ∇ŕ∇ '(part of) leg' ('calf of leg', 'thigh', 'knee'?) > pre-T *t‘ǚːŕ > **t‘Eüŕ > pT *t‘i̯ǖŕ > NaT *t‘ǖz and Chv чӗр čьr 'knee', N *Ḳaīʕa 'throw, leave' > pre-T *ḱ‘ǎːl > **ḱ‘Eal > T *ḱi̯āl- > NaT *ḱāl- and Chv jul-; N *śihaŕu 'dirt' > pre-T *sEaŕ > T *si̯āŕ > NaT *sāz 'swamp' and Chv šur 'swamp, quagmire'.

§ 3. Grammar. The proto-Nostratic language was analytic. Its grammar was based on a rigid word order, auxiliary words and pronouns.

All words belonged to one of the three classes: [1] lexical words, [2] pronouns, [3] auxiliary words. These classes differ in their syntactical functioning. But some pronouns may follow syntactical rules of lexical words, too.

A. The word order may be described by the following rules:

1. The predicate is the final lexical word of a sentence. It may be followed by personal and demonstrative pronouns (*ʔitê mi 'I eat'), but not by other lexical words.

2. Attributive (expressed by a lexical word) precedes its head.

3. Direct object immediately precedes its verb. Other objects precede the verb, too.

4. Pronominal subject follows the predicate.

5. Pronominal attributive ('my', 'this') may follow the noun.

6. Case markers follow the noun.

The only place left for the (non-pronominal) subject (and for adverbial modofiers) is before the verb with its objects.

This word order survives in U, T, M, Tg, Ko, J, D, K, C as word order and in all daughter-languages as the order of morphemes within words. It was preserved in proto-IE (and its most ancient descendants) as the unmarked word-order, but when the IE words became syntactically autonomous (marking their syntactic function by their morphological form [obligatory cases, etc.]), the former rigid word order disappeared, so that the word order began functioning as a means of focalization. In S, B and Eg the old word order was displaced by a new one (originally emphatic, e.g. attributives following their head).

B. There was a very rich system of pronouns, among them:

[1] personal pronouns: *mi 'I' and *ṭˈüˈ ~ *śˈüˈ 'thou' in the direct case, other pronouns in oblique cases (*H₂oy∇ 'by me, my', *k∇ ~ *g∇ 'thee, thy'), as well as pronominals (i.e. lexical words replacing the pronouns, e.g. *ʔˈoˈkE 'self' functioning as a lexical replacement for

*mi 'I', whence IE *eg͡oH ~ *eg͡H-, Semitic *-āku, etc.), pers. pronouns of 1 pl. excl. (*n̄∇) and of 1 pl. incl. (*g⌐U⌐), as well as compound pronouns: *mi ʔa 'we', *mi ʔa 'ye' (with the plurality marker *ʔa),*mi ṭ⌐ü⌐ 'we' incl.;

[2] interrogative pronouns: *Ḳo 'who?', *mi 'what?', *ya 'which?', as well as *w∇ 'who?' and *ŋ⌐U⌐ [1] 'thing', [2] 'what?' (most probably, a phonetic reduction or an ellipsis from *ya ŋ⌐U⌐ or *ʔäy∇ ŋ⌐U⌐ 'which thing?');

[3] deictic particles *ha, *h⌐e⌐, *⌐h⌐i, *⌐h⌐u indicating the degree of proximity to the interlocutors (hic-deixis, iste-deixis, ille-deixis, etc.), demonstrative pronouns: for active (animate and the like) beings\objects and for inanimate objects (*sE 'he, she' for animate [active], *Ḳ⌐ü⌐ id., and *ṭä 'it' for inanimate, as well as *p̣⌐ä⌐ 'ille, another [animate]', *yi 'he', [?] 'that' [anaphoric], and *r∇ [< *ʔ∇⌐r⌐∇?] [theme-focalizing particle]), for distal deixis: *ča 'that', for distal or intermediate deixis: *ćE 'that'; demonstrative pronouns for collectivity (*ʔa⌐h⌐a, *⌐h⌐a, *ʔEḷA, *ḷA, *g̣∇), for collectivity-plurality (*n̄⌐ä⌐, *r∇ yE), for plurality (*t∇ of plurality, *⌐ʔ⌐∇ś∇ 'they' and *yE [= y⌐i⌐?] 'these, they' for animate beings, *ʔa, pc. of plurality, *kU, pc. of plurality [used mostly with pronouns], as well as *ʔ∇q∇ ['thing, things' and prn.\n. of plurality]), for duality (*nE 'they [two]', *⌐h⌐æ ~ *°⌐h⌐U for animate beings, *yi for inanimate objects), for individualization (*⌐y⌐iyo 'which' [relative], 'that which, related to', *r∇ = a theme-focalizing pc.), etc. In the descendant languages these pronouns and particles were transformed into personal endings of the verb (1st and 2nd persons from personal pronouns, 3rd person from demonstratives), into pronominal possessive suffixes, into markers of the nominative case (e.g., IE nominative *-s [for nouns of the active gender] from the N demonstrative active *sE), into affixes of plural, dual and collectivity. In some languages (IE, K, and probably S) the genitive case, too, is based on pronouns (e.g., the genitive case endings: IE *-oi̯os, K *-iš̀ < *⌐y⌐iyo sE 'which is' [× N *ya sE id.], originally 'that which is X', as well as probably the S genitive ending *-i < N *⌐y⌐iyo [and\or ? *ya 'which']).

C. Auxiliary words: [1] postpositions and locative adverbs (in many cases functioning also as preverbs): *nu 'of, from', *mA and *ṭ∇ (particles of marked accusative), *Ḳ∇ (= *ḳ∇?) 'towards' (→ 'to'), *k∇ 'out of, from', *ś∇ 'to, towards', *t⌐ä⌐ 'away (from), from', *ć̣æ

'away, downward', *ʅA (locative pc.), as well as lexical words transformed into postpositions/preverbs: *ʔiň⌐A⌐ 'place' (→ 'in'), *dˌoyˌa 'place (within, below)' (→ locative particle *da 'in'); [2] negations: *ǹi 'not', *ʔe (~ ? *ʔä) 'not', *mä ¬ *mäh⌐o⌐ 'do not' (prohibitive pc.) and 'not' (negative); [3] auxiliary words with meaning of tenses and aspects: *d⌐i⌐, marker of imperfective (< an auxiliary verb?), *H₂i, pc. of past (preterite), *c̣|c̣i or *ĉ̣|ĉi, a marker of verbal frequentativity\iterativity, ; [4] auxiliary words of other meanings: *ʔa 'to become, be', *ʔA, a marker of the male sex (← '[young] man'??), *bA, adjectival pc. forming compound (→ derived) names of quality bearers, nomina posessoris, animal names, *b∇, pc. forming compound (→ derived) nomina abstracta, *c̣a or *c̣a (= *Hic̣|c̣a?), a marker of relative constructions, *č⌐ü⌐ 'that of..., that which', *ko 'whereas, but, also; doch' (adversative-thematic and reminding enclitic conjunction), *Ḳa (a pc. of request, *ʅ∇ (a pc. with diminutive meaning, *í∇ (< *ʔâí∇?) (pc. in deverbal nominal constructions, nominalizing the verbal action), *mA (marker of nominalized syntactic constructions [≈ subordinate sentences], nominalizer [originally a prn.] that formed analytic equivalents of nomina actionis, nomina agentis and other derived nouns), *ň∇ (a marker [pronoun] that formed analytic equivalents of passive participles), ṭi (syntactic pc. that builds analytical nomina actionis), *ṭ∇ (a marker of passive participial constructions), *wa 'also, same', *y∇ (particle of hypocoristic [?] address [vocative]).

§ 4. **Grammatical typology.** As we can see, proto-Nostratic was a highly analytic language. In this point there is a certain disagreement between Illich-Svitych and myself. Illich-Svitych, albeit recognizing the analytical status of many grammatical elements in N, still believed that some grammatical elements were agglutinated affixes: the marker of oblique cases *-n (= my *nu 'of, from'), the formative of marked accusative *-m (= my *mA), the plural marker *-NA (= my *ň⌐ä⌐ of collectiveness and plurality), and several others. This interpretation is hardly acceptable because the N etyma in question still preserve traces of their former analytic status: [1] they preserve some mobility within the sentence (a feature of separate words rather than affixes), [2] several pN particles are still analytic in some descendant languages, [3] N etyma with grammatical and derivational function are sometimes identical with autosemantic words. Thus, the element *nu 'of, from' functions in the daughter-languages not only as a case suffix (genitive

in U, T, M, Tg, formative of the stem of oblique case in the IE heteroclitic nouns, part of the ablative case ending in T, K and in IE adverbs), but also as a preverb of separation/withdrawal in IE (Baltic), as an analytic marker of separation/withdrawal (ablative) in B (functioning in postverbal and other positions). The element *mA is still analytic in Manchu (be, postposition of the direct object, cf. Hrl. 35, 74-5) and Japanese (OJ wɔ > J o). On the analytical status of J o (< N *mA), no (< N *nu) cf. Vrd.JG 278-82. The element *n̄ᵓäᵓ functions not only as a postnominal and postverbal marker of pl. (> pl. suffix of nouns in K, HS and A, ending of 3 pl. of verbs in K, part of the IE ending *-nti ~ *-nt of 3 pl.), but also as the **initial** marker of pl. or abstractness (← 'collectiveness) in U and Eg pronouns: F nuo pl. 'those' ↔ tuo sg. 'that', ne pl. 'those' ↔ se sg. 'that', Eg nꝫ abstract 'this' and 'these (things')' ↔ pꝫ 'this' m. ↔ tꝫ f. The animate plural deictic element (?) *yE 'these, they' functions not only as the postnominal marker of plural (> plural ending in IE, U, A and C), but also as a prenominal and prepronominal plural marker (in B, Bj and OEg). The affix forming causative verbs in HS may both precede the verbal root and follow it (e.g. in deverbal nouns), which points to an original analytic status of the corresponding N etymon. HS *tw- [prefix of reflexivization in derived verbs > B *tw- → t- id., S prefix and infix *(-)t-, etc.] and the AnIE reflexive particle *-ti (> HrLw -ti 'sich', Lw -ti, Lc -ti, reflexive pc., Ht z-, -za id.) are etymologically identical with N *t∇w∇ 'head' (preserved with this neaning in K and Om), which proves the analytic origin of the marker of reflexivization.

In the descendant languages most of these grammatical auxiliary words and some pronouns turned into synthetic affixes (agglutinative in Early U and A, inflectional [fusional] in IE and to a certain extent in HS and K).

One remark about the opposition of tenses in Arabic: the traditional view is that the opposition kataba vs. yaktubu is that of aspects: kataba is perfective, yaktubu is imperfective (cf. Rdr. EVS). But there is another theory that treats this opposition as temporal (Xrak. OKY). Without trying to solve the problem, I preserve the terminology that refers to kataba as perfective (pf.) and to yaktubu as imperfective (ip.).

§ **5. Derivation.** The original status of the N etyma underlying derivational affixes of the daughter-languages is less clear than that of grammatical morphemes. For some of derivatioanl affixes the analytic origin is obvious. Thus, the etymon *mA (that underlies affixes of nomina actionis and nomina agentis in the descendant languages) was a

separate word, which is evidenced by its position: in HS and K it is found both in front of the verb and after it (while in IE, U, D and A its position in the word is always final). The same is true of the etymon *ṭi (> suffixes and prefixes of verb; nomina actionis and other deverbal nouns, infinitives, etc. in the daughter-languages). The adjectival particle *bA forming animal names and other names of quality bearers (IE *eln̥-bʰo-s 'deer', S *θaʕla-b- 'fox', U *ora-pa 'squirrel', Tg *kȫr-be 'mail reindeer', Manchu oŋgo-ba 'forgetful') is interpreted as analytic on the evidence of its phonetic behaviour: the regular reflex of the N intervocal *-b- in U is *-w-, but in the word *orapa 'squirrel' (> F orava) we find *p, which is regular in the word-initial position only. But for many other etyma of this sort we are not yet able to draw conclusions. Of course, we cannot rule out an ancient synthetic origin of some enigmatic "root extensions" ("Wurzeldeterminative", "élargissements des racines") that have lost their former meaning in IE, HS and other lgs. and are represented by an additional consonant at the end of roots or by synonymous roots (N words) that differ by their final (usually third) consonant. These "root extensions" need serious investigation.

In proto-Nostratic there are groups of phonetically and semantically words, e.g. (1) *w∇d̮∇ʲʕ∇ 'walk, go, set out for' and *wedhA 'to cause to go (to drive, to lead)', (2) *w⸢o⸣nṭ∇ 'belly' and *wañ∇ʒ⸢a⸣ 'belly', (3) *ç̟iryâ 'to scratch\chisel, to shape (an object), to fashion' and *ĉ̱âR̮iʲp∇ (~ *-b-?) 'scratch\chisel, hew, cut', (4) *ṭûÍE (or *ṭæwûÍE?) 'extend, stretch, be(come) long' and *ṭaLP∇ (= *ṭaLb∇?) 'vast; room', (5) *čEg∇ 'to prick' (→ 'to butt'), *čika 'to prick, to split', and *čAk̟⸢U⸣ 'to prick (stechen), to gouge', (6) *ĉodh∇ 'break (esp. a body part), strike' and *ĉ∇d̮∇ʲg∇ ¬ *ĉ∇g∇d∇ 'to tear, to split', (7) *ʿwûŕû 'to scrape', *waHr∇ ¬ *warH∇ 'to scratch, to scrape', and *wAŕĉ∇ 'rub, scrape'. The origin of this similarity is not yet known. Three sources of the similarity may be suggested: (1) ancient (pre-Nostratic) derivation, (2) lexical attraction: phonetically similar words influence the meaning of each other, (3) ideophony. The problem is still to be investigated. By the way, similarity of this kind exists in many languages (if not in all of them). Compare, for instance, English scratch, scrape, scrub, or Russian прыскать 'to sprinkle' and брызгать 'to splash, to sprinkle', or Russian скрипеть 'to squeak, to creak' and скрежетать 'to grind'.

§ 6. The place of Hamito-Semitic. In modern long-range comparative linguistics there are two opinions as to the place of Hamito-Semitic

(Afro-Asiatic) among the languages of the world: (1) the traditional view among the long-range-comparativists (H. Pedersen, V. Illich-Svitych, the present author, etc.) is that HS belongs to the Nostratic macrofamily as its branch, (2) recently several scholars have expressed a different opinion: HS is coordinate with N rather than subordinate to it. Joseph Greenberg believes that HS, Kartvelian and Dravidian do not belong to "Eurasiatic" (his term for Nostratic) as its branches but are coordinate with it. Recently Sergei Starostin has also expressed an opinion about the coordinate relationship between HS and "N proper".

   J. Greenberg's opinion is based on comparison of words of different families within a list of arbirarily chosen items. Before receiving the lexical volume of his book *Indo-European and its Closest Relatives* (Stanford, 2002) I had to judge upon Gr.'s theory from a short list of these items and words for the "Eurasiatic" languages that were published by Ruhlen (Ruhlen OLs 16-17). It is a list of 30 lexical items. It is not free from mistakes and very subjective conjectures. The main IE word for 'eat' is not *tap (found in Tokharian only, but registered as the representative of IE in Greenberg-Ruhlen's list), but *ed- (found in almost all branches if IE: Latin edo, Germanic *it-, Sanscrit at-, Hittite it-, etc.), which is related both to Altaic (Mongolian ide 'eat') and to HS *ʔit- 'to eat' (in East Cushitic and West Chadic). The ancient word for 'what?' is not that represented by IE *yo-~*ye-, Uralic *yo-, etc. (which is an ancient N word, but it means 'which'), but *mi, which is represented not only in Uralic, Altaic (Chuvash), but also in HS (all branches), Kartvelian and probably in Dravidian (cf. IS II 66-68). IE *tek- 'to touch' (adduced in the list in the item 'arrive') corresponds exactly to HS *√tḳ (cf. here s.v. N *ṭakalæ 'to touch'). If this list is corrected, enlarged and compared with roots of different branches of HS (as well as Kartvelian and Dravidian), we will see that all these languages are much nearer to "Eurasiatic" than believed by Greenberg and Ruhlen (see Table I):

Table I. "Eurasiatic cognates" (Ruhlen OLs 16-7) and their cognates in Hamito-Semitic, Kartvelian, and Dravidian

Mean	"Eurasiatic cognate"	Ham.-Sem.	Kartv.	Drav.
I	IE *mē-, U *m, etc.	Highland East Cush. *-m	*me	
I	IE *-x 1 sg. marker in verbs	? *ʔ- id.	*χw- id.	
thou	[1] IE *tu-~te, Ur. *t-, etc.,	*t-		
	[2] IE *-s, Turk. *sän		*si-	

Mean	"Eurasiatic cognate"	Ham.-Sem.	Kartv.	Drav.
pronoun (actually 'I','ego')	IE *eĝʰo-m 'ego'	S *-āku 'I' & cognates in B, Eg.		
who?	IE *kʷo-, Ur. *ku, etc.	preserved in Om, Beja, and Ch, but replaced by *m- 'what?' elsewhere		replaced by *y∇ 'which?'
what?	Ur. *mi, Chv. man, etc.	*m-	*min 'who? (< N *mi 'what?')	replaced by *y∇ 'which?'
which?	IE *yo-/*ye-, Ur. *yo-, etc.	? S *ʔayy-		*y∇
this	IE *ḱ-, etc.	Cush. *k-		
that	IE *to-, etc.	*t, fem. & inanim. demonstr.		*-t inanim.
not	IE *ne-, etc.	Eg. n	*nu 'do not!'	
not, do not	Ur. *äla 'do not'	S *ʔal 'do not'		*all- 'not to be'
plural	Ur., Turk. *-t, etc.	S *-āt pl.	*-ta	
two	IE *dwō, etc.	S *tuʔm- 'twin'	*ṭqu-m- 'twin', ṭqu-č̣- 'double'	
eye	IE *okʷ-	Agaw *√ʕḳʷ- 'see', Geez *ʕuk- id.		
see (not 'eye')	Yukaghir nugie 'have seen', etc.			*nik- 'be seen'
bark	Ur. *kopa, Turkic *kāp-, etc.	Cush. *ḳapp, Ch. *√ḳHp	Georgian ḳep̣- 'sheat of paper'	
bark, skin	IE *ker-, FU *kere-, Tung. *xere-	S *√ḳrm	Georg. kerk-	
feather	Ur. *tulka, Turk. **däl∇kʼ- > *jäläkʼ-	Glavda (Ch.) dīākʷà	*bur-ṭq̇]-	
star	IE *Hastēr	*ʕaθtar- 'Venus'		
moon	Korean tal (-l < *-r)	Ch. *√tr		
fish	Ur. *kala, Tung. *xol-sa, etc.	Ch. √klp	Svan ḳalmaχ	*kol(l)-

Mean	"Eurasiatic cognate"	Ham.-Sem.	Kartv.	Drav.
wolf	Ur. *loka 'fox', Mong. *noqa 'dog'		*leḳʷ- 'dog'	*nakka 'jackal, fox'
elder brother	Turkic *āka, etc.	S *ʔaχ- < **ʔaq-		
edge	Ur. *käćä, etc.	S *ḳiçç-	Svan ḳäcχ	
wet	Ur. *ńōre, etc.	?S *√nhr 'river'		*nīr 'water'
dark	[1] Ur. *pol▽, etc.			*pul(l)- 'brown'
	[2] FU *rüm▽	Ch. *rim-	*rum-	
speak	IE *kel-, etc.	Ar qāla 'say', etc.	Sv. gul- 'say'	
sleep	Ur *uni-, etc.	S *ʕūn-		
eat	IE *ed-, Mong.ide-	Cush. *ʔit-, Ron *ʔet-		
arrive	FUr. *tule	S *√dχl, Ch. *ḍ▽l		
take, grasp	IE *kap-, etc.	Cush. *ḳab-		*kap(p)-
wash	Ht. arra-	Arab. √ʔry 'pour'		? *ur-
wash		S *rḥṣ́	*rc₁χ-	

The lexical volume of Greenberg book *Indo-European and its Closest Relatives* (that reached me after the text of this dictionary was already written) did not change anything in my opinion about Gr.'s Eurasiatic theory. Most of his valid comparisons between IE, U and A have exact cognates in HS and/or K and/or D. This can be easily seen from my etymological entries that include references to Gr.'s book (after the sign ◇ or ◎). For instance, he compares IE *kʷasi̯o- 'basket' with U *koća-~*kuća- 'drinking vessel' and OJ kasipa 'container for food\drink'. But reflexes of the same N word are found in Semitic *ˈkaʔas- 'vessel', Berber *kʾūˈss- 'pot, drinking vessel', Cushitic (Xamir kŭskŭsā 'Wasserkrug'), Kartvelian (Georgian ḳvaçia 'small earthern pot') and South Dravidian *kuǀoč-a- (+ suffixes) 'potter' (see my entry #993 [*ḳoʔać▽ 'basket'] and Gr.'s entry #75 of the second volume).

In the first (grammatical) volume of the same book J. Greenberg enumerates the grammatical morphemes that are common to several branches of the Eurasiatic macrofamily. Most of these morphemic parallels are real. But here again we see that the arbitrary exclusion of Hamito-Semitic, Kartvelian and Dravidian is not justified. Almost all "Eurasiatic" morphemes mentioned by Greenberg are shared by Hamito-Semitic and/or Kartvelian and partially by Elamo-Dravidian. For

instance, the "second-person T" (to use Greenberg's notation) is found not only in IE, Uralic, Mongolian and Gilyak, but also in all branches of Hamito-Semitic (e.g. Semitic *ta-), in Kartvelian *tkwen 'ye', 'vester', in proto-Elamic *-ti of the 2nd person and in Drav. *-N-ti, pers. ending of 2s non-past of verbs (see Gr. I 71-4 and the entry *ṭˈüˈ 'thou' of the present dictionary). The "interrogative M" (Grb. I 229-31) is found not only in Uralic, Altaic and some Indo-European languages (Brythonic, Tocharian, Hittite), but also in five branches of Hamito-Semitic (Semitic, Egyptian, Berber, Cushitic, Chadic), in Kartvelian and Dravidian (see here s.v. *mi 'what?'). Greenberg's "Eurasiatic" negation ELE (my *ʔäla) is typical of HS (much more than of IE, where its presence is extremely problematic). To judge by these two last examples, Hamito-Semitic looks more "Eurasiatic" that even Indo-European! Greenberg's book is entitled "Indo-European and its Closest Relatives". To judge from the above examples, Hamito-Semitic is closer to IE than IE itself! Greenberg's book actually proves that in this respect (exluding HS from Nostratic) he is wrong.

Starostin's hypotehsis on HS as a sister-language rather than a daughter-language of N is based on his measurement of shared and replaced vocubulary (of Semitic, IE, Uralic, Turkic, etc.) within Swadesh's list of 100 words (the so-called "basic vocabulary"). Starostin concluded that Semitic (taken as a representative of HS) diverged from N earlier than the "Strictly-N" daughter-families from one another. As it is known, the glottochronological method of measuring linguistic relatioship is based on the unproved assumption that languages replace words of the "basic vocabulary" at a constant rate. But glottochronology cannot serve as a reliable instrument of genetic classification of related languages at least for two reasons: (1) it fails to distinguish between cladistic proximity (German and Swedish are nearer to each other than to Italian and Spanish, because the former go back to Proto-Germanic, while the latter are descendants of Latin, hence German is a "sister-language" of Swedish, but a "cousin-language" of Italian) and dialectal areal proximity (adjacent dialects of a language share innnovations without going back to a special intermediate proto-language, e.g. Czech is nearer to Polish than to Bulgarian, but there was no Proto-West-Slavic, i.e. it cannot be claimed that Polish diverged from Czech **later** than from Russian, Bulgarian or Slovene and that it is **genetically nearer** to Czech than to Bulgarian; on the other hand, Russian is nearer to Polish than to Czech, but there was no Proto-Russian-Polish), (2) it fails to take account of major structural (phonological & morphological) factors encouraging word replacement in some languages (in contrast to other lgs. where these factors do not exist). For instance, in French some phonological factors (loss of many intervocalic consonants and of the posttonic syllables) encouraged

homonymy and replacement of lexical unites (even belonging to the sacro-saint "basic vocabulary" of 100 words): N *ʔ⌐e¬γ∇ 'come, go' is preserved in Proto-IE *ei- 'to go' and in Latin ī- 'to go', but is lost in French, because the phonetic laws in the history of French do not allow this verb to exist: it would have yielded *oi [wa] 'goes' undistinguisable from many other ancient verbs which would have merged in *oi [wa] unless the language had expelled these potential homonyms. The same is true of N *ʔitê 'eat' > IE *ed- > Latin ed-, which would have yielded the same *oi [wa], unless it had been lost in the prehistory of French. Now, let us take just the same N words and see what happened to them in HS and in Semitic: N *ʔ⌐e¬γ∇ 'come, go' yielded HS *ʔiy- 'come' (preserved in Egyptian and Cushitic), but could not survive in Semitic: due to Semitic historical phonology and morphology, 'he went' would have been *ʔā in Arabic and *ʔā in Hebrew (because Semitic verbal roots were devocalized and the intervocalic *-y- was lost); N *ʔitê 'eat' survives in HS (namely in Cushitic and Chadic), but because of the devocalization of verbal roots it was lost in Semitic (otherwise it would have been undistinguishable from other verbs with the same historical consonants, such as *ʔat⌐ʔ¬∇ 'come'). The alleged constant rate of lexical replacement is a hypothesis at variance with the structure of languages. If in Swadesh's list the percentage of words shared by Semitic and IE, Semitic and Uralic, Semitic and Turkic, etc. is indeed lower than that shared by IE and Uralic, IE and Turkic, etc. (as Starostin claims), it may be due to the structural history of Semitic rather than to the date of separation of HS from other daughter-families of Nostratic.

The present author shares the opinion of those who are sceptical about the reliablity of lexicostatistics as a source of chronology. For more details see my paper "Sources of linguistic chronology" (AD SLC) in *Time Depth in Historical Linguistics* (*TDHL* [2OOO]: 4O1-9).

If Proto-"Nostratic proper" (without HS) had ever existed, it would have lead to creation of a specific "Strictly-N" word stock, not found in HS (just as there is a Proto-Germanic word stock that includes roots not found in other branches of IE). But among the 2998 N words registered in this "Nostratic Dictionary" the overwhelming majority (more than 2700) do appear in HS (including cases with a query). The N words found in several daughter-families but not in HS (which could have justified a hypothesis of "N proper") are even fewer than those found in several branches but not in IE, but nobody will exclude IE from N! *Therefore the traditional Nostraticist view considering HS as a branch of N is still valid.*

§ 7. Using etymological dictionaries. The etymological dictionaries of daughter families (such as UEW for Uraluc, DQA for Altaic, OS for Hamito-Semitic, P and WP for Indo-European, D for Dravidian, etc.) have proved to be extremely useful in our research. This does not mean that the present author agrees with all etymological proposals and hypotheses of the quoted colleagues. If I want to indicate that only a part of the proposed comparisons is acceptable, I use the symbol "≈" before the abbreviated name of the source. In quoting DQA, I use the abbreviation "incl." to indicate the acceptable comparisons between sub-branches. For instance, in the entry ***ˉ̗Eˈgˈˌumê** 'wet\cold weather, dew' (→ 'wet snowfall' → 'snow') I quote DQA #1232 in the following way: "≈ DQA #1232 (A *ĺṻńi; incl. Tg, Ko)", which means that I agree with the comparison (found in DQA) between NTg *lūńe 'wet snow' and pKo *nūn 'snow', but not with another comparison in the same entry of DQA, namely that with the erroneously reconstructed M ⁑dün in HlM дүн(г) хүйтэн 'extreme cold' - in fact, literally 'full cold' with dün 'full'. I usually do not quote the untenable or unconvincing comparisons of my colleagues and do not explain the reasons of my doubts, because this is beyond the scope of the present dictionary.

§ 8. The Nostratic symposium. Remarks of my colleagues and methodology. The Nostratic Symposium (Cambridge, the McDonald Institute, July 1998), the discussion and the remarks of my colleagues have been very helpful in improving the quality and the exact formulation of the etymologies in this Nostratic Dictionary. This is true not only about the remarks with which I agree (and which are taken into account), but also about those with which I disagree. They are important because they suggest the necessity of explicit and more precise formulation of the ideas concerning etymologies. One example: in AD NM 28 I state that "milk as food exists only in societies with husbandry". I meant there milk as food for adults rather than mother's milk for babies. I supposed that this is obvious. But now I see (from D. Sinor's reaction) that there may be misunderstanding, so that a more explicit statement is needed. A further example is the use of capital letters to denote unspecified phonemes of a certain class. They are used not in order to conceal conflicting evidence in daughter languages (as one of the colleagues suggested), but first of all to refer to cases when the extant evidence is not enough for identifying a phoneme (see below) or when details of positional representation of phonemes are not yet known. Here also explicit formulation of the usage will help to avoid misunderstanding.

Therefore it will be useful now to dwell on some questions of methodology:

§ **8.1. The purpose of the book "The Nostratic Macrofamily and Linguistic Palaeontology**". The book was not intended to be a proof of the relationship between the Nostratic languages. Alexander Vovin is quite right in stressing that "Dolgopolsky's goal in the book is to reconstruct Nostratic homeland and habitat and not to prove the hypothesis itself". The hypothesis was proved more than 30 years ago by V. Illich-Svitych in his "Essay of Comparison of the Nostratic Languages".

In order to prove genetic relationship, one must compare words of the basic vocabulary and grammatical morphemes. That is what Illich-Svitych did (IS I 3-37). But in a paper concerning linguistic palæontology the basic vocabulary and the grammatical morphemes are of no use. If I find that IE *ed- 'to eat' is cognate with Mongolian ide 'to eat', East Cushitic *ʔit- and Ron Chadic *ʔet 'to eat' and I reconstruct N *ʔitê 'eat', this will add nothing to the study of the life, habitat, homeland and culture of the speakers of proto-Nostratic. The same is true of reconstructing proto-Nostratic pronouns for 'I', 'thou', 'who?', 'this' and the Nostratic markers of genitive and accusative. Even without comparative linguistics one expects that the speakers of that ancient language had concepts for 'to eat', for 'I', 'thou', 'what?', etc., and had syntactic means to build a sentence. In linguistic palæontology we work with words and roots belonging to culture and to geographically bound natural phenomena, which is not a basis for proving genetical connections between languages. Usually what is important for the demonstration of genetic relationship of languages is often irrelevant for linguistic palæontology, and viceversa.

Unfortunately, some of my colleagues ignored the goal of NM and tried to draw conclusions about the validity of the Nostratic theory on the basis of the etymologies quoted in NM. This is like trying to check the existence of the Indo-European linguistic family by analizing the etymologies found in Bn. VIIE (*Le vocabulaire des institutions indo-européennes*), which is a study in IE linguistic palæontology.

§ **8.2. Morphology as a criterion of genetic relationship.** Some of my distinguished colleagues stressed the crucial importance of morphology for the demonstration of genetic kinship of languages. This is an old idea, expressed already by Antoine Meillet. This idea is acceptable if the concept "morphology" includes both synthetical and analytical grammatical morphemes. Actually, the same morpheme may be analytical earlier and synthetical later. One of the essential parts of IE morphology is the personal conjugation of verbs such as Old Indian 1 sg. bharāmi - 2 sg. bharasi - 3 sg. bharati and Greek 1 sg. δίδομι - 2 sg. δίδος - 3 sg. δίδοσι. But already Franz Bopp, one of the founders of IE comparative linguistics, payed attention to the fact that the

marker of 1 sg. *-mi in the IE verbs is etymologically identical with the stem of the 1 sg. pronoun (in the oblique cases: cp. Latin mē, Sanskrit mā, English me). It is obvious that the IE personal endings go back to personal pronouns of the 1st and 2nd person and to a demonstrative pronoun (for 3 sg.). What happened in the prehistory of IE, happened also in some Mongolic languages - but not in the prehistory, but almost before our eyes, in the recent centuries: in proto-Mongolic and in Classical Mongolian there is no synthetic personal conjugation, but in Buryat, Kalmuck, Dagur and Moghol it has been formed from a predicative word + personal pronoun (Buryat ɣerexe-b 'I shall come', Kalmuck ɣoβ-na-β 'I go', Dagur ičim-bē 'ich fahre, werde fahren', Moghol rå-nån-bi 'I come, am coming' with -b, -β, -b̓ē and -bi < proto-Mongolic *bi 'I'; Kalmuck garβ-č, Buryat garba-š 'you [sg.] went out', Moghol irån-či 'you come', Dagur ɣawbei̯-ši 'you will go' with -č, -š, -či and -ši < proto-Mongolic *či 'thou').

But if we define morphology as a system of synthetic morphemes only, it will be wrong to claim that "morphological correspondences provide the key to the reconstruction of any proto-language" (to quote D. Sinor). Shall we exclude Sino-Tibetan and other languages without synthetic morphology from comparative linguistics? Prof. Sinor believes that "a comparative dictionary of Nostratic languages will never bring proof of their genetic relationship, a task that only comparative morphology could accomplish" (D. Sin. NT 8). In the case of Nostratic (an analytic language with grammatical particles and pronouns changing into synthetic morphemes in dauhter-languages) the term "comparative morphology" is valid only if it means analysis of the system of grammatical particles and pronouns with their subsequent transformation into synthetic morphemes. Such comparative morphological analysis was begun by Illich-Svitych, especially in the introductory part of his "Essay of Comparison" (IS I 10-18), although his position as to the status of the grammatical morphemes was different from mine (see above § 4).

§ **8.3. Capital letters.** Prof. Comrie suspects that the capital letters (used in Nostratic reconstructions as signs of unspecified phonemes of certain classes) are a refuge for cases with conflicting evidence provided by different daughter languages. He quotes (with indignation) the Nostratic etymon *ḲERᐁ for leguminous plants (AD NM 54), where all letters are capital! In fact what stands behind the capital letters is lack of specific information indispensable for distinguishing between certain phonemes. The symbol *Ḳ means "*ḳ or *q̇". The distinction between the velar *ḳ and the uvular *q̇ has survived in Kartvelian only and has been lost in all other branches of Nostratic. Hence, if a word is

not attested in Kartvelian, we have to use the capital letter Ḳ (or to write explicitly "*ḳ or *g̣"). In the entry in question the Kartvelian reflex is unknown, therefore we use *Ḳ. The unspecified R means "*r or *ŕ" (and not "all kinds of r-sounds", as Comrie erroneously believes). The distinction between the reflexes of *r or *ŕ has survived in Turkic and Dravidian only. If the word (as *ḲER∇) is not attested in Turkic and Dravidian, we have to use the capital letter *R. The symbol *E is used here instead of *e|ä because both Indo-European and Hamito-Semitic (the only languages where this word is attested) have lost the former phonological distinction between N *e and *ä. Here I admit that it would have been more accurate to symbolize the reconstruction as *Ḳe|äR∇ (in order to rule out *i and *ü). Therefore in the present dictionary I have used a more accurate notation (with the sign æ for *e|ä): N *ḲæR∇ '≈ ∈ pod, fruit of a leguminous plant'. The symbol *∇ (for unspecified vowel) is used here because no information for indentifying the final vowel is available. The use of capital letters is not a refuge but rather a convenient method for distinguishing between the known and the unknown. Of course, the unknown includes also cases in which the extant data do not allow us to identify certain phonemes of the word or of the morpheme.

§ **8.4. Merger of homonyms.** One of my colleagues has indicated cases of overlapping etymologies and has even considered them "a common error in purposes of distinct linguistic relationships" (Campbell IB 11). The distinguished scholar has not payed attention to the extremely typical phenomenon of homonymic merger in the history of languages. Every new speaker of a language reconstitutes the language on the basis of utterances he heard (and read). It is true of any speaker and of any generation of speakers of any language. If a language has inherited (or borrowed, derived) several homonyms and if it is possible to bridge between their meanings (according to the typical patterns of polysemy - like metonymy, metaphore, ellipsis, broadening or narrowing of meanings, etc.), the homonyms will inevitably merge into one word. I shall cite only several examples (from hundreds and thousands found in the history of languages).

In Russian there is a word сало 'lard, tallow, animal fat' and a corresponding adjective сальный 'made of tallow, of animal fat'. In the 19th century Russian borrowed from French the adjective sale 'dirty', that accoding to the laws of Russian morphology turned into сальный (souris sale 'dirty smile' ↦ сальная улыбка). But for any speakers of Russian (including those knowing French, like myself) сальный in both meanings is the same word. If in Russian we

hear сальная улыбка (as of a man looking at a woman with indecent thoughts), we imagine a face stained with dirty fat.

In Georgian there is a word quli 'slave' (an old loan from Turkic qul; -i is a suffix of nominative). In the 19th century Russian borrowed the word кули from English coolie (of Dravidian origin). The word won popularity in Russia (probably due to the translation of the English novel "Coolie" by the Indian writer Mulk Raj Anand, preceded by occasional mentioning of this word in "Fregat Pallada" by Goncharov and in short stories by other Russian authors), and in the famous song "От края до края" ("From border to border", by the poet Lebedev-Kumach) there are words: Поют эту песню и рикши и кули, поёт эту песню китайский солдат 'This song (about Stalin) is sung by rikishas and coolies, this song is sung by a Chinese soldier". From Russian the word penetrated Georgian. But in Georgian it coalesced with quli 'slave'. For speakers of Georgian this is obviously the same word, because the meanings 'slave' and 'coolie' are very near. A formal proof of this coalescence is the uvular consonant q- in quli 'coolie' (rather than the velar ḳ- that usually renders Russian к-).

The Spanish subjunctive sea (of the verb for 'be') goes back both to siat and sedeat (subjunctive forms of the Latin verbs for 'to be' and 'to sit'), while the Spanish infinitive ser 'to be' is from Latin sedēre 'to sit' without homonymic merger.

In IE there is a verb *bʰer- that means both 'carry, take, bring' (> Latin fer-ō, Greek ϕέρ-ω, Old Indian bharā-mi 'I carry', Slavonic ber-ǫ 'I take', Armenian berem 'I carry, bring') and 'give birth to' (Gothic baíran, English bear 'to give birth to', Albanian mberat 'pregnant'). It goes back to two or three different Nostratic words: [1] *baˤʔˤeri 'hold, take' (> Mongolian bari- 'hold'), [2] *berEʔa 'give birth to; child' (> Dravidian *per̠- v. 'beget, bear (a young)'), as well as possibly to [3] *bärʔ∇ 'give' (> Turkic bēr- 'give', proto-Tamil *paric- 'gift'). In IE, due to the apophony, the vocalic distinction between N words with *a, *ä, and *e was lost (see above § 2.3), the laryngeal *ʔ was also lost, so that the two or three Nostratic etyma became homonyms. The semantic distance between 'hold, take' and 'give' was small ('give' can be interpreted as metonymy from 'hold' ⇾ 'bring'), but even 'give birth to' could be understood as metonymy from 'hold, carry', so that the three (ot two) Nostratic words merged into one root. In many Indo-European languages the root preserved the original meanings as polysemic variants (such as Gothic baíran 'carry, bring,

give birth', Old Irish ᛒᚱᛖᛏᚻ 'fait de porter/emporter, fait de porter un enfant').

Dravidian *cɪv∇ŋkɪ 'leopard' (or sim.) goes back to N *ʒ̣ɪw∇mˌ∇ˌɡE 'leopard', but N *ĉ̣ˁɪˀb∇y̆∇ 'hyena' merged with it (because in Dravidian in the word-initial position the voiceless and emphatic affricates coalesced, and so did the intervocalic *-w- and *-b-), and as a result Dravidian *cɪv∇ŋkɪ means both 'leopard' and 'hyena, tiger-wolf'.

Hence overlapping etymologies is not an error but an inevitable result of the merger of homonyms - which is a universal law.

§ 8.5. "Isolated cognates" and the amount of preserved phonological information.

Sergey Starostin's comments on my book (S SNM) are a brilliant contribution to long-range comparative linguistics. In these comments, together with some other papers, he found Sino-Caucasian parallels to Nostratic etyma, which are the first step for establishing a Macro-Eurasian super-family covering both Nostratic and Sino-Caucasian (as well as probably some other families). But I have some methodological reservations as to his approach and results.

One reservation (shared by A. Vovin [Vv. AEN 376-8]) concerns "isolated cognates", i.e. words represented in only one of many (three or more) branches of a family. According to Starostin, "in families like this the probability of a common root being preserved in only one branch is quite small, so that a root present only in Turkic or Japanese has a very little chance to be actually Common Altaic (i.e. going back to proto-Altaic - A.D.)" (S SNM 1). Practical application of this principle (not applied by Starostin himself - e.g. in his book on Altaic and Japanese [S AJ]) will bring about disaster to etymological research. One would have to reject all Gothic reflexes of IE words unless they are found in other Germanic languages, and all Lithuanian reflexes of IE words and roots unless they are represented in Latvian and/or Prussian.

The Nostratic etymon *käluǀü 'woman of the other exogamous moiety (of the same age or younger than ego)' is represented in Semitic *kall-at- 'bride, daughter-in-law' (AD NM 84-87), but is not attested by certain cognates elsewhere in Hamito-Semitic. Shall we dismiss this Semitic cognate or find it unreliable only because it is not known in Omotic or Chadic? Shall we share Starostin's strange opinion that such a root "has a very little chance to be" proto-Hamito-Semitic? Let us not forget that all other branches of HS (except Egyptian) are represented by modern languages only, so that a word which might have existed in proto-Omotic or proto-Libyan-Berber was lost several thousand years ago (just as it has been lost in all modern Indo-European languages outside the Slavic subbranch). By the way, recently possible (but not

certain) cognates of this word have been found in Chadic and East Cushitic (cf. the entry *kälû in the present dictionary).

The Nostratic word *qantֿ∇ 'forehead, front' was reconstructed by Illich-Svitych (IS MS 354, IS SS 336) on the basis of IE, Altaic and Egyptian. The Semitic reflex of the word was not known to Illich-Svitych because the languages preserving it were not yet described in 196O's (when IS MS ad IS SS were written). But according to the laws of Nostratic comparative phonology (discovered by Illich-Svitych) the Semitic reflex has to be *χanֿt-. To-day, due to the late Prof. Johnstone's research, we know that in Jibbali (a Semitic language in Southwestern Oman) there is a word χanֿti 'front, front part of anything' (Jo. J 303). Both the phonetic shape and the meaning of the word correspond exactly to what was predicted by Illich-Svitych. Actually this story resembles Leverrier's prediction of the existence of Neptune long before it was actually discovered, or Saussure's hypothesis of the proto-IE "sonantic coefficients" predicting the laryngeals long before they were discovered in Hittite. Shall we neglect or underestimate *qantֿ∇ (an extremely importanr cognate) and deny its proto-Semitic origin only because it is absent in the Semitic languages outside the Southeastern branch (Jibbali, Mehri and Harsusi)?

The IE word *memso- 'meat' is known to have survived in Gothic mimz 'meat', but not in any other Germanic languages. Shall we deny the proto-Germanic origin of this Gothic word (that in fact goes back to NaIE *mēms- 'meat' and to N *ˀʔ'omśa 'flesh, meat')? Shall we deny the proto-Germanic antiquity of the Gothic verb hlifan 'to steal' (obviously from IE *klep- 'steal, hide') only because it has been lost by all other Germanic languages?

"A root present only in Turkic and Japanese has a very little chance to be actually Common Altaic" (Starostin). By "Common Altaic" Starostin means "proto-Altaic". Is this statement true? When he speaks about Japanese, I can understand it, but for other reasons: the Japanese language has lost very much of the proto-Altaic phonological information, so that the probability of chance coincidence in Japanese is rather high. With Turkic the situation is different: Turkic preserves much of the phonological information of proto-Altaic, so that proto-Turkic *tolu 'hail' is a legitimate cognate of IE *del- 'rain, dew' and probably of FU *tälwä 'winter', in spite of its absence in all other branches of Altaic, and hence it must have existed in proto-Altaic. If a root is preserved in Tungusian (a phonologically conservative branch with *x- going back to N *ḳ- and *q- only) and has extra-Altaic cognates in other Nostratic languages, is has much more than "a very little chance" of being proto-Altaic: Tungusian *xodi- 'to finish, stop' (a

cognate of Dravidian *kōṭ˳/*koṭṭ- 'end, summit, top', IE *kʷe(ː)d-/*kʷo(ː)d- 'sharp point', Semitic *°√ḳtw‖y ~ *ḳutṭ- > Geez ḳʷaṭṭ, ḳʷaṭṭa 'butt end of spear', etc.) is very likely to have existed in proto-Altaic, though we find no traces of this root in the other branches of Altaic.

Of course, at the initial stage of research of a possible genetic connection between some languages we are justifiably recommended to be careful with such "isolates" as the only argument of the common origin of language families. But later, when the genetic connection has been proved beyond reasonable doubt and we know the basic phonological correspondences between the languages in question, we may and must use the isolates (especially if they are rich enough in phonological information) to elucidate etymology of words.

I have already mentioned the preservation of phonological information as an important factor in evaluating attested words as sources of etymology. Words that preserve much phonological information (Spanish tiempo 'time' - with **all** infornation of phonemes of Latin tĕmpus, except for the final -us) are more important than those with little information (as French [tɑ̃] spelled as temps). Words with loss of phonological information may go back to different alternative etymons (as French [tɑ̃] going back to several Latin words: tempus 'time', tantum 'so much', tendit '[he] stretches', etc.) and hence cannot prove much. This linguistic factor is much more important than the mechanical factor of "isolatedness".

Starostin's statistical conclusion based on the principle of "isolatedness" and aimed at determining the taxonomic place of Hamito-Semitic (S SNM 14-15) has no real value, because the principle of "isolatedness" is wrong.

§ **8.6. Etymological doublets.** In very interesting remarks of Alexander Vovin there is one theoretical postulate that cannot be accepted. For Vovin it is methodologically impossible that two different roots of a language go back to the same Nostratic etymon (cf. Vv. AEN 369). In my opinion, the postulate is wrong. Etymological doublets do exist in languages, if a root is found in different phonetic conditions (incl. phonetic influence of adjacent morphemes or different levels of stress), undergoes lexical attraction, analogy, etc. - cf. English off and of, life [laɪ̯f] and live [lɪv], wife [waɪ̯f] and woman ['wu-mən] / pl. women ['wɪ-mɪn], French homme and on, Hebrew 'leḇ 'heart' and lē'ḇāḇ id. - both from *'libab-um.

§ **8.7. External comparative evidence and "teleological reconstruction".** On several occasions A. Vovin mentions "teleological

reconstruction" as an illegitimate procedure (Vv. AEN 378, 382-3). By "teleological reconstruction" he means reconstruction of elements (in an intermediate proto-language) that cannot be proved by direct evidence of the descendant languages, but are suggested by external comparison. An example: in M *qoruβ|ɣu 'film, cataract' I prefer the variant *qoruβu which is in regular correspondence with Tungisian, Kartvelian, HS and IE, though the attested M languages have lost the phonetic distinction between earlier *-β- and *-ɣ-. Another example is *K in proto-Tungusian *ⱼiᵢKⱼ-kte 'berries'. The element *-kta/e is a suffix of nomina collectiva, but there is no direct evidence for the preceding *K. If the Altaic word goes back to N *dik̲∇ 'edible cereals or fruit' (reflected in K *dik̲- and in HS *d∇k̲-), we have to expect in Tungusian *ⱼiᵢKⱼ-kte > *ⱼikte. I cannot share his attitude to external comparative evidence. The procedure labeled by him "teleological reconstruction" is known well in comparative linguistics and is quite legitimate. In the proto-Slavic noun *sъnъ n. 'sleep, dream' there is no *p before *n, but we must suppose its existence in the pre-history of Slavic (and its subsequent loss due to the Slavic law of open syllables) on the external comparative evidence of other Indo-European languages: Greek ὕπνος, Old Indian svapnaḥ, etc. In proto-Italic we reconstruct *pes-ni-s (> Latin pēnis 'tail, penis'), though the preconsonantic *s has not been attested in any Italic language, but its presence (and subsequent loss due to phonetic laws) is suggested by the external comparative evidence of Old Indian pasas-, Greek πέος 'penis'. If a proto-language lost phonemes in certain environments (e.g. in consonant clusters) without leaving traces in descendant languages, we sometimes may suppose their former existence by analyzing other cognate (especially ancient) languages ("sisters" of the proto-language). In reconstructing the history of languages we cannot afford neglecting evidence of any source.

§ **8.8. Trisyllabic etymons.** In IS's reconstruction most lexical etymons (but not pronouns or grammatical morphemes) are dysillabic. Bur even IS recognized the existence of some N trisyllabic words: *K̲awing∇ 'arm-pit' (IS I 344), *p'aliHma 'palm of hand' (IS III 93-5) and probably *purč∇(g∇) ~ *pülč∇(g∇) 'flea' (IS II 99-100). In the present dictionary trisyllabic etyma are numerous. In my opinion, trisyllabic and even quadrisyllabic words (> roots) are not an exception, but one of the existing types of syllabic structure (⇔ DbT NJ 339). Hence I cannot accept the rejection of trisyllabic words as an argument against some of my reconstructions (cf. MichM #13 about *dˈoTˈgiHU 'fish' [= *doTgiʔû in the present dictionary]). I suppose

that contraction of trisyllabic words into dysillabic is a common phenomenon in the later history (daughter-families of Nostratic), which explains the loss of *-ᴦoTᴧ- of that N word (originally in an unstressed syllable?) in HS, IE and A. Compare similar phenomena in the history of many languages, such as the fate of Latin digitus, cubitum, calidus and frigidus in the Romance languages.

§ **9. Alphabetical order of entries.** The alphabetical order of consonants is as follows: *ʔ (incl. *ˀ), *ʕ (incl. *ʕ̣), *b, *c (incl. *ć, *c̄, *c̣, *C), *c̦ (incl. *ć̦, *c̦̄, *c̣̦, C̦), *č, *č̣, *ĉ, *ĉ̣, *g, *ǧ, *ɣ (incl. *Γ), *h, *H (incl. *ʜ, *ʜ₁, *ʜ₂), *ħ, *k (incl. *K, *k̲), *ḳ (incl. *Ḳ), *l (incl. *ḷ, *l̄, *L), *ĺ, *m, *n (incl. *ṅ, *n̄, *n̲, *N), *ń, *ŋ, *p (incl. *ṗ, *P, *P̱), *q (incl. *Q), *q̦, *r (incl. *R), *ŕ, *s (incl. *ś, *s̄, *ṡ, *ṣ, *S), *š, *ŝ, *t (incl. *T), *ṭ, *w, *χ (incl. *X), *y, *z (incl. *ź, *z̄, *ż, *Z), *ž, *ẑ, *з (incl. *з́, *з̄, *з̣, *З), *ǯ, *ĵ.

On details of the alphabetic arrangement of entries see our *List of Nostratic entries and of Indo-European roots*.

§ **10. Nostratic etyma and cross-references.** The reconstructed Nostratic etyma (including in cross-references) are printed in bold script. It refers only to reconstructions either proposed or accepted by the present author in this paper(rather than to those quoted from other scholars).

If in same entry there are several cross-references to the same Nostratic etymon, its meaning is often defined only once. It means that if in a cross-reference a Nostratic etymon is mentioned without semantic definition, it must be understood that it has the same meaning as mentioned earlier (within the same entry). Whenever necessary, such a meaning is denoted by an anaphoric sign '⇑'.

§ **11. A note on reconstructions.** If in a reconstruction of descending proto-languages the name of a family is followed by a name of a branch (e.g. **D**: SD) or a name of a branch is followed by that of a sub-branch (e.g.: "S: CS", "FU [in FP]", etc., e.g., **D**: SD *toṭṭ° 'point, nipple', FU: FP *kōče v. 'crawl, clime, run'), it means that the word is attested in one branch of the family or in one sub-branch of the primary branch only, but it is reconstructed on the pD, pS, and pFU level (using formulas of sound changes for the respective family or primary branch as a whole). But if the reconstructed form is preceded by the name of a (sub)branch only (e.g. S *ˈyad- 'hand'), the

reconstruction is based on rules and formulas of the respective (sub)branch only. For instance, EC *ḳadħ- ~ *ḳudħ- 'thorn' is not reconstructed on a pHS or a pC level, but rather on the East Cushitic one. This is true of all cases except the branches of S and FU: the reconstruction preceded by the abbreviations WS, CS, SS, SES, EthS, FP and FV are reconstructed on the pS or pFU (= pU) level.

§ 12. Was Nostratic a root-isolating or a stem-isolating language?

There remains a question: were all Nostratic words monomorphemic, or some of them consisted of two (or even more) morphemes?

We have no ready answer to that question. Here we can discuss two structural problems.

First, we can see that Nostratic words (except for monosyllabic grammatical particles and some pronouns) are polysyllabic. They consist of two, three and sometimes four syllables. In this respect they do not resemble the known root-isolating languages, which are usually monosyllabic (as Chinese) or mono- and bi-syllabic (as Vietrnamese). This fact seems to suggest that Nostratic is more likely to have been a stem-isolating language (with some bi-morphemic words), either at the final stage of its existence (that we are reconstructing) or more probably at an earlier stage of its history. But we cannot be sure about this, because we have not succeeded (so far) to identify the constituent morphemes of Nostratic words. A very rare case of a presumably derived (or compound?) word is Nostratic *mat▽ˈʔ˺▽R▽ 'rainy season' (item no. 1496), that is likely to be derived from (or compound with?) N *mat▽ˈʔ˺▽ 'moisture'.

Secondly, in many cases we reconstruct Nostratic words with optional root extensions (denoted as "+ext."). These words with extensions may be interpreted as bi-morphemic. But this is not the only possible interpretation. Such words with extensions may have been syntactic combinations of words. What we denote as extensions may actually have been unstressed words (bi- or even tri-syllabic words?) functioning as the second member of word-combinations. These unstressed words have been reduced to one syllable (or even consonants without vowels) in the descendant lgs., which caused widespread homonymy among these "extensions" with obliteration of their original meaning. Such phenomena are well known in many languages, especially in the history of Germanic languages, as well as in Slavic, Hungarian, etc. Unfortunately we have no means of proper reconstruction of these unstressed words and their original meaning. If the extensions were unstressed words, our reconstructions of Nostratic words with extensions do not prove that there were bi-morphemic words in Nostratic.

§ 13. On transcription

H. Fleming wrote in his review of AD NM (*AL* XLI/3: 422): "The presentation of the ... etymologies is not user‑friendly. An incredible blizzard of idiosyncratic symbols buries the basic data. ... One must fight one's way through several pages of explanatory notes for symbols that one forgets soon after... The reader is presumed to be as erudite as the author, and so one is confronted with forms written in Hebrew, Greek, Arabic, Russian, Old Church Slavonic, etc. - but not in IPA".

I am going to justify my use of symbols and scripts. One cannot be equally friendly with all kinds of readers. Both NM and this dictionary are written mainly for those linguists who are interested in *languages* (shall we call them "Sprachforscher?) rather than for "general linguists" who deal with the human language as a whole and not with particular languages and language families. More specifically, I write for historical linguists rather than for those who describe modern languages without reference to their history. It is easier for the Sprachforscher (Orientalists, Slavicists) to recognize an Arabic, Hebrew, Armenian, Slavonic or Russian word written in their usual spelling than in IPA. Besides, the traditional spelling often provides us with etymological information lost in the actual pronunciation of the words. The Arabic verb banā 'he built', if written phonetically, gives us no information of the root-final etymological consonant, which is preserved in traditional spelling (letters b, n and y). But, taking into account the interest of those readers who are not Slavicists or Orientalists, I always accompany every non-Latin-based national spelling (other than Greek and modern Cyrillic) with its transcription or transliteration. As to Greek and modern Cyrillic scripts (for Russian, etc.), any professional philologist is expected to know these two alphabets. If he does not, let him consult the Encyclopedia Britannica on his book-shelf (s.v. "Greek Language" and "Slavic Languages").

Now about IPA. This transcription system is almost never used in comparative and historical linguistics, it is usually absent in etymological and comparative dictionaries of any language families of Europe, Asia and Africa. This is not by chance. IPA has intrinsic drawbacks making its use unpractical and even impossible in reconstruction of the history of language families:

[1] Its basic principle: "one symbol for every phoneme (as far is possible)" - is wrong and practically Europocentric (or, better to say, French-English-Germano-centric). The above pronciple is the only reason to prefer ʃ to the analytical symbol š, which is found in the spelling of Czech, Slovak, Croatian, Slovene, Lithuanian, Latvian, and which is the usual traditional symbol in Semitic, Slavic, Finno-Ugric, Turkic, Mongolian, Caucasian, etc. linguistics. It is often used in Cushitic and Chadic linguistics, including in Fleming's own papers. For the affricate č IPA uses either the digraph tʃ (which is misleading,

because č is one consonant rather a consonant cluster and because in many languages [such as Russian and Polish] there is phonemic opposition č ↔ tš) or the clumsy sign ʧ, instead of the generally understood č (which is used both in practical spelling of many languages and in many kinds of traditional transcription). The principle "one symbol for every phoneme" is counter-productive because it ignores the systemic structure of phonology. In many parts of consonantism the analytic principle ("one symbol for one distinctive feature") is much more practical. Cp. my system of sibilant consonants:

	F r i c a t i v e s		**A f f r i c a t e s**		
	Voiced	Voice-less	Voiced	Voice-less	Glottal-ized
Hissing	z	s	ʒ	c	c̣
Hissing-hushing (like in Kartvelian)	z̀	s̀	ʒ̀	c̀	c̣̀
Palatal	ź	ś	ʒ́	ć	ć̣
Hushing	ž	š	ǯ	č	č̣
Lateral	ẑ	ŝ	ʒ̂	ĉ	ĉ̣

It has only 9 symbols for the whole system, it is easily learned and understood. IPA will have to use 20 or more different symbols: z, s, ʣ, ʦ, ʑ, ɕ, ʥ, ʒ, ʃ, ʤ, ʧ, ɧ, ɬ, etc. Where do we see more "incredible blizzard"? What is more "idiosyncratic"?

IPA is unable to denote many phonemes existing in languages without inventing new symbols. In Twi there is a voicelsss domal infradental infralabialized sibilant, which is denoted in IPA by the symbol ɭ. But how shall one denote the corresponding voiced sibilant (as in Jibbali)? In my system ɭ is denoted as š̰, and its voiced counterpart as ž̰, without necessity of any special explanation of the symbol ž̰. If necessary, the corresponding affricates will be naturally denoted as č̰ and ǯ̰.

[2] IPA may be used only if we know (or claim to know) the exact pronunciation of phonemes in a language. This is possible for modern languages. But what shall we do with ancient languages, with reconstructed proto-languages, where the exact pronunciation is unknown? We do not know if Classical Greek σ was pronounced as s, or š, or an apico-alveolar ṡ (like in New Greek). What shall we do if one language has different dialectal variants? How shall we transcribe the Arabic phoneme ج? In Cairo it is pronounced [g], in Bedouin and Iraqi Arabic [ʤ], in Urban Syro-Palestinian and Maghrebine Arabic [ʒ], in Sudanese Arabic as palatal [ɟ], etc. (to use the IPA transcription). In this

particular case of ǯ I have chosen to use a special super-dialectal transliteration symbol ǧ (and the symbol ǧ in super-dialectal transcription of Arabic). What shall we do with reconstructed words if we cannot be sure about some phonetic feature of the phoneme in question (e.g., we know that IE *s is a voiceless sibilant, but we cannot specify it as [s], [ʃ], [ʂ], [ç] or some other voiceless sibilant)? Historical and comparative linguistics has to cope with three kinds of uncertanties: (1) the phoneme is known, but its exact phonetic realization cannot be or has not been established, (2) there are different realizations of the same phoneme in different dialects of a language, so that we need a super-dialectal transcription (such as exists in traditional spelling of languages), (3) in some words or roots we cannot reconstruct some distinctive feature for a class of phonemes, so that we need symbols for unspecified phonemes (e.g. unspecified voiced sibilant, unspecified laryngeal, etc.).

I have tried to create a system of transcription which copes with all these problems. Since this is a unified transcription for several hundred languages (including those with highly complicated system of sounds), it cannot be very simple. Therefore some users will find it not friendly enough. I am sorry about it, but nothing better can be done.

I have done my best in using basic elements of traditional transcriptions: the Orientalistic Transcription, Finno-Ugric Transcription, traditions of transcription of Altaic, Caucasian, Slavic and African languages, as well as IPA. Yes, I have used IPA in those parts of it which are good - especially in denoting vowels (symbols ɔ, ɛ, ɜ, ɯ, ʌ). *Feci quod potui, faciant meliora potentes.*

§ **13.1. On transliteration and traditional spelling.** Data from written languages that use traditional script (other than Latin) are quoted in transliteration (except for Greek and some languages using modern Cyrillic script). Data from languages with traditional Romanized spelling are quoted as in the sources. If a language has rival spelling systems, I have tried to use that of the most authoritative sources or that of standard dictionaries. For instance, for Anglo-Saxon ("Old English") I have used the spelling of Holthausen's dictionary. In quoting Serbo-Croatian the Cyrillic and Roman national scripts indicate the Serbian vs. Croatian variants of their common language; if both variants are identical, the Roman script is used.

§ **14. On references.** In the present dictionary the references are indicated by abbreviations (explained in Bibliography). I have preferred this system to the popular American system of referring to the used literature by names of scholars and data. I did it because my system spares more space: "P" (for the *Indoeuropäisches etymologisches*

Wörterbuch by Pokorny) is shorter then "Pokorny 1959", "BK" is shorter than "Biberstein-Kazimirsky 1860", "Kln. SAH" is preferable to "Klingenheben 1927-1928". The more so for papers of collective authorship: "KRPS" is shorter than "Karaimsko-russko-pol'skij slovar' 1974" or "Baskakov, Zajączkowski, Szaptał (eds.) 1974". I have used this system also for articles in reviews (though in some very rare cases, when the article is unaccessible to me at the moment of submitting this dictionary, I had to use the commonly used practice of quoting by the author's name and abbreviation of the periodical).

§ 15. On epochs and dialects of languages. One of serous problems in compiling a comparative dictionary is ascribing words to particular periods in the history of some languages and to particular dialects. For instance, J. Vendryes and J. Pokorny differ in periodization of the history of Irish. J. Vendryes's "irlandais ancien" includes both Old Irish *stricto sensu* (his "vieil irlandais") and Middle Irish (cf. Vn. *A*, p. IX), while other scholars (e.g. Pokorny) distinguish between these two stages. Many words included by Vendryes in his "Lexique étymologique de l'irlandais ancien" are labeled by Pokorny as Middle Irish. I have prefered to use Vendryes's (and Thurneysen's) periodization and label both "vieil irlandais" and Middle Irish as OIr (Old Irish). Among words that are usually characterized as Old High German there are those belonging to the Upper German dialects ("oberdeutsch") rather than to High German ("hochdeutsch") *stricto sensu*. I have to follow this practice (in spite of its deficiency) except for cases when the difference between dialects is essential for the etymology, so that I sometimes have to use the abbreviation "OHG U" (i.e. Upper German dialects of OHG = "altoberdeutsch"). Let us hope that these problems will not jeopardize the understanding of the etymology and the history of words and roots. Another difficult case is that the so-called "Chagatay language", a term used by different authors in different senses. When quoting Radloff, I used the label "Chg {Rl.}" wherever Radloff uses the language name "Dsch.", though in fact it is often applied to a later literary lge. of Turks (probably better named as East Turki).

§ 16. On infinitives and "pseudo-infinitives" in our vocabulary entries. The pN etyma with verbal meaning are often written with the infinitive particle 'to'. It does not mean that the etyma in question are infinitives or verbal nouns. It is merely an artificial way of indicating that their lexical meaning is verbal.

But in registering the lexical items of the descending languages the same infinitive particle 'to' has its usual meaning. It is used with infinitives and similar verbal nominals (masdar, verbal noun, etc.) only. With the Nenets verbs it is used for the indefinite gerund (неопределенно-деепричастная форма with the suffix -ś ~ -ź ~ -ć) that has, among its functions, that similar to the infinitive.

"Quasi-infinitives" appear also when the verbal meaning is discussed in general terms, without reference to a particular language (e.g. "valid if the primary meaning is 'to pick'", cf. entry no. 453a), as well as sometimes when we mention pIE (and pWIE) roots and pHS, pS or pB consonantic roots, while the English homonymy prevents us from describing the meaning without 'to' (as in the case of 'to fly', that has to be distinguished from 'fly' ['musca']). Similar quasi-infinitives appear also in quotations of etymological hypotheses of other scholars.

§ **17. On indicating the meaning of words and forms.** If in an etymological entry the meaning of items in a branch or sub-branch is indicated with the etymon (pCh, pB, pT, etc.) only and not indicated with every one of the descending languages, it means that the descending languages have preserved the meaning of the proto-form. Cf., for instance, the entry no. 2141 (*S∇WH$_2\nabla$ 'drink'), in which the meaning of the word in the Chadic languages is indicated for pCh only (Ch $^*\sqrt{}$S$_2$Wh v. 'drink') and not for every lge. of the Ch sub-family, which means that the members of the Ch sub-family have preserved the meaning of pCh $^*\sqrt{}$S$_2$Wh.

CLASSIFICATION OF THE NOSTRATIC LANGUAGES

This is not a comprehensive classification of *all* Nostratic languages. For obvious reasons I have not find it necessary to include many of those modern or young languages which are irrelevant for long-range comparison because their stock of roots and affixes goes back entirely to well known and well described ancient or reconstructed langages. It was not necessary to include here such languages as Afrikaans, Sinhalese, Farsi-Kabuli or to give a comprehensive classification of all modern Indo-Aryan and West Iranian languages. On the other hand, even minor languages in families and sub-families without sufficient ancient linguistic documentaton are relevant for deep etymology and have been used in our etymological research. They are represented in this classification.

Names of primary families of languages (Indo-European, Hamito-Semitic, Uralic, etc.) are printed in bold type italics. Names of secondary families of languages (such as Semitic, Berber, Finno-Ugrian, Anatolian Indo-European) are printed in italics.

I. IE = *Indo-European:*
ppIE = Early proto-Indo-European

I.1. NaIE = *Narrow IE* ("IE proper", subfamily including all IE languages except Hittite-Luwian):

I.1.1. Ary (= IIr) =   Aryan (Indo-Iranian):

I.1.1.1.   MtA = Mitannian Aryan (⟿ hippological words and theonyms in Ht and Hurrite)

I.1.1.2. PAry = Pontic (Tauro-Pontic) Aryan (an branch of Aryan that remained in the region north of the Black Sea and preserved in place names; discovered by Trubachev, but misinterpreted by him as part of InA)

I.1.1.3. InA = Indo-Aryan:

Ass = Assamese

Bhr = Bihari

Bngl = Bengali

Gp = Gipsy; Gp A = Asiatic dls. of Gp; Gp Eu = European dls. of Gp

Hnd = Hindi; Hnd Bhj = Bhojpuri Hindi

Lhn = Lahnda

Mld = Maldivian

Mrt = Marathi

NInA = New Indo-Aryan lgs.

Npl = Nepali

OI = Old Indian: Sk (= OI Sk) = Sanskrit, OI BdSk = Buddhistic Sanskrit, OI BHS = Buddhistic Hybrid Sanskrit, OI ClSk = Classical Sanskrit, OI EpSk = Epic Sanskrit, OI ltSk = Late Sanskrit, OI Vd = Vedic

Ori = Oriya

Pali

pInA = proto-Indo-Aryan (reconstructed from InA lgs.)

Prkr = Prakrit (Middle Indian)

Sin = Sindhi; dl.: Sin J = Jatki dialect

Vd = Vedic (= OI Vd)

I.1.1.3.1. Drd = Dardic lgs.:

Khw = Khowar

Kls = Kalasha

Kshm = Kashmiri

Shina

Shm = Shumashti

I.1.1.4. Irn = Iranian:

Alan

Av = Avestan; variants: Av G = Gatha Avestan, YAv = Young Avestan

Awr = Awroman

Bct = Bactrian

Blc = Balochi (Beluji)

ClNPrs = Classical New Persian

EIrn = East Iranian (branch of the Irn family)

GAv = Gatha Avestan

Ish = Ishkashimi

KhS = Khotan Saka

Krd = Kurdish; dls.: Krd K = Kurmanji (= Northern Kurmanji), Krd Sr = Sorani (Southern Kurmanji)

Luri

Med = Median

MIrn = Middle Iranian (cover name for several Irn lgs.)

MPrs = Middle Persian; MPrs T = Middle Persian of Turfan, MncMPrs = Manichaean Middle Persian

NPrs = New Persian, NPrs B = Bakhtiyar dl. of NPrs

MPrt T = Middle Parthian of Turfan

OPrs = Old Persian

Orm = Ormuri

Oss = Ossetic; dls.: Oss D = Digor dl., Oss I = Iron dl.

Phl = Pehlevi, Pahlavi (Middle Persian)

Pmr = Pamir Iranian (common name of Ish, Shgn, Srk, Wx, Yzg, Oroshori, Bartangi, Rushani, and Khufi)

Prc = Parachi

Psh = Pashto, Afghani

Prs = Persian; MPrs Trf = Middle Persian of Turfan

Prt = Parthian

Psh = Pushtu, Pashto, Afghani

Saka

Sct = Scythian

Sgd = Sogdian; variants: BdhSgd = Sgd of the Buddhist texts, ChrSgd = Sgd of the Christian texts, Sgd M = Sgd of the texts of Mug, MncSgd = Sgd of the Manichaean texts, Sgd OL = Sgd of the Old Letters

Shgn = Shugnani

Srk = Sarikoli, Sariqoli (сарыкольский язык)

Tati

Tjk = Tajik

Xuri (Khuri, Chur) (a WIr dialect)

Xwr = Xwarezmic (Iranian)

Wx = Wakhi (ваханский яз.)

Ydg = Yidga

Ygn =Yaghnobi

Yzg = Yazgulami

ZPhl = Zoroastrian Pehlevi

I.1.1.5. Nrs = Nuristani (Kafir) subgr.:

Ash = Ashkun

Kati

Pra = Prasun

Wgl = Waigali

WIrn = West Iranian (branch of the Irn family)

I.1.2. Gk = Greek; Gk Hl = Hellenistic Greek; dls.: Gk A = Attic, Gk AC = Arcado-Cypriote, Gk Ae = Aeolic, Gk Ar = Arcadian sdl., Gk Arg = sdl. of Argos, Gk B = Bœothian sdl., Gk Cp = Cyprian, Gk Cr = Cretan sdl.,Gk Crc = Corcyrian (Cercyrian) sdl., Gk Crn = Corynthian sdl., Gk D =

Doric, Gk Dl = Delphian sdl., Gk El = Elian sdl. of Gk D (Gk of Elis), Gk Ep = Epic Greek, Gk Epr = Epirotic sdl., Gk Hm = Homeric Gk, Gk I = Ionic, Gk L = Lesbian sdl., Gk Lc = Laconian sdl., Gk Lr = Locrian sdl., Gk Mc = Mycenæan, Gk OA = Old Attic, Gk P = Pamphilian sdl., Gk Ph = Phocæan sdl., Gk R = Rhodian sdl. (sd. of Rhodes), Gk Sr = Syracusan sdl., Gk Th = Thessalian sdl.

MGk = Middle Greek (of the Roman and Byzantine periods)

NGk = New Greek; variants: NGk D = Dhimotiki, NGk K = Katharevusa

I.1.3. Itc = Italic:

I.1.3.1. Latin-Faliscan:

Fls = Faliscan

L = Latin; dls.: L Prn = Praenestian dl., L Ln = Lanuvian, L Sc = Sicilian L; EpL = Epigraphic Latin

ltL = Late Latin (= proto-Romance)

MdL = Medieval Latin

OL = Old Latin (= ArcL, Archaic Latin), dl.: OL Pr = Old Latin of Praenesta

VL = Vulgar Latin; dl.: VL Gl = Gaulish Latin

I.1.3.1.1 Rom = Romance languages:

AfR = Afro-Romance (a lge of L origin, surviving in North Africa up to the 10th-11th c., according to Lewicki LRA)

Ctl = Catalan

Dlm = Dalmatian

Fr = French; dl.: Fr Lr = Lorrainese dl.

Frl. = Friulan (= RhR F)

Gsc = Gascon

It = Italian; dls. and subdialects: Ab = It of Abruzzi, Cl = Calabrian, Lm = Lombardian dls., Mdn = sdl. of Modena, Ml = Milanese, Np = Neapolitan, P = Piemontese, Pv = sdl. of Piverone (Piemonte), Sr = sdl. of Sora, STs = South Toscanian, Tr = Trentine (d. of Trento), V = Venetian

McdRm = Macedo-Rumanian

Occ = Occitanian (Modern Provençal); dl.: Occ Lm = Limousin dial. (incl. Occ Cr = sdl. of the département de Creuse)

OIt = Old (Medieval) Italian; sdls.: OIt Ml = Old Milanese, OIt Pv = Old Pavian (dialetto antico pavese), OIt V = Old (Medieval) Venetian

OFr = Old French

Port = Portuguese; Port Mrn = subd. of Miranda

Prv = Provençal (Classical Provençal)

RhR = Rhaeto-Romance; dls.: RhR F = Friulan, LE = Lower Engadin, RhR Srm = Surmiran, RhR Srs = Sursilvan, RhR Sts = Sutsilvan, RhR TL = South Tirol Ladin, RhR UE = Upper Engadin

Rm = Rumanian

Sp = Spanish; Sp Mrg = Spanish of Maragateria

Srd = Sardinian; dls.: Srd Cm = Campidanian, Srd L = Logudorian

I.1.3.2. OsUm = Osco-Umbrian:

Mrc = Marrucinian

Mrs = Marsian
Osc = Oscan
Pæl = Paelignian
Sbn = Sabine
Um = Umbrian
Vls = Volscan
I.1.4. Clt = Celtic:
Gl = Gaulish
CnC = Continental Celtic (cover name for Gl, CltI, and some other Clt
lgs.)
CltI = Celtiberian
I.1.4.1 Goidelic:
Ir = Irish
MIr = Middle Irish
Mx = Manx
NIr = New Irish
OgIr = Ogam Irish (archaic Irish in Ogamic inscriptions)
OIr = Old Irish (= Vendryes's "irlandais ancien", i.e. including Middle
Irish)
ScGl = Scottish Gaelic
I.1.4.2. Brtt = Brittonic (Brythonic) Celtic:
Br = Breton; dl.: Cr = dl. of the diocese of Cornouaille, L = dl. of Léon,
T = Trégorrois (diocese of Tréguier), V = dl. of Vannes
Crn = Cornish
MBr = Middle Breton
MW = Middle Welsh
OBrth = Old Brythonic
OBr = Old Breton
OCrn = Old Cornish
OW = Old Welsh
W = Welsh
I.1.5. Gmc = Germanic:
ORu = the language of the oldest Runic inscriptions
I.1.5.1. NrGmc (= Scn) = Scandinavian, North Germanic:
Dn = Danish
Far = Faroese
Gtl = Gotlandic (a dl. intermediate between Swedish and Danish)
Ic = Icelandic
NIc = New Icelandic
NNr = New Norwegian (nynorsk)
Nr = Norwegian (BNr = bokmål; NNr = New Nr, i.e. nynorsk; Nr Δ =
Norwegian dls.)
ODn = Old Danish
OGtn = Old Gutnish
ON = Old Norse; ON R = ON of the Runic inscriptions
ONr = Old Norwegian

OScn = Old Scandinavian

OSw = Old Swedish; OSw Ru = Old Swedish of Runic inscriptions

OWN = Old West Norse

pScn = proto-Scandinavian (proto-North-Germanic)

Sw = Swedish

I.1.5.2. East Germanic:

Brgn = Burgundian

Gt = Gothic; dialect: Gt Cr = Crimean Gothic

I.1.5.3. WGmc = West Germanic:

AS = Anglo-Saxon (= Old English); dl.: AS A = Anglian

Dt = Dutch (= Netherlandic, Dutch-Flemish); variants: Dt Fl = Flemish, Dt. N = Dutch of the Netherlands; dls.: Dt G = Dutch dl. of Gelderland, Dt H = Dutch of Holland, Dt Lm = dl. of Limburg.

Frs = Frisian

…HG = … High German (e.g., OHG, MHG, NHG)

LG = Low German (cp. MLG)

Lngb = Langobardian

MDt = Middle Dutch

ME = Middle English

MHG = Middle High German; MHG U = Upper German dialects of MHG

MLG = Middle Low German

MMG = Middle Middle German (Middel German dialects of MHG)

NE = New English, dls.: NE Ork = Orkney English, NE Sc = Scottish English, NE Shetl = Shetland English

NGr = New German (ds.): NGr Al = Alemannic, NGr Als = Alsatian German (elsässisch), NGr B = Bavarian (Bayrisch), NGr EP = dls. of East Prussia, NGr Gtn = dl. of Göttingen; NGr Hs = Hessisch (d. of Hessen), NGr HsN = dl. of Hessen-Nassau, NGr M = Middle German dls., NGr NrF = North Franconian German, NGr Ö = Austrian sdls., NGr OP = NGr of East Prussia, NGr S = NUG = Southern (Upper) German dls. (oberdeutsch), NGr Sb = Swabian German, NGr Sw = Swiss German, NGr Trl = Tirolean German, NGr WF = West Franconian German, NGr Wph = Westphalian German

NHG = New High German

OFrk = Old Franconian (= OHG F)

OHG = Old High German; OHG Al = Alemannic dl., OHG F = Franconian dl. (= OFrk), OHG U = Upper German dl. (oberdeutsch); OHG R = OHG of the Runic inscriptions

OLF = Old Low Franconian (altniederfränkisch)

OMG = Old Middle German (in OsS's terminology)

OSx = Old Saxon

OWGmc = Old West Germanic (preserved in proper names in L sources)

Yid = Yiddish; dl.: Yid C = Central Yiddish (Poland), Yid NB = West Yiddish of the Netherlands and Belgium, Yid N = Northern Yiddish (Lithuania, Belorussia), Yid S = Southern Yiddish (the Ukraine, Rumania)

I.1.6.1. Blt = Baltic:

I.1.6.1.1. EBlt = East Baltic

Cur = Curonian

Lt = Lithuanian; dls.: Lt A = Aukshtaitian (High Lt, Aukshtaitish) (with sbds.: EA = East Aukštaitis, WA = West Aukštaitis; subsubdialect of EA: Lt U = sbd. of Ukmergė), Lt D = Dzuki, Lt P = dts. of former Prussian Lithuania, Lt Z = Zhemaitian (Low Lt, Shamaitish, Samogitian; sdl.: Lt K = sdls. of the area of Klaipeda [former Memel-Gebiet])

Ltv = Latvian; dls.: Ltv Ltg = Latgalian, Ltv H = hochlettisch

I.1.6.2. Pru = Prussian

Ytv = Yatvingian (ятвяжский), incl. the language of the supposedly Yatvigian glossarium described by Zinkievičius (Zink. LJZ).

I.1.6.2. Sl = Slavic:

BChS = Bulgarian Church Slavonic

Blg = Bulgarian

Blr = Belorussian

ChS = Church Slavonic

Cz = Czech; dls.: Cz L = Lakh (lašský), Cz M = Moravian, Cz MS = Moravian-Slovak, Cz SEB = Southeast Bohemian (= Czech-Moravian) dl.

HLs = High Lusatian (High Sorbian)

Kshb = Kashubian

LLs = Low Lusatian (Low Sorbian)

McdS = Macedonian (a Slavic language)

MR = Middle Russian

OCrt = Old Croatian (a dialect of OSCr), OCrt K = Kajkav dls. of OCr

OCS = Old Church Slavonic

OP = Old Polish

OR = Old Russian

OSCr = Old Serbo-Croatian

P = Polish

Plb = Polabian

R = Russian; dls.: R Ar = dls. of the Arkhangelsk province (gubernija), R Dn = dls. of the Don region, R Kl = dls. of the Kaluga province (gubernija), R Ks = Kostroma dl., R Ng = Novgorod dl., R Ol = R of the former Olonets province, R Prm = dialect(s) of the Perm province, R Psk = Pskov dl., R Rz = R of the Ryazan region \ province, R Rzh = Rzhev dl., R S = Southern dls., R Sib = Siberian dls., R Sml = dls. of the Smolensk region, R Tv = dls. of the Tver province, R Vlg = dls. of the Vologda region, R Vt = R of the Vyatka region, R W = Western dls.

RChS = Russian Church Slavonic

SCr = Serbo-Croatian; variants: SCr Ch = Chakav dls., SCr Cr = Croatian, SCr K = Kajkav dls., SCr MN = Montenegro dls., SCr Sr = Serbian

Slk = Slovak; dialect: Slk MS = Moravian-Slovak (moravsko-slovenské nářečí)

Slv = Slovene

Slvnz = Slovinzian (Slowinzisch, słowiński język,†словинский язык)

SrChS = Serbian Church Slavonic

Uk = Ukrainian; dl.: Uk B =Bukovina dl., Uk P = Polesye (Полесье) dls.

I.1.7. Thracian branch:

Al = Albanian; pAl = p0pto-Albanian; Al G = Geg, Al T = Tosk; subdialects: A = Arbanasi Geg (Dalmazia), Ba = Barile Tosk, Be = Berat Tosk, Ç = Çamërian Tosk (Çamërisht), D = Geg of Dushman, Db = Geg of Dibër, Dr = Southern Geg of Durrës, Elb = South Geg of Elbasan, F = Falconara Tosk (Italy), Fr = Tosk of Frashër, Gj = Labërian Tosk of Gjirokastër, Hm = Tosk of Himarë, Kr = Southern Geg of Krujë; Lb = Labërian Tosk (Labërisht), M = Malësian Geg, Mn = Mandres Tosk, Mt = Geg of Mat, Mz = Tosk of Myzeqe, OT = Southern Geg of Old Tiranë, P = Prishtinë Geg (Kosovo), Prm = Tosk of Përmet, SG = Southern Geg, Sf = Sofiko Tosk, Sh = Shkodër Geg, Sl = Salamis Tosk (Greece), SM = San Marzano Tosk (Italy), U = Ukrainian Tosk, V = Vaccarizzo Tosk (Italy), Z = Zadrimë Geg

MAl = Middle Albanian (up to the 17th cent.)

DM = Daco-Moesian (= Dacian)

StAl, StAl T = Modern Standard Al (based mainly on Tosk)

StAl G = Standard Geg Al

Thrc = Thracian

I.1.8. Arm = Armenian (= Old Armenian, Grabar)

ClArm = Classical Old Armenian

eOArm = Early Old Armenian

NArm = New Armenian

NEArm = New East Armenian

NWArm = New West Armenian

I.1.9. Mcd = Macedonian (an ancient IE language)

I.1.10. Ilr = Illyrian (lge. or lgs.)

I.1.11. Msp = Messapic

I.1.12. Pnn = Pannonian

I.1.13. Phr = Phrygian

NPhr = New Phrygian

OPhr = Old Phrygian (= {BajO} старофригийский)

I.1.14. Tc = Tocharian lgs.: Tc A, Tc B

I.1. 15. Vn = Venetic

I.2. AnIE = *Anatolian Indo-European:*

Car = Carian

Ht = Hittite

Ld = Lydian

LycIs = Lycaonic-Isaurian

Pal = Palaic

I.2.1 SAn = South Anatolian Indo-European:
HrLw = Hieroglyphic Luwian (= Hieroglyphic Hittite)
Lc = Lycian (= Lycian A); dl.: Lc M = Milyan (= Lycian B)
Lw = Luwian (= Cuneiform Luwian)
1.2 or 1.1. Pls = "Pelasgian" (Pelastian) of IE origin (pre-Greek IE language[s] of Greece) = Philistine

II. HS = *Hamito-Semitic (Afroasiatic):*
II.1.S = *Semitic:*
II.1.1.WS = West Semitic:
II.1.1.1.CS = Central Semitic:
II.1.1.1.1. NWS = Northwest Semitic (a controversial taxonomic unity)
II.1.1.1.1.1. Cn = Canaanite (Macro-Canaanite):
II.1.1.1.1.1.1. SCn = Canaanite proper (South Cn):
Amn = Ammonite
BHb = Biblical Hebrew; variants: BHb B = BHb with Babylonian vocalization; BHb T = BHb with Tiberian traditional (masoretic) vocalization
Ed = Edomite
Hb = Hebrew
ltHb = Late Hebrew (second half of the 1st mill. A.D.); ltHb B = Babylonian ltHb, ltHb J = Jerusalemite ("Palestinian") ltHb, ltHb T = Tiberian (Northern) ltHb
M'b = Moabite
MdHb = Medieval Hebrew
NHb = New Hebrew (19th - 20th c.)
OHb = Old Hebrew (the language of the 2nd and the 1st mill. BCE, undelying BHb and EpHb)
OCn (= OSCn) = Old South Canaanite; OCn Sn = Cn of the Old Sinaitic inscriptions, OCn TA = Cn of the Tell-el-Amarna leters
PBHb = Post-Biblical Hebrew (e.g. Ben-Sirah, apocryphical literature of the 2nd and the 1st cent. BCE)
Ph = Phoenician; dls.: Ph By = Byblian, Ph OBy = Old Byblian, OPh = Old Phoenician
Pun = Punic
SmH = Samaritan Hebrew
II.1.1.1.1.1.2. Ug = Ugaritic
II.1.1.1.1.1.3. Amr = Amorite
II.1.1.1.1.2. McAram = Macro-Aramaic:
II.1.1.1.1.2.1. Aram = Aramaic:
BA = Biblical Aramaic
BzJPA = Jewish Palestinian Aramaic of the Byzantine period
ChrPA = Christian Palestinian Aramaic ("Syro-Palestinian")
DSA = Aramaic in Demotic script
Htr = Hatra (an ancient Aramaic dialect)

IA = Imperial Aramaic, Official Aramaic; IA F = Aramaic words in Frahang-i-Pahlavik (glossaries of Aramaic heterograms in Phl)

JA = Jewish Aramaic (common name for JEA and JPA)

JEA = Jewish East Aramaic (Babylonian Aramaic)

JPA = Jewish Palestinian Aramaic (Kutscher's "Galilean Aramaic"); JPA B = JPA of the Byzantine period

Md = Mandaic (incl. ClMd [= Classical Mandaic] and NMd [New Mandaic, Modern spoken Mandaic])

MNA = Mlahso Neo-Aramaic

Nbt = Nabataean

NNEA = Norteastern Neo-Aramaic ("Modern Assyrian"); dls.: H = Hertevin, JIA = Jewish NNEA of Iranian Azerbaijan, JZ = Jewish NENAr of Zakho, U = Urmiya dial.

NSr = Neo-Syriac

OA = Old Aramaic

PA = Palestinian Aramaic (incl. JPA, JPA B, ChrPA)

Plm = Palmyrene

SmA = Samaritan Aramaic

Sr = Syriac

TA = Turoyo Neo-Aramaic; dls.: TA M = TAr of Mîdin, TA Mt = TAr of Midyat

Ww = the dialect of the "Waw" inscription (belonging to Aramaic?)

II.1.1.1.1.2.2. DA = the language of the Deir-Alla inscription

II.1.1.1.1.2.3. Yd = Ya'udic, Samalian

II.1.1.1.2. Macro-Arabic:

Ar = Arabic; dls.: Ar AT = Arabic of Algeria and Tunisia, Ar CA = Central African dl., Ar CB = dl. of the coastal part of Batina (Northern Oman), Ar ChCS = Arabic dl. of Chad and Central Sudan, Ar ChrNG = Christian fallah dl. of northern Galilea, Ar Cr = Cairo Arabic, Ar D = Dathina dl., Ar Df = Dofar (Zfar) dl., Ar Eg = Egyptian dl., Ar G = Gulf Arabic (the United Arab Emirates, Qatar, and Bahrain), Ar Hdr = Hadramauti dl., Ar Hm = Hamata Arabic, Ar IB = dl. of the inland part of Batina, , Ar Ir = Iraqi Arabic, Ar Lb = Arabic of Libya, Ar Mgr = Maghrib Arabic (North Africa), Ar Mrc = Moroccan dls., Ar. Ng = Ar. of Nigeria, Ar NY = North Yemenite dls. (including   sdls.: Ar NY K = k-subdialect, Ar NY SE - Southeastern sdl., Ar NY S = Southern sdls., Ar NY T = Tihamah sdl.), Ar O = Oman Arabic, Ar OY = Old Yemenite Arabic of 10-11 c. A.D. ([in al-Hamdânî's & Nashwân's works], incl. Himyarite loans), Ar P = Palestinian Arabic dls., Ar SA = South Arabian dls. (Yemen, southern Oman), Ar Sd = Sudanese dls., Ar SL = Syro-Libanese dl., Ar Sp = Arabic of Spain (8th through 15th c.), Ar Y = Yemenite dls. of Ar, Ar Zhl = Arabic of Zahle (Lebanon); dialect groups: Ar B = Bedouin Arabic, Ar F = Fallah (rural) Arabic, Ar Ur = Urban Arabic; PsClAr = Post-Classical Literary Arabic

Lh = Lihyanic

Malt = Maltese

OAr = Old Arabic; dialect areas: OWAr = OAr H = Old West Arabic (Hijaz dls.), OEAr = Old East Arabic; dls.: A = Asad, 'A = Al-'Aliyah, An = Ansar, Hd = Hudhail, Hm = Himyarite OAr, Hr = dl. of Hauran, Hz = Hawazin, Md = dl. of Medinah, Mk = dl. of Mekka, Nj = Nejd dls., O = 'Oman, Q = Qais, Qd = Quda'ah, Qr = Quraysh, R = Rabi'ah, Sl = Sulaim, T = Tamim, Tj = Tajji', TR = Taim ar-Ribab, Y = Yemenite

ONA = Old North Arabian, Frühnordarabisch (Lh, Sf, Tmd, Hs')

Sf = Safaitic

Tmd = Thamudic

II.1.1.2. SS = South Semitic:

II.1.1.2.1. SWS = Southwestern Semitic:

II.1.1.2.1.1. ESA = Epigraphic South Arabian (= OSA)

OSA = Old South Arabian (= ESA); ds. (labeled as separate languages): Hdr = Hadrami, Mn = Minaean, Qtb = Qatabanian, Sb = Sabaic, Sabaean.

Hdr - Hadrami

Hmr = Himyarite (Ḥimyarite [up to the X-XI c. CE])

Mn = Minaean, Minaic (מעינית), Madhabian

Qtb - Qatabanian

Sb = Sabaean, Sabaic (dialect of OSA)

II.1.1.2.1.1 or II.1.1.1.2. OYmn = Old Yemenite, common name for words of Himyarite and of the Old West Arabic dialect of Yemen (10th and early 11th cent.), the distinction between them being unfeasible today

II.1.1.2.1.2. EthS = Ethiosemitic:

Eth = Ethiopian (= EthS)

Amh = Amharic

Arg = Argobba

Gft = Gafat

Grg = Gurage; Gurage languages: Grg Ch = Chaha, Grg Ez = Ezha, Grg Ed = Endegeñ, Grg En = Ennemor, Grg Go = Gogot, Grg Gt = Gyeto, Grg Mh = Muher, Grg Ms = Masqan, Grg Sl = Selti, Grg So = Soddo, Grg Wl = Wolane, Grg Z = Zway

Gz = Ge'ez

Har = Harari

OEth = Old Ethiopian (a spoken lge., which was the basis of Ge'ez)

Tgr = Tigre

Tgy = Tigray, Tigrinya; dl.: Tgy H = Hamasien Tigray

II.1.1.2.2. SES = Southeast Semitic (= the ancestor of the Modern South Arabian languages: Mh, Hrs, Jb, Hbt, Bth, Sq):

SEA = Southeast Arabian (common name for Modern South Arabian: Mh, Hrs, Jb, Hbt, Bth, Sq)

Bth = Bathari (Baṭḥari)

Hbt = Hobyot

Hrs = Harsusi

Jb = Jibbali, Gəblēt, Śaḥri, Šẖawri, dls.: C = Central, E = Eastern, EM = Mehrizing subd. of Jb E

Mh = Mehri (Mahri); dls.: Mh D = Mh of Dhofar, Mh J = Eastern Mh of Jadib and Hawf, Mh Ng = Nagd Mehri, Mh Q = Qishn Mehri, Mh QB = dl. of Qishn bedouins, Mh SW = Southwestern Mehri

Sq = Soqotri; dls.: Sq M = Mountain (Central) dl., Sq N = Northern dl. (incl. Sq HS = Ḥadiboh-Suq), Sq S = Southern dl.

II.1.2. ES = East Semitic:

Ak = Akkadian; dls. : Ak A = Assyrian, Ak B = Babylonian, Ak LB = Late Babylonian (spB}, Ak MA = Middle Assyrian (nA), Ak MB = Middle Babylonian (nB}, Ak NA = Neo-Assyrian (nA), Ak NB = Neo-Babylonian (nB), Ak OA = Old Assyrian (aA), Ak OB = Old Babylonian (aB}, Ak StB = Standard Babylonian (used in Asssyria of the Neo-Assyrian period), Ak YB = Young Babylonian (jB), OAk = Old Akkadian (aAK), Ak Bg = Ak of the Boghazköy texts, Ak M = Ak of Mari, Ak Nz = Ak of Nuzi, Ak RS = Ak of Ras-Shamra tablets.

Ebl = Eblaic, Eblaite

II.2. LbB = *Libyco-Berber* (Old Libyan + Berber + Guanche) (= ливийско-гуанчские языки):

II.2.1. B = Berber:

II.2.1.1. NrB = North Berber:

ASgr = Ayt-Seghrushen (a dl. of Tmz, treated here as a separate lge.)

Assh = Ashasha (a B dialect)

Awj = Awjila

BHlm = Ben-H_alima (dialect of Ulad-Ben-H_alima, a B "Zenatian" dialect of Central Morocco)

BMn = Beni-Menacer

BMs = Beni-Messaud

BSlh = Beni-Salah

BSn = Beni-Snus

Btw = Bettiwa

CA = Berber of Central Algeria

CM = Berber of Central Morocco

Dbl = Dyebayli (a dl. of Nfs?)

Dmn = Demnat

Fgg = Figuig, Figig

Gd = Ghadamsi

Grr = Gurara (Berber dls. of the "ksurs" [villages near Timinun and Badrian])

Hlm = Halima (a B dialect)

Hrw = Harawa (a B dialect)

Iz = Izayan (a B dialect)

Izd = Ayt-Izdeg (a dialect of Tmz, treated here as a separate lge.)

Izn = Beni-Iznacen

Jrb = Berber of Jerba

Kb = Kabyle; dls.: Kb AX = Ayt-Khalfun, Kb AZ = Ayt-Ziyan, Kb Ir = Irjen, Kb GK = dls. of Grande Kabylie, Kb M = At Mangellat (= Kb {Dl.}), Kb PK = dls. of Petite Kabylie, Kb Z = Zwawa

Mtm = Matmata (maṭmāṭa)

Mz = Mzab, Mozabite

NdA = Ndir-Abes

Nfs = Nefusi

Ntf = Ntifa (a B dialect)

Rf = Rif Berber dls.; dls.: Rf A = Beni-Amret, Rf B = Boqqoya (Iboqqoyen), Rf Bt = Bettiwa, Rf K = Kebdana (Ikhbdhanen), Rf Q = Gela'ia (Iqr'ien), Rf S' = Beni-S'id (Aith-Sghidh), Rf T = Beni-Tuzin (Aith-Thuzin), Rf Tf = Beni-Itteft (Aith-Itteftth), Rf Tm = Beni-Temsanan (Aith-Themsanan), Rf U = Beni-Uriaghel (Aith-Uriaghen), Rf Wr = Rif Beni-Waryaghel

Shl = Tashelhit, Shl T = Tashelhit of Tazerwalt (Tashelhit of Semlal [Destaing's "Tachelhit du Sous"] are treated here as a separate lge., see below Sll)

Shnw = Shenua, Shenwa

Shw = Shawiya

Si = Siwa

Skn = Sokna

Sll (= Shl Sm) = Tashelhit of Semlal (Destaing's "Tachelhit du Sous")

Snd = Sened (= Zenatia de Qalaât es-Sened, a B dialect)

SrSn = Srair Senhazha (Senhaja de Sraïr); dl.: SrSn Gz = Taghzut, SrSn AA = Aït-Ahmad

Tgn = Tuggana

Tmm = Timimun (Gurara of Timimun, a Zenetic Berber language)

Tmz = Tamazight; dls.: AA = Ayt-'Ayyash, AH = Ayt-Hadiddu, AM = Ayt Myill, AN = Ayt-Ndhir, AS = Ayt-Sadden, Iz = Iziyan; ASgr (Ayt Seghrushen) and AIzd (Ayt-Izdeg) are treated as separate lgs.

Wrg = Wargla, Wargli

Wrs = Warsenis (le Zenatia de l'Ouarsenis)

ZAS = Berber dialect of Zayan and Ayt-Sgugu

Zgw = Zaghawa

Zkara

Zkr = Ida-u-Zikri

Zmr = Zemmur (a B dialect)

Zn = Zayan

Zwr = Zwara (a B dialect)

II.2.1.2. SB = South Berber:

Adgg = Twareg of Adghagh

Ah = Twareg of Ahaggar

ETwl = Eastern Tawellemmet

Gh = Ghat

Tdq = Tadghaq (a B dialect of Adghagh of Ifoghas)

Tnsl = Taneslemt

Ttq = Taïtoq

Tw = Twareg; dls.: Tw D = Tadraq, Tw M = dislects of Mali, Tw Ng = dialects of Niger (Twl, Ty), Tw U = Tudalt, Tw Ud = dls. spoken in

Udalan (NE Burkina-Faso, i.e. Tw D and Tw U); Ah, ETwl, Ty and Tnsl are treated here as separate lgs.

Twl = Tawellemet (common name of ETwl and WTwl)

Ty = Tayert, Tayrt, Twareg of Air (Ayr); subd.: Ty KU = Kel-Ui

WTwl = Western Tawellemmet

II.2.1.3. WB = West Berber:

Zng = Zenaga (tuɗ̣ɗ̣ũg̣iya)

II.2.2. Lb = Libyan:

ONum = Old Numidian (= Old East Numidian, Old Libyan)

II.2.3. Gnc = Guanche (dialect cluster); dls.: Fv = Fuenteventura, G = La Gomera, GC = Gran Canaria, Hr = Hierro (Ferro), L = Lanzarote, P = La Palma, T = Tenerife

II.3. *Egyptian* branch:

Cpt = Coptic; OCpt = Old Coptic, dls.: Cpt A = Akhmimic; Cpt F = Fayumic; Cpt B = Bohairic; Cpt L = Lycopolitan (Sub-Akhmimic), Cpt P = the dialect of the *Books of Proberbs* , Cpt S = Sahidic;

DEg = Demotic Egyptian

Eg = Egyptian; stages and variants: Am = Eg of Amarna Texts, BD = Eg of the Book of the Dead ("Totb."), CT = Coffin Texts (Sargtexte), D = Demotic, Eth = Eg of Ethiopian Inscriptions, G = Eg of the Greek-Roman times, LL = Eg of the late and latest (mostly religious) literature (EG's "Lit. Sp."); Md = Eg of Medical Texts, MK = Middle Kingdom Eg, MKL = Eg of Middle Kingdom literature, MP = Eg of the Mathematical papyri, NK = Eg of the New Kingdom, NKL = New Kingdom literature, OK = Old Kingdom Eg, P = Pyramid Texts, RNK = Eg of the ritual texts of the New Kingdom, RT = Eg of the Royal Tombs of Thebae, St = Saite Dynasty (26th Dynasty), Wc = Eg of the Westcar papyrus (spoken Middle Eg), XVIII = 18th Dynasty, XIX = 19th Dynasty, XX = 20th Dynasty, XXII = 22nd Dynasty; L = Late Egyptian (EG's "Sp."), M = Middle Eg, N = New Egyptian (EG's "Nä."), O = Old Eg, fOK = from Old Kingdom on, fP = from the Pyramid Texts on, fMK = from Middle Kingdom on, fNK = from New Kingdom 0n, fO = from Eg O on, fM = from Eg M on, fMd = from Eg Md on, fN = from Eg N on, fXVIII = from the 18th Dynasty on (in other cases, if a word is present in different periods of the history, the most ancient is mentioned)

II.4. C = *Cushitic:*

II.4.1. Bj = Beja; dls.: Bj A = Amar'ar , Bj B = Bishari, Bj Br = Bj of Barka, Bj Hd = Hadendawa, Bj N = the northern dialect (acc. to Reinisch), Bj R = the dialect described by Reinisch (Halanga?), Bj Rp = the dialect described by Roper

II.4.2. Ag = Agaw (Central Cushitic):

Aw = Awngi; dls.: Aw D = Dangela, Aw K = Kwakera

Bln = Bilin

Dmb = Dembiya

Dmt = Damot

Km = Kemant

Knfl = Kunfäl (a language of the Agaw subgroup)

Q = Qwara, Kwara; dialect: Q F = {Flad} 'Falashan'

Xm = Xamir (Hamir, Xamtanga, Khamtanga); dls.: Xm {R} = Hamir, Xm {Ap.} = Khamtanga, Xm T = CR's Hamta, Xm K = Kaïliña, Xm Wg = Xamir of Wag

II.4.3. EC = East Cushitic:

II.4.3.1. LEC = Lowland East Cushitic:

II.4.3.1.1. AfS = Afar-Saho lgs.

Af = Afar; dialect: Af N = Northern dl., Af S = Southern dl.., Af Tjr = Tajurah dl.

Sa = Saho; dls.: Sa HA = Sa of High Assaorta, Sa I = Irob

II.4.3.1.2. SLEC = Southern Lowland East Cushitic (= Omo-Tana):

II.4.3.1.2.1.

Arr = Arbore

Dsn = Dasenech (= Geleba)

Elm = Elmolo

II.4.3.1.2.2. Sam = Sam, Macro-Somali (subfamily of LEC):

pSam = proto-Sam {Heine}

Bn = Boni; dls.: Bn Ba = Baddey, Bn Bi = Bireri, Bn Bl = Bala; Bn Bu = Bura, Bn J = Jara, Bn K = Kili, Bn Kj = Kije, Bn Sa = Safare

Rn = Rendille

pSml = proto-Somali {Lamberti}

Sml = Somali; dls. and sdls.: Ab = Af-Abgaal, Aj = Af-Ajuraan, Ash = Ashraaf dls., 'Aw = Af-'Awramale', B = Benaadir, Bi = Af-Bimaal, C = Central, D = Darood, Db = Af-Dabarre (Doborre), Dg = Af-Degodiya, Dgl = Digil, Dl = Af-Dolbohaante, Dx = Af-Daakhteri, E = Eastern, Ga = Af-Galja'aal, Ge = Af-Geedabuursi, Gn = Af-Gendershi, Gr = Af-Garre, He = Af-Helleedi, Hw = Hawiyya {after R and C}, I = Isaaq, Af-Isaaq (= Sml N), 'I = Af-'Iise, J = Jabarti, Ji = Af-Jiidu, Md = Mudug, Me = Af-Merka, Mj = Af-Majerteen, Mr = Af-Marrehaan, Mt = Max-aad-tiri, My = Af-May, N = Northern dls., NC = North-Central, NE = North-Eastern, Og = Ogaden (Af-Ogaadeen), Oj = Af-Oojji, Or = Af-Oroole, S = Southern, Sha = Af-Shabelle, Shi = Af-Shingaani, T = Af-Tunni, UJ = Upper Jubba, Wr = Af-Wardeyg, X= Af-Xamari

II.4.3.1.2.3. Bs = Baiso

II.4.3.1.3. McOr = Macro-Oromo (Oromoid):

II.4.3.1.3.1. Or = Oromo; dls.: Or B = Borana, Or BI = Isiolo sdl. of Or B, Or Brr = Barareta, Or Gj = Guji dl., Or H = Harar dl., Or O = Orma, Or M = Macha, Or S = Southern dls., Or T = Tulama, Or Wt = Wata, Or Wl = Wälläga; Or AM = Afan Monyo (Karakara) sdl. of Or O

II.4.3.1.3.2. KG = Konso-Gidole:

Bss = Bussa

Di = Dirasha (ḍirā̆ša)

Gato

Gdl = Gidole (Dirayta)

Kns = Konso

Mos = Mossiya

Msl = Mashile (a dialect of the Konso-Geleba subgr.)

Turo

II.4.3.2. Dl = Dullay dialect cluster (= "Werizoid"):

Cm = Tsamako, Tsamay, Sāmakko

Dbs = Dobase

Dihina

Gaba

Gln = Gollango

Grs = Gorrose (a dialect of the Dullay dialect continuum)

Gwd = Gawwada, dial.: Gwd D = Gawwada Dalpena

Hr = Harso

II.4.3.3. Ya = Yaku (= Yaaku, Mogogodo)

II.4.3.4. HEC = Highland East Cushitic:

Alb = Alaba

Brj = Burji

Ged = Gede'o (= Darasa)

Hd = Hadiy(y)a; dialects: Hd Lb = Libido

Kmb = Kambatta

Qbn = Qabenna

Sd = Sidamo; dialect: Sd Hb = Sidamo of Habiela

Tmbr = Tembaro

II.4.3.5. Dhl = Dahalo (belongs either to EC or to SC)

II.4.4. SC = South Cushitic

II.4.4.1. Rt = Rift (subgr. of South Cushitic)

II.4.4.1.1. WRt = West Rift (subgr. of Rift within South Cushitic)

Alg = Alagwa

Brn = Burunge

Grw = Gorowa

Irq = Iraqw

II.4.4.1.2. ERt = East Rift

Asa

Kz = Kwadza, Ngomvia

II.4.4.1.3. Mb = Mbugu, Ma'a (a Bantu language with many SC loans)

II.4.5. ? Klk = the Kuliak languages:

Ik = Ik (= Teuso), a Kuliak language

Ny = Nyang'i (= Nyangiya), a Kuliak language

So = So (= Tepeth, Tepes), a Kuliak language

II.5. Om = *Omotic:*

II.5.1. NrOm = North Omotic (= Kefa-Gimojian subgr. of WOm in FlB NSL 47)

II.5.1.1. Gng = Gonga (a subsubgr. of NOm):

Amuru = Amuru, Amurru

Anf = Anfillo (Southern Mao, after Grotanelli)

Gjb = Gojjebi

Kf = Kaffa

Mch = Mocha (Shekko)

Mnj (one of the sources\components of Kf, according to Fl. MEH)

Shn = Shinashsha (Bworo), dl.: Shn D = Dangur Shinashsha

II.5.1.2. Omt = Ometo (a cluster of dls.\languages) (when unspecified [Zs, Wlt, etc.], Omt {Moreno} is meant)

Bdt = Badditu, Baddito, Koyra

Bsk = Basketo (an Omotic language)

Cha = Chara

Cnc = C'ancha Ometo

COmt = Central Ometo (cover name for several dialects, incl. Gf)

Dc = Dache (an Ometo dialect)

Dk = Doka (an Ometo dialect)

Drz = Dorze-Jo, Dorze (a dialect of the Ometo cluster)

Dwr = Dawro (Kullo), a dialect of the Ometo cluster

Dz = Doze (an Omotic dialect within the Ometo dialect cluster)

Gm = Gamo, Gamu (a dialect of the Ometo cluster)

Gdc = Gidicho (NOmt)

Gemu (an Ometo lge., related to Gf)

Gf = Gofa (an Ometo language)

Gnj = Ganjule (= Ganjawle, a dialect of East Ometo)

Hrr = Haruro

Kcm = Kachama (= Gatsama, a dialect of Ometo)

Krt = Koorete (= Amarro)

Male

Malo

Oyda

Wl = Wolaytta (an Ometo lge.), Wolamo

Zl = Zala

Zrg = Zergulla (a dialect of Ometo)

Zs = Zayse (a dialect of Ometo)

II.5.1.3. Ym = Yemsa (Janjero)

II.5.1.4. Gmr = Gimirra:

Bnc = Bench (Gimirra-Bench, Benesho)

She

II.5.1.5. Ma = Mao (a NrOm dialect cluster) (when it is not specified [BMa, HzMa, etc.), Mao {Grotanelli} is meant)

BMa = Bambes(h)i Mao (= Bambassi)

DMa = Diddesa Mao

GaMa = Ganza Mao

GeMa = Gebsi Mao

HzMa = Hozo Mao

MdMa = Madegi Mao

NrMa = Northern Mao

Sz = Sezo Mao (Seze), Sz1 & Sz 2 = two subdialects of Sezo

II.5.2. Dzd = Dizoid (a subgroup of NOm languages = Maji subgr. of WOm in FlB NSE 47):

Mj = Maji (= Dizi, an Omotic language), d: DJ = Dizi-Jeba

Na = Na'o (Nayi) (a Dizoid language)

Shk = Shako (a Dizoid language, = AY's and Fl's Sheko)

II.5.3. SOm = South Omotic, Aroid (= Ari-Banna, = EOm of FlB NSL 47):

Ari = Ari (a SOm language); dls.: Ari B = Bako (Baka), Ari G = Galila, Ari J = Ari-Jinka, Ari U = Ubamer

ArJ = Ari-Jinka (dialect of Ari)

Bk = Bako (= Ari B)

Dm = Dime

Hm = Hamer (Hamar); dls.: Hm B = Hamar-Ban(n)a, Hm K = Karo (Kara)

II.6. Ch = *Chadic:*

II.6.1. WCh = West Chadic:

II.6.1.1. HAB = Hausa-Angas-Bolewa (Hangbole):

II.6.1.1.1. Hausa gr.:

Gw = Gwandara; dls.: Gw Cn = Chanchara dl. (Arabishi), Gw G = Gitata dl., Gw K = Karshi dl., Gw Kr = Koro dl. (Gwagwa), Gw Nm = Nimbia dl., Gw T = Toni dl. (Garaku)

Hs = Hausa; StHs = Standard Hausa; dls.: Hs B = Bausanchi (Bauchi dl.), Hs D = Dauranchi (Daura dl.), Hs Dm = Damagaranchi (Damagaram dl.), Hs G = Gobiranchi (Gobir dl.), Hs Hd = Hadejia, Hs K = Kananchi (Kano dl.), Hs Kc = Katsinanchi (Katsina dl.), Hs Kt = Katagum, Hs Skt = Sakkwatanchi (Sokoto dl.), Hs Z = Zazzaganchi (Zaria dl.)

II.6.1.1.2. AG = Angas-Goemay (Angas-Sura):

Ang = Angas; dls.: Ang H = High Angas, Ang K = Kabwir dl.

Cp = Chip

Gmy = Goemay, Ankwe

Kfr = Kofyar; dialect: Kfr M = Mernyang (= Merniang, Mirriam)

Mnt = Montol

Mpn = Mupun

Su = Sura

Tal

Ywm = Yiwom (= Gerka)

II.6.1.1.3. BT = Bole-Tangale gr.:

Bele

Bl = Bolewa, Bolanchi, Bole; dialect: BlF = Bolewa of Fika {Meek}

Dr = Dera, Kanakuru

Gera

Glm = Galambu

Grm = Geruma

Kpt = Kupto

Krf = Kirfi, Kirifi

Krkr = Karekare

Kwm = Kwami

Maha

Ngm = Ngamo

Pr = Pero

Tng = Tangale; dialect: Tng B = Biliri dl.

II.6.1.2. Ron lgs.:

Bks = Bokkos

Btr = Butura

Df = Daffo, DfB = Daffo and Butura

Fy = Fyer

Klr = Kulere

Sha

Tmbs = Tambas

II.6.1.3. NrBc = North Bauchi:

Cg = Tsagu

Dir = Diri

Jmb = Jimbin

Kry = Kariya, Kariyanchi

Mbr = Mburku, Mburkanchi

My = Miya, Miyanchi

P' = Pa'a, Pa'anchi

Sir = Siryanchi, Siri

Wrj = Warji

II.6.1.4. SBc = South Bauchi:

Bbr = Bubburè

Bg = Boghom (Burrum)

BG = Bu-gàlàmbu

Bot = Bot, Boot (a South Bauchi language)

Brw = Barawa

Buli

Ds = Dass; dls.: Ds B = Bodli (Zumbul), Ds Bn = Bandas (Dur), Ds D = Dïkshi; Dwat and Wangday are treated as languages

Dw = Dwat (Dwot, Zodi), a dialect of Dass treated as a language

Gj = Geji; dls.: Gj B = Bu (Zaranda), Gj G = Geji proper (Gyanzi), Gj Mg = Migang (Bolu, Pelu)

Grn = Guruntum; dls.: Grn G = Guruntum proper, Grn Mb = Mbaru

Jm = Jimi

Kir = Kir; dls.: Kir K = Kir (Kiir), Kir L = Lar (Balar), Kir Mn = Mansi (Mangas)

Plc = Polchi; dls.: Plc B = Barang (Baram, Dir = Baram Dutse), Plc Ny = Nyamzax (Langas) and Lundur, Plc P = Polchi proper (Posï)

Sy = Saya (Seya, Sayanchi) dialect cluster; dls.: Sy B = Bot (Boot), Sy Zk = Zakshi, Sy Z = Zari

Tala; dls.: Tala L = Lungu (Tala), Tala Sh = Sho (Ju), Tala Z = Zangwal (Sor, Zangwal of Zungur)

Tule

Wnd = Wangday (a dialect of Dass)

Zar = Zar (Zaar, Sigidi); dls.: Zar GL = Zar of Gambar-Lere, Zar K = Zar of Kal, Zar L = Zar of Lusa

Zem = Zem (Zeem); dls.: Zem Ch = Chari, Zem D = Dokshi (Lushi), Zem Z = Zem proper; Tule is treated as a language

Zul (Dira, Diri)

II.6.1.5. NgzB = Ngizim-Bade gr. of languages:

Bd = Bade

Du = Duwai

Ngz = Ngizim

II.6.2. CCh = Central Chadic:

II.6.2.1. McTr = Macro-Tera (subbranch):

Bk = Boka

G'nd = Ga'anda

Gbn = Gabin

Hw = Hwona (Hona)

Jr = Jara

Pdl = Pidlimti

Tr = Tera

II.6.2.2. BM = Bura-Margi subbranch:

Bu = Bura, BuP = Bura Pele

Cb = Chibak

Hld = Hildi

Klb = Kilba (H_ba)

Mrg = Margi; Mrg L = Margi of Lasa {Meek}, Mrg M = Margi of Minthla {Meek}; Mrg P = Margi Putai; Mrg Pl = Plain Margi {Meek}

Ngx = Ngwaxi (Ngwakhi, Ngwahyi)

Wmd = Wamdiu, Wamdiu Margi

WMrg = West Margi

II.6.2.3. McHigi = Macro-Higi (Higi subbranch of CCh):

FlG = Fali Gili

FlK = Fali Kiria, Fali of Kiria

Higi (cluster of dls.)

Hg… = Higi … (dialect cluster); dls.: HgB = Higi Baza, HgF = Higi Futu; HgG = Higi Ghye; HgHm = Higi Humsi {Meek}; HgK = Higi Kamale (= Kps); HgMd = Higi Moda {Meek}; HgMk = Higi Makulu {Meek}; HgNk = Higi Nkafa; HgSn = Higi Sinna {Meek}; HgWl = Higi Wula {Meek}

Kps = Kapsiki (= Higi Kamale)

II.6.2.4. BB = Bata-Bachama subbranch:

Bcm = Bachama

Bt = Bata

BtG = Bata-Garua

BtD = Bata-Demsa

BtM = Bata Malabu

BtZ = Bata Zumo

FlB = Fali of Bwagira

FlJ = Fali of Jilbu

FlM = Fali of Muchella

Gude (Cheke)

Gudu

Mln = Mwulyen

Nz = Nzangi

II.6.2.5. Lmn = Lamang; dls.: Lmn Hd = Hidkala, Hitkalanchi, Lm Vz = Vizik, Lmn A = Alataghwa

II.6.2.6. McMdr = Macro-Mandara (Mandara subbranch):

Dgh = Dghwede, Duxwide, Zeghvana

Gdf = Guduf

Glv = Glavda

Gmrg = Gamergu

Gv = Gava

Mdr = Mandara (Wandala); dialect: Mdr Mr = Mora

Ngs = Ngweshe (= Gvoko, Gboko, Glanda?)

Nkc = Nakatsa

Pdk = Padokwo, Paduko

II.6.2.7. Suk = Sukur

II.6.2.8. McMtk = Macro-Matakam = Matakam subbranch:

Gzg = Giziga; dls.: Gzg D = Giziga Dogba, Gzg Mj = Giziga Mijivin, Gzg Mt = Muturua

Hrz = Hurza

Mada

Mbk = Mboku

Mf = Mafa

MfG = Mofu-Gudur; dialect: MfG M = Mokong

MfM = Mofu-Meri

Mkt = Muktile, Muktele

Mlk = Moloko

Mofu

Mtk = Matakam

Myn = Muyang

pMM = proto-Mafa-Mada

Vm = Vame

Zlg = Zulgo

II.6.2.9. McDb = Macro-Daba, Daba subbranch:

Db = Daba; d: Db H = Daba-Hina, Db K = Kola (treated as a separate lge.)

Kola

Msy = Musgoy

II.6.2.10. Gdr = Gidar

II.6.2.11. McKtk = Macro-Kotoko = Kotoko subbranch:

Bdm = Buduma (Yedina)

Glf = Gulfei

Ktk = Kotoko; dls.: Ktk Af = Affade, Ktk Kl = Klesem, Ktk Ks = Kuseri, Ktk Mk = Makeri

Lgn = Logone; dialect: Lgn M = Mandague
Ngl = Ngala
II.6.2.12. McMsg = Macro-Musgu = Musgu subbranch:
Bld = Baldamu
Mbara
Msg = Musgu; Msg G = Musgu Girvidik (= Munjuk), Msg Ng = Musgu
Ngilemong, Msg P = Musgum-Pus (= Munjuk de Pouss); Mulwi is treated
as a separate language
Msk = Muskum; Msk {Lk.} = Lukas's "Muzgum-Stadt" (Lk. ZSS 142-4)
Mlw = Mulwi (= Vulum, Mogrum, a dl. of Msg)
II.6.2.13 McMs = Macro-Masa = Masa subbranch:
Azm = Azumeyna (Banana-Marba)
Bana {Lukas ZSS}
Bnn = Banana
BnnM = Banana-Mouseye {ChL, Lk. ZSS}
Lame
LamP = Lame-Peve
Ms = Masa
Msm = Misme
Zm = Zime
ZmB = Zime-Batna
ZmD = Zime-Dari
II.6.3. ECh = East Chadic:
II.6.3.1. KwK = Kwang-Kera subbranch:
Kwn = Kwang (Modgel); dial.: Kwn M = Mobu
Ke = Kera
II.6.3.2. Lai = Lele-Kabalay subbranch.:
Dormo = Lukas's "Dormo"
Gabri = Lukas's "Gabri", Bentons "Gabri = Chire"
Kbl = Kabalay (Lukas's "Kaba")
Ll = Lele
Nng = Lukas's "Nangire"
Tbn = Tobanga
II.6.3.3. McSmr = Macro-Sumray = Sumray subbranch.:
Nd = Ndam; dialect: Nd D = Ndam Dik
Smr = Sumray, Somray (Sibine); dls.: Smr G = Gabri, Gaberi
(recorded by GD [DLOuCh 292-301] and by AF and Nacht. [Lk. ZSS 86-
8])
Tmk = Tumak
II.6.3.4. McSkr = Macro-Sokoro, Sokoro subbranch:
Skr = Sokoro
Mw = Mawa
Bar = Barein
II.6.3.5. McDng = Macro-Dangla (Dangla gr., Dangla-Migama)
Dng = Dangla, Dangaleat (a common denomination for West Dangla
& EDng)

EDng = East Dangla (treated here as a separate lge)
Bdy = Bidiya
Mgm = Migama (Jonkor)
II.6.3.6. Mkl = Mokilko, Mokulo
II.6.3.7. McMu = Macro-Mubi (Mubi subbranch):
Brg = Birgit
Jg = Jegu
Kjk = Kajakse
Kjr = Kujarke
Mjl = Minjile
Mu = Mubi

III. K = *Kartvelian*:

III.1. GZ = Georgian-Zan:

III.1.1. OG = Old Georgian; dls.: OG H = the dialect represented in the Haemeti texts, OG X = the dialect of the Xanmeti texts

MG = Middle Georgian (XII-XVI centuries) (Shota Rustaveli, *Kartulis cxovreba,* etc.)

eNG = Early New Georgian (XVII-XVIII c.) (Sulhan-Saba Orbeliani's dictionary)

G = NG = (New) Georgian; dls.: G A = Acharuli (= Ajar) G, G F = Pereidnuli (Fereydan) G, G Gm = Gudamaqruli G, G Gr (and G G) = Guruli G, G HA = High Ajar G, G I = Imeruli G (sdls.: HI = High Imeruli, LI = Low Imeruli), G Imx = Imerxeuli G, G Ing = Inglouri G, G J = Javaxuri G, G K = Kartluri G (incl. G ArX = Aragva Gorge sdl. [Aragvis xeobis kartluri]), G Kx = Kaxuri G, G Kzq = Kiziquri subd. of G Kx, G Lch = Lechxumuri G, G M = Moxeuri G, G Ms = Mesxuri G, G Mt = Mtiuluri G, G P = Pshauri G, G R = Rachuli G, G T = Tushuri G, G Ti = Tianuri G, G UA = Upper Ajar, G X = Xevsuruli; a period: eNG = Early New Georgian (17th c., as registered by Sulxan-Saba Orbeliani)

III.1.2. Zan languages:

Lz = Laz (Chan); dls.: Lz A = Atinuri, Lz Ar = Arxaburi, Arkaburi (архавский), Lz Art = Artashenian, Lz Ch = Chxaletian (чхальский говор), Lz V = Vicuri, Lz VAr = Vicur-Arxaburi, Lz X = Xopuri

Mg = Megrelian; dls.: Mg SmZ = Samurzakan-Zugdidian, Mg Sn = Senakian; Mg BM = Bandza-Martvil sdl. of Mg Sn, Mg Z = Zugdidian sdl. of Mg SmZ

III.2. Sv = Svan; Sv L = Lashxuri dl., Sv LB = Lower Bal dl., Sv Ln = Lentexuri dl., Sv UB = Upper Bal; sdls. of LB: Sv Bc = Becho, Sv Ch = Chubexeuri, Sv Ec = Etseruli, Sv P = Pari, Sv T = Tavrari, Sv Lx = Laxamuluri; sdl. of Sv L: Sv Chl = Choluri; sdls. of UB: Sv I = Ipari, Sv U = Ushguluri (= Ushkuli), Sv Lt = Lat'aluri, Sv M = Mulaxi-Mestia (= Muzhali-Mulaxi, Central UB); dialect areas: LSv = Lower Svan (incl. Sv L and Sv Ln) , USv = Upper Svan (incl. Sv LB and Sv UB).

IV. U = *Uralic*:

IV.1. FU = *Finno-Ugrian*   (= Fenno-Ugrian)

IV.1.1. FP = Finno-Permian,

IV.1.1.1. FV = Finno-Volgaic

IV.1.1.1.1. FL = Finno-Lappish

IV.1.1.1.1.1. BF = Balto-Finnic

Es = Estonian; dilects: Es N = Northern dl., Es S = Southern dl., Es SVl = sdl. of Southern Viljandimaa, Es V = Võru dl.

F = Finnish; dls.: F H = Häme dl., F MNB = Middle- and North-Bothnian dls. (keski- ja pohjoispohjalaiset murteet), F N = Northern dls. (Peräpohjolan murteet, hinterbottnische Dialekte), F SB = South Bothnian (eteläpohjalainen), F SE = Southeastern dls., F Sv = Savo dls., F SW = Southwestern, F U = dls. of Nyland (Uusimaa) (incl. F I = Iitti sdl.); eF = Early Finnish (16th - 18th c.)

Ing = Ingrian (ижорский язык); dls.: Ing Hv = Hevaha dl., Ing O = Oredezh dl., Ing Sk = Soikkola dl.

Krl = Karelian; dls.: Krl A = Aunus (Olonets Krl, livvin kieli), Krl K = Karjala(inen), Krl KA = {SKES} karjala-aunus; Krl Ld = Lude (lyydiläismurteet), Krl Tv = Tver Karelian (sd. of Krl K}

Lv = Livonian; dls.: Lv E = Eastern, Lv W = Western, Lv I = Ira sdl., Lv Slc = Salaca sdl.

Vo = Vote, водский язык, vatja, wotisch

Vp = Veps

IV.1.1.1.1.2. Lp = Lapp, Lappish; dls. & dialect areas: Lp A = Akkala Lp (= бабинский диалект) (belongs to Lp E), Lp Å = Åsele Lp (belongs to Lp S), Lp E = Eastern dls. of Lp (Lp Klt, Lp Kld, Lp T), Lp I = Inari Lp (belongs to Lp E), Lp K = Kola Lp (dialect area of Lp E, including Lp Kld and Lp T), Lp Kld = Kildin dl. (of Lp E), Lp Klt = Koltta Lp (= Kolta Lp, Skolt Lp, belongs to Lp E), Lp L = Lule Lp, Lp N = Norwegian Lp, Lp OSw = Old Swedish dls. of Lp (XVIII c.) (incl. Northern Lp L & Northern Lp S) (after {LÖ}), Lp P = Pite Lp, Lp S = Southern Lp, Lp Sw = Swedish dialect area of Lp (= {SKES} lp R), Lp T = Ter (Turja) Lp (= Kert's йоканьгский диал.) (belongs to Lp E), Lp U = Uume Lp; subdialects: Lp Fi = Finnmarken sdl. (of Lp N, = Ruijanlappi), Lp J = Jemtland sdl. (of Lp S), Lp M = Maritime sdl. (of Lp N), Lp Nd = Neiden (Näätämö) sdl. (of Lp Klt), Lp Nt = Notozero sdl. (of Lp Klt), Lp Pa = Paatsjoki (Boris-Gleb) sdl. (of Lp Klt), Lp Sn = Snåsa sdl. (of Lp S), Lp Snk = Suonikylä sdl. (of Lp Klt), Lp To = Tornio sdl. (of Lp N); sub-subdialects: Lp Ar = Arjeplog ssd. (of Lp P), Lp En = Enontekiö ssd. (of Lp Fi), Lp Fr = Frostviken ssd. (of Lp Å), Lp Gr = Gratangen ssd. (of Lp To), Lp Hr = Härjedalen ssd. (of Lp J), Lp Jk = Jokan'g (Yokostrov) ssd. (of Lp T), Lp Krs = Karasjok ssd. (of Lp Fi), Lp Krsv = Kaaresuvanto ssd. (of Lp To), Lp Kt = Koutokeino ssd. (of Lp Fi), Lp Ml = Malå ssd. (of Lp U), Lp Mr = Meråker ssd. (of Lp J), Lp O = Offerdal ssd. (of Lp J), Lp P = Polmak (Pulmanki) ssd. (of Lp Fi), Lp Pr = Parkalompolo ssd. (of Lp To), Lp Rr = Røros ssd. (of Lp J), Lp Tf = Tysfjord ssd. (of Lp L), Lp Tn = Tännäs ssd. (of Lp J), Lp Ut = Utsjoki

ssd. (of Lp Fi), Lp Vfs = Vefsen ssd. (of Lp Å), Lp Vl = Vilhelmina ssd. (of Lp Å).

IV.1.1.1.2. Chr = Cheremis; dls.: Chr B = Chr of Birsk; Chr Ch = Chr of Cheboksarï; Chr E = Eastern Chr dls.; Chr H = High Chr (= Hill Chr, горно-марийский); Chr K = Chr of Kosmodemyansk (subd. of Chr H); Chr L = Low Chr (= Meadow Chr, лугово-марийский) (to-day StChr L is oficially labelled "Meadow-Eastern, лугово-восточный марийский", but is actually based on Chr L); Chr M = Chr of Malmïzh; Chr NW = Northwestern Chr.; Chr P = Cheremis of the former Perm province; Chr U = Cheremis of Urzhum; Chr Uf (= Chr E Uf) = Cheremis of the former Ufa province; Chr V = Chr of Vetluga; Chr Y = Chr of Yaransk; Chr YO = Cheremis of Yoshkar-Ola (Carevokokshaysk); Chr YU = Chr of Yaransk & Urzhum

IV.1.1.1.3. Mr = Mordvin languages:

eMr = Early Mordvin (18th c., according to SJRN)

Er = Erzya Mordvin; sdls.: A = Alatïr sdl. (the area of the river Алатырь), BI = sdl. of Bolshoye Ignatovo, Iv = Ivancevo sdl., Kal = Kalyayevo sdl., LP = Lower Pyana sdl., Trb = Torbeyevo (former Kazhkïtka) sdl.

Mk = Moksha Mordvin; dls.: P = Mk of the former Penza province

pMr = proto-Mordvin

IV.1.1.2. Prm = Permian

OPrm = Old Permian (древнепермский язык)

Prmk = Permyak; Prmk In = Inva dl. (иньвенский говор), Prmk K = Kochevo dl. (кочёвский говор), Prmk N = Northern dl., Prmk Zz = dl. of the Zyuzdincï (зюздинский диалект)

pZ = proto-Ziryene (proto-Komi) (ancerstor of Z, Prmk, and Yz)

Vt = Votyak; dls.: Vt C = Central Votyak (средний диалект), Vt N = Northern Votyak, Vt S = Southern Votyak, Vt SW = Southwestern Votyak; subdialects: Vt B = Beserman Vt, Vt G = Glazov sdl., Vt Ks = Kosa sdl. (of Vt N), Vt Kz = Vt of the former Kazan province, Vt M = Malmïzh sdl., Vt MU = Malmïzh-Urzhum sdl., Vt Sh = Shoshma subdialect of Vt SW, Vt Sl = Slobodskoy sdl. (district of Slobodskoy, Vyatka oblast), Vt Sr = Sarapul subd., Vt Sm = Vt. of the former Samara province, Vt Tl = Tïlovay subd. of Vt C, Vt Uf = Vt. of the former Ufa province, Vt Ur = Ursïgurt subdialect of Vt SW, Vt Y = Yelabuga sdl.

Yz = Yazvian, Yaz'va dialect (in the Prm subbranch of FU) = коми-язьвинский диалект

Z = Ziryene; dls.: Z EV = Eastern Vïchegda sdl. (of Z UV), Z I = Izhma dl., Z K = Kerchemya dl. (керчемский говор), Z Le = Letka dl., Z LI = Lower Izhma dl., Z LL = Luza & Letka dl., Z Lu = Luza dl., Z LV = Lower Vïchegda dl., Z MS = Middle Sïsola dl., Z MV = Middle Vïchegda dl., Z Mz = Mezen' dl., Z N = Northern dls., Z P = Pechora dl., Z Pr = Prupt dl., Z Sk = Sïktïvkar dl., Z Ss = Sïsola dl., Z Ud = Udora dl., Z US = Upper Sïsola dl., Z UV = Upper Vïchegda dl., Z V = Vïchegda dl., Z Vm = Vïm dl., Z Vsh = Vishera sdl. (of Z UV).

IV 1.2. Ur = Ugric

IV.1.2.1. ObU = Ob-Ugric

Os = Ostyak; dls. & dialect areas: Os Ag = Agan sdl. (of Os Sr), Os B = Beryozovo sdl. (of Os O), Os Cng = Cingala sdl. (of Os I), Os D = Demyanka dl., Os E = Eastern dialect area, Os I = Irtïsh dls. (collective denomination), Os K = Konda dl., Os Km = dl. of Kaminskoye, Os Kr = dl. of Krasnoyarskie, Os Ks = Koshelevsk sdl. (of Os I), Os Kz = Kazïm dl., Os LD = Lower Demyanka dl. (= DT), Os Lk = Likrisovskoye dl., Os LK = Lower Konda sdl., Os MY = Malïy Yugan dl., Os N = northern dialect area, Os Nz = Nizyam dl., Os O = Obdorsk (Salehard) dl., Os Pïm = Pïm (Pim) dl. (sd. of Os Sr), Os Pt = Pitlyar dl. (sd. of Os N), Os Sh = Sherkalï dl. (= Middle Ob dl.), Os Shr = Shuryshkar dl., Os Sl = Salïm dl., Os Sn = Sïnya dl., Os Sr = Surgut dl., Os Ty = Tremyugan dl., Os UA = Ust-Agan dl., Os UD = Upper Demyanka dl. (= DN), Os Uy = Ust-Yugan dl., Os V = Vakh dl., Os VK = Verknhe-Kalïmsk dl.;,Os Vrt = Vartovskoye dl., Os Vy = Vasyugan dl., Os Y = Yugan dl. (sd. of Os Sr), Os Z = Zavodniye (Zavodinskiye) dl.

OVg = Old Vogul (18th cent.); dls.: OVg E = Eastern, OVg N = Northern, OVg S = Southern, OVg W = Western; sdls.: OVg E TM = a sdl. (of OVg E) labelled by Honti as "ᴛᴍ", OVg I = Is sdl. (of OVg W [?]), OVg L = Lyalya sdl. (of OVg W [?]), OVg N Ber = Berezovo (OVg N; three variants: OVg N BerG, OVg N BerO and OVg BerK), OVg N B and OVg N Chd = sdls. (of OVg N) labelled by L. Honti as "ʙ" and "Čd", OVg N NSs = Northern Sosva sdl. (OVg N), OVg N SoG = a sdl. of OVg N labelled by Honti as "ꜱₒɢ", OVg N SoO = a sdl. of OVg N labelled by Honti as "ꜱₒO", OVg S Chus = Chusovaya sdl. (of OVg N; two variants: OVg S ChusO and OVg S ChusM), OVg S Kg = Kungur sdl. (of OVg S), OVg S SSs = Southern Sosva sdl. (OVg S), OVg S Tg = Tagil sdl. (OVg S), OVg S Tr = Tura sdl. (OVg S), OVg S Vt = Verxoturye (Верхотурье) sdl. (OVg S), OVg Str = a sdl. (of OVg W [?]) labelled by Honti as "ꜱᴛʀ", OVg Tb = a sdl. (of OVg E or [less plausibly] OVg S?) labelled by Honti as "**Tob**" (= Tobol?), OVg W P = Pelïmskoye sdl. (OVg W), OVg W Sol = Solikamsk sdl. (OVg W), OVg W UsU = Ust-Ulsuy (Усть-Ульсуй) sdl. (OVg W) of the 19th cent. (cf. Kann AWD); the sigilla "ʙ", "ꜱₒɢ", "ꜱₒO", "ꜱᴛʀ", "Tob" and "ᴛᴍ" (used by Honti after J. Gulya) remain enigmatic because their source (Gulya's manuscript paper "Altwogulische Dialekte") has not yet been published and is not available to the present writer; the queries "[?]" belong to Honti

Vg = Vogul; dialectal areas: Vg E (= Vg K) = Eastern (Konda) Vogul (incl. dls.: Vg LK = Lower Konda dl., Vg MK = Middle Konda dl., Vg UK = Upper Konda dl., Vg MO = Middle Ob [Sherkal] dl., Vg Yk = Yukonda dl.), Vg N = Northern Vogul dialect area (incl. Vg UL = Upper Lozva dl., Vg Ss = Sosva dl., Vg Sg = Sïgva dl.), Vg S = Southern Vogul (Vg T = Tavda dl., Vg TCh = sd. Chandïri of the Tavda dl., Vg TG = sd. Gorodok of the Tavda dl., Vg TY = sd. Yanïchkova of the Tavda dl.), Vg W = Western Vogul (incl. Vg LL = Lower Lozva dl., Vg ML = Middle Lozva dl.,

Vg NV = North Vagilsk dl. [sds: Vg NVK = sdl. of the village Kama, Vg NVZ = Zaozërnaja sdl.], Vg P = Pelïmka dl., Vg SV = South Vagilsk dl.; Vg V = Vagilsk dls.]

IV 1.2.2. Hg = Hungarian; d: Hg S = South Hungarian (sd.: Hg O = Ormánság sdl.)

OHg = Old Hungarian

IV.2. Sm = *Samoyed*

IV.2.1. NrSm = North Samoyed subgr.

Ne = Nenets; dls.: Ne F = Forest Nenets (sdls.: Ne F K = Konda, Ne F Ks = Kiselevskaya, Ne F L = Lyamin, Ne F Ny = Nyalina), Ne T = Tundra Nenets (sd.: Ne BZ = Bol'shaja Zemlja (ssd: Ne Sd = Ne BZ registered in the area of†the Syaida river), Ne Kn = Kanin sdl., Ne O = Obdorsk (Salehard) sdl., Ne Ym = Yamal sdl.)

Ng = Nganasan (Tavgi)

En = Enets; dls. : En B = Baikha dl. (Bay); En K = Karasino dl.; En M = the dl. of Mangazeya (XVIII c.); En T = Enets of Turukhan region; En Tn = Tundra dl. ({Hl.}); En X = Khantaika (Хантайка) dl. (Somatu, Madu)

Yr = Yurak (an extinct lge. akin to Ne and En)

IV.2.2. Slq = Sölqup (Selkup); dls.: Slq B = Baikha dl., Slq Ch = Chaya dl., Slq Chl = Chulïm dl., Slq F = Farkovo sdl. (of Slq Yn), Slq Kar = Karasino dl., Slq Ke = Ket' dl., Slq LKe = Lower Ket' dl., Slq LO = Lower Ob dl., Slq LTz = Lower Taz dl., Slq MKe = Middle Ket' dl., Slq MO = Middle Ob dl., Slq MTm = Middle Tïm dl., Slq MTz = Middle Taz dl., Slq NP = Nat-Pumpokolsk dl., Slq Nr = Narïm dl., Slq O = Ob dls., Slq Tm = Tïm dl., Slq Tur = Turukhan dl., Slq Tz = Taz dl., Slq UKe = Upper Ket' dl., Slq UO = Upper Ob dl., Slq UTz = Upper Taz dl., Slq V = Vakh dl., Slq Vy = Vasyugan dl., Slq Yel = Yeloguy dl., Slq Yn = Yenisey dl.

IV.2.3. Kms = Kamassian

Koyb = Koybal (a Samoyed language, actually a dialect of Kms)

IV.2.4. Mt = Mator (= Mator-Taygi-Karagas); dls.: Mt T = Taigi, Mt M = Mator (Motor) proper, Mt K = Karagas; Mt A = Abakan (a Sayan Samoyed dialect or a conglomerate of data mostly of Mt origin, after Msr. and Strl.)

IV.3. Y = *Yukagir;* Y = Yukagir language group

OY = Old Yukagir; dls.: Ch = Chuvan, K = Kolïma, O = Omok, NW = Northwestern Y (Ust-Yansk)

Y = Yukagir; dls.: Y K = Kolïma Yukagir (Jochelson's Upper Kolïma Y), Y T = Tundra Yukagir

V. A = Altaic

Hun = Hunnic (Hsiung-nu)

V.1. T = *Turkic*

ppT = Early proto-Turkic

V.1.1. NaT = Narrow Turkic, Common Turkic (the proto-language of all T languages except Bulghar and Chuvash)

V.1.1.0. OT = Old Turkic; dls. (after Clauson and other authors): OT O = Orkhon dl., OT Og = Old Oghuz, OT OY = Orkhon and Yenisey dialect(s) (Kök-Türkisch), OT Qp = Old Qïpchaq, OT QU = Qarakhanid Uyghur (Xakani), OT Tü = Türkü, OT U = Old Uyghur, OT Y = Yenisey dialect

MT = Middle Turkic

MU = Middle Uyghur (d. of MT)

V.1.1.1. Og = Oghuz (= Southwest Turkic)

'AQ = 'Ali-Qurchi Turkic (a SOg dialect, to the south of Arak, Iran)

Az = Azeri; dls.: Az Erz = Erzerum dl., Az Mgn = Mughan sdl., Az Nx = Nukha dl., Az Qb = Quba dl., Az Qz = Qazakh dl., Az Shm = Shamakhï dl., Az Sl = Salyan dl. (сальянский диалект), Az Tbr = Tebriz dl.

Afsh = Afshar

AfshN = SOg of Afshar-e Nanakchi (near Kabul)

FA = Firuz-Abad Turkic (a SOg dialect, Iran)

FX = SOg of Qal'aye Farhad-Xan (to the NE of Kermanshah, Iran)

Ggz = Gagauz

HAS = SOg of Hoseyn-Abad-e Sarmashad (near Kazerun, Iran)

MOg = Middle Oghuz; dl.: MOg Tkm = Türkmen dl. of MOg ("Old Türkmen")

MOsm = Middle Osman Turkic

MT Tkm = Türkmäni Middle Turkic (a dl. of MOg)

NEXT = Northeastern Khorasan Turkic (dialect cluster); dls.: G = Gujgi, J = Jonk, L = Langar, M = Mareshk

NWXT = Northwestern Khorasan Turkic (dialect cluster); dls.: A = Asadli, B = Bojnurd, ShT = Sheykh-Teymur

NXT = Northern Khorasan Turkic (dialect cluster); dls.: D1 = Dara-Gaz 1, D2 = Dara-Gaz 2, Dg = Dougha'i, L = Lotf-abad, Q = Quchan, Shi = Shirwan, Shu = Shurak, Ze = Zeyarat, Zo = Zourum

OOsm = Old Osman Turkic

Osm = Osman Turkic

Prdm = Paradomba Turkic (a SOg dial., to the west of Borujin, Iran)

Qrw = Qorwa Turkic, SOg of Qorwa (to the NE of Kermanshah, Iran)

SA = Soleyman-Abad Turkic (a SOg dialect, Iran)

SEXT = Southeastern Khorasan Turkic (dialect cluster); dls.: ChS = Charam-Sarjam, K = Kalat, R = Ruh-abad, XO = Kharwe-'Olya

Shhr = Shahrak Turkic (a SOg dialect, to the east of Shahre-e Kord, Iran)

Slr = Salar; dls.: Slr A = Slr of Altiyuli, Slr X = Slr of Khanbakh, Slr U = Slr of Ujirem, Slr Ul = Slr of Ullaghïl

Snqr = Sonqor Turkic (a SOg dialect, to the NE of Kermanshah, Iran)

SOg = Southern Oghuz dls.

SWXT = Southwestern Khorasan Turkic (dialect cluster); dls.: H = Hokm-abad, J = Joghatay, PK = Pir-Komaj, QB = Qara-Bagh, SA = Soltan-abad

Tk = Turkish; dls. and sdls.: Tk An = Anatolian dls., Tk Çr = Çorum sdl., Tk Er = Erzurum sdl., Tk İç = İçel sdl., Tk Ist = Istanbul sdl., Tk Kn = Konya sdl., Tk Qrpp = Qarapapaq (Karapapak) dl., Tk Rh = Tk of the Rhodope Mountains, Tk WAn = Tk of Western Anatolia

Tkm = Turkoman (Türkmen); dls.: Tkm NC = North Caucasian Turkoman (туркменский говор Ставрополья), NY = North Yomud dl.

XT = Khorasan Turkic (a collective name for NEXT, NWXT, NXT, SEXT, and SWXT)

XwT = Xwarezmic Turkic

V.1.1.2. Qïpchaq (= Northwest Turkic) lgs.:

Blq = Balqar

Brb = Baraba (= STt B); sdl.: Brb Tk Tarmakül sdl.

Bsh = Bashkir (Bashqort)

Cmn = Cumanic (= MQp Cm)

CrTt = Crimean Tatar

Kr = Karaite (Karaim); dls.: Kr Cr = Crimean dl., Kr G = Galich dl., Kr L = Luck (Луцк) дл., Kr T = Trakai (Troki) dl.

MQp = Middle Qïpchaq; variations: MQp A = Armeno-Qypchaq, MQp Cm = Cumanic (of the Codex Cumanicus), MQp Mm = Mamluq-Qypchaq (Egyptian Qypchaq)

MsTt = Mishär Tatar

Nog = Noghay; dls.: Nog A = Aqnoghay, Nog P = Noghay proper, Nog Q = Qaranoghay

Qmq = Qumïq

Qp = Qïpchaq

Qq = Qaraqalpaq; Qq X = Ramstedt's "Chagatay of Xiwa (Хива)"

QrB = Qarachay-Balqar

Qzq = Qazaq

SbTt = Siberian Tatar; dls.: SbTt B = Baraba (= Brb), SbTt Ichk = Ichkina dialect (d. of the Ichkina river, SW-Siberia), SbTt TI = Tobol-Irtysh dialect (incl: SbTt Tb = Tobol Tatar [= TbTt], SbTt Bkl = Baykalovo subsubd., SbTt Kk = Kükrände subsubd. [= Cheburga subsubd. of TbTt], SbTt Ltm = Laytamak subsubd., SbTt Tr = Tara sdl., SbTt Tv = Tevriz Tt, TbTt = Tobol Tatar [= SbTt Tb]), SbTt Tö = Tömen Tt (тюменский дл.), SbTt Tom = Tomsk dialect (incl.: SbTt EuCh = Eushta-Chat subd., SbTt Ql = Qalmaq, SbTt OCh = Orsk Chat, орский подговор чатов)]

VTt = Volga Tatar (= Kazan Tatar); dialects: VTt K = Christian Tatar (кряшенский), VTt Ks = Kasymov dl. (касымовский), VTt W = Western dl., VTt M = Middle Tatar dl., VTt E = Eastern dl.; subdialects: VTt H = Highland sdl. of VTt M (= тау ягы сөйләшләре, говоры нагорной стороны Татарстана), VTt I = Ichkina sdl. of VTt M (ичкинский говор) in West Siberia, VTt Mn = Menzelya sdl. of VTt M (мензелинский говор), VTt Smb = Simbirsk sdl., VTt TYK = Christian Highland Tatar sdl. (тау ягы керәшеннәре сөйләше =

подберезинский говор), VTt TYT = Tarkhan sdl. of Highland Tatar (тау ягы, тархан сөйләше = тарханский говор), VTt U = Ural Tatar (говор уральских татар)

V.1.1.3. QrgA = Qïrgïz-Altay languages (Central-Eastern Turkic):

Alt = Altay-Kizhi; StAlt = горно-алтайский, ойротский

Ln = Lobnor Turkic

Qmn = Qumanda, Kumanda Tatar

QK = Lebed' Tatar, Quu-Kizhi (Chalkan, язык лебединских татар)

Qrg = Qïrgïz; Qrg S = Southern dls. of Qrg; Qrg T = Talas dl.

StAlt = горно-алтайский, ойротский

Tb = Tuba (диалект черневых татар, туба-кижи)

Tln = Telengit (Tälängit)

Tlt = Teleut (Tälängät)

V.1.1.4. SET = Southeast Turkic (Baskakov's "Qarluq Turkic"):

ET = East Turkic (= New Uygur); dls.: ET G = Guma dl., ET H = Hami dl., ET K = Kashghar dl., ET Kc = Kucha dl., ET Ta = Tashmaliq dl., ET Tr = Taranchi, ET X = Khotan dl., ET Y = Yarkand dl.V.

Chg = Chagatay (West Türkistan Islamic literary lge, late XIV-XX); Chg Xw = Chagatay of Xwarezm

QT = Qaraxanid Turkic (West Türkistan, XII-XIV) (the tafsirs, Rabghuzi, Ibn-Muhanna)

Tki = Türki (traditional literary language of East Turkistan)

Uz = Uzbek; dls.: Uz Af = Uz dls. in Afganistan, Uz Nm = Namangan dl., Uz NmA = Namangan-Andizhan dls., Uz Qp = Qïpchaq dls., Uz Srt = dls. labelled "Sart" in the old literature (≈ Uz NmA), Uz U = urban (Iranized) dls., Uz XrOg = Xwarezmic-Oghuz dls.

V.1.1.5. NET = Northeast Turkic:

Bltr = Beltir (today turned to a dialect of Xk)

Chl = Chulïm (чулымский); dls.: Kü = Küärik (treated as a separate lge), Chl U = Upper Chulïm, Chl M = Middle Chulïm

SY = Sarïg-Yugur (= Western Yugur)

Kü = Küärik or Küärük (dialect of Chulïm)

Qb = Qoybal (= Koybal Turkic); dl.: Qb Sl = Salbin dl.

Qc = Qacha (качинский диалект «хакасского» языка)

Qzl = Qïzïl (кызыльский диалект т. н. «хакасского» языка)

Sg = Saghay (сагайский диалект «хакасского» языка)

Shor (шорский язык + шорский диал. «хакасского» языка)

Xk = Khakas (Xakas, хакасский яз., Abakan Turkic) (dialect cluster; StXk is based on Sg & Qc)

V.1.1.6. Tuva-Tofalar (Sayan Turkic)

Tf = Tofalar

Tv = Tuva; Tv NE = Norteastern dl. (Toju dl., тоджинский диал.)

V.1.1.7. Xlj = Xalaj, Khalaj

V.1.1.8. Yakut sg.

Yk = Yakut

Dlg = Dolgan

V.1.2. pBlgh = proto-Bulghar

Blgh = Bulghar

Chv = Chuvash; dls.: Chv H = High Chuvash (верховой [вирьял] диалект; Chr K = Kurmïsh sdl. of Chv H; Chv KA = Krasnoarmeysk sdl. of Chv H; Chv Mr = Morgaush sdl. of Chv H; Chv L = Low Chuvash (низовой [анатри] диалект); Chv M = Morgaush sdl. of Chv H; Chv MK = Malo-Karachkino dl. of Chv; Chv V = Vurnar sdl. of Chv H

V.1.3 (non yet classified): Xzr = Xazar (Khazar, хазарский яз.)

V.2. M = *Mongolic*

Ba = Bao'an (Pao'an, баоаньский)

Brt = Buryat; dls.: Brt A = Alar dl., Brt Ag = Aga dl. (агинский говор), Brt E = Eastern dls. of Brt, Brt NU = Nizhneudinsk dl.

WrM = Written Mongolian (Script Mongolian, Schriftmongolisch, Classical Mongolian)

WrO = Written Oyrat

Dg = Dagur; dls.: Dg B = Butha D, Dg Cc = Cicikar Dg, Dg Hl = Hailar Dg

Dx = Dongxiang, Tunghsiang, Santa, дунсянский язык

HlM = Halha-Mongolian

IM = Mongolian dialects of Inner Mongolia; IM H = Hejing dialect

Kl = Kalmuck; dls.: D = Dörböt, Ö = Ölöt, T = Torgut

Mgl = Moghol; dialect: Mgl Mr = Marda

MM = Middle Mongolian; dls. (variants): MM E = Eastern MM, MM W = Western MM

MMgl = Middle Moghol

Mnr = Monguor; dls.: Mnr E = Eastern Monguor (Dongbuyuguyu), Mnr M = Minhe Monguor (= Sanch'uang, Potanin's сань-чуан), Mnr H = Huzu Monguor (Uyangpu, Potanin's у-ян-бу, Tuzuyu), Mnr Nr = Naringol sdl. of Mnr H (SM's Monguor)

Oyr = Oyrat; dls.: Oyr B = Bayit (Bayat), Oyr T = Torgut, Oyr ET = East Torgut (in Sinkiang, Rm's "Osttorgutisch")

OM = Old Mongolian

Ord = Ordos Mongolian (a dialect treated here as a separate lge.)

PClWrM = Pre-Classic Written Mongolian

Shrn = Shirongol (a collective name for Monguor, Dongxiang, and Baoan)

ShY = Shira-Yughur (= Shera-Yöghur, Jegün Yogur, Eastern Yugur)

Trgt = Torgut (a dialect of Kalmuck and Oyrat)

V.3. Tg = *Tungusic (= Tungusian), Tungus-Manchu*

V.3.1. NrTg = North Tungusian (subgroup of the Tungusian language family)

Ewk = Ewenki; dls.: Ewk E = Eastern, Ewk N = Northern, Ewk S = Southern; subdialects: Ewk A = Ayan subd. (of Ewk E), Ewk Ag = Agata & Bolshoy Porog subd. (of Ewk N), Ewk Ald = Aldan subd. (of Ewk E), Ewk B = Baykit subd. (of Ewk S), Ewk Bnt = Baunt subd. (of Ewk S), Ewk Brg = Barguzin subd. (of Ewk S), Ewk Chlm = Chulman subd. (of Ewk E), Ewk

Chmk = Chumikan subd. (of Ewk E), Ewk D = Dudinka subd. (of Ewk N), Ewk Hng = Hingan (Хинган) subd. (of Ewk E), Ewk I = Ilimpeya subd. (of Ewk N), Ewk Kch = Kachug subd. (of Ewk E), Ewk M = Maya subd. (of Ewk E), Ewk NB = Northern Baykal subd. (of Ewk S), Ewk Nk = Nakanna subd. (of Ewk N), Ewk Np = Nepa subd. (of Ewk S), Ewk Nr = Nercha subd. (of Ewk E), Ewk O = Olekma subdialects (of Ewk E), Ewk PT = Podkamennaya-Tunguska subdialects (of Ewk S), Ewk Skh = Sakhalin subd. (of Ewk E), Ewk Sm = Sïm subd. (of Ewk S), Ewk Tk = Tokko subd. (of Ewk E), Ewk Tkm = Tokma subd. (of Ewk S), Ewk Tmt = Tommot subd. (of Ewk E), Ewk Tng = Tungir subd. (of Ewk E), Ewk Tp = Timpton subd. (of Ewk E), Ewk Tt = Totta subd. (of Ewk E), Ewk U = Uchami subd. (of Ewk S), Ewk UA = Upper Amur subd. (of Ewk E), Ewk Ucr = Uchur subd. (of Ewk E), Ewk UL = Upper Lena subd. (of Ewk S), Ewk Urm = Urmi subd. (of Ewk E), Ewk V = Vanavar subd. (of Ewk S), Ewk Vl = Vilyuy subd. (of Ewk E), Ewk Vtm = Vitim subd. (of Ewk E), Ewk Y = Yerbogachen subd. (of Ewk N), Ewk Z = Zeya subd. (of Ewk E)

Lm = Lamut (Ewen, эвенский язык); dls.: Lm A = Arman d., Lm C = Central Lamut, Lm E = Eastern Lamut, Lm W = Western Lamut; sds.: Lm An = Anyuy subd. (of Lm C), Lm And = Anadïr subd. (of Lm E), Lm B = Bïstraya subd. (of Lm E), Lm KO = Kolïma-Omolon subd. (of Lm E), Lm M = Moma subd. (of Lm C), Lm N = Northern Lm (a subd. of Lm E), Lm O = Okhotsk subd. (of Lm E), Lm Ol = Ola subd. (of Lm E), Lm P = Penzhina subd. (of Lm E), Lm Sk = Sakkïrïr subd. (of Lm W), Lm T = Tompon subd. (of Lm C), Lm Tg = Tügesir subd. (of Lm W) Lm Y = Yukagir subd. (of Lm W)

Neg = Negidal; dls.: H = High Amgun (верхнеамгуньский), L = Low Amgun (нижнеамгуньский),

V.3.2. AmTg = Amur Tungusian (subgroup of the Tungusian language family)

Nn = Nanay (Gold); dls.: Nn A = Amur dialect gr. (Nn Nh, Nn SA & Nn G), Nn B = Bikin dl., Nn G = Garin dl. (= Samar dl.), Nn K = Kili (a dial. area, incl. Nn KU, Nn Sn & Nn UU [Doerfer considers Kili a separate language]), Nn KU = Kur & Urmi dl., Nn Nh = Naykhin dl., Nn SA = Sakachi-Alan, Nn Sn = Sungari dialect gr. (incl. Nn B), Nn UU = Upper Ussuri dl.

Orc = Orochi

Ork = Orok

Sln = Solon

Ud = Udihe (Ude, удыхейский, удэйский, удэгейский); sdls.: Ud A = Anyuy sdl., , Ud B = Bikin sdl., Ud I = Iman sdl., Ud K = Koppi sdl., Ud X = Khor (Xor) sdl., Ud Sm = Samarga sdl.

Ul = Ulcha, Olcha

V.3.3. STg = South Tungusian

Jrc = Jurchen

WrMc = Written Manchu (Script Manchu, Classical Manchu)

Mc = Manchu; dls.: Mc N = Modern Northern Manchu, Mc Sb = Sibe Manchu

PClWrMc = Pre-Classical Written Manchu (according to *TF*)

V.4. Ko = *Korean;* dls.: Ko Chs = Chŏngsando, Ko Chj = Chejudo, Ko Chl = Chenla, Ko Hm = Hamgyŏngdo, Ko Ks = Kyŏngsangdo (incl. Ko Ks S = South Kyo_ngsangdo), Ko Kw = Kangwŏndo, Ko N = Northern dls., Ko NW = Northwestern dls., Ko Ph = Phyŏngyang, Ko PhN = Phyŏngyang-Namdo, Ko Sl = Seoul (Sŏul)

ClKo = Classical Korean (Korean in Rm. SKE) (early NKo that still distinguishes between ɑ and ʌ)

MKo = Middle Korean

NKo = New Korean, Modern Standard Korean

OKg = Old Koguryo

OKo = Old Korean (Silla)

V.5. JK = *Japanese-Kogurö* family

ClJ = Classical (Literary) Japanese (Bungo, based on ltOJ)

J = Japanese; dls.: J Ak = Akida, J Ht = Hateruma, J Ik = Ikema, J Is = Ishigaki, J Kg = Kagoshima, J Km = Kameyama, J K = Kyoto, J Kt = Keto, J Nk = Nakasuji, J Ns = Nase, J Rk = Ryukyu dls., pRk = proto-Ryukyu, J Sh = Shuri, J Sz = Sudzu, J Tk = Tokyo (= StJ), J Ty = Toyama, J Y = Yonakuni

Kgr = Kogurö, Kokuryo_

ltOJ = Late Old Japanese (9th-11th c.)

MJ = Middle Japanese (12th-16th c.)

OJ = Old Japanese (to the 8th c.); OJ E = Eastern Old Japanese

pJ = proto-Japanese

VI. D = Dravidian

VI.1. SD = South Dravidian

AlK = Alu-Kurumba 【Ālu Ku<u>r</u>umba】 (a Dravidian dialect of the Nilgiri area)

Bel = Belari

Brgd = Burgandi

Irl = Irula

JKr = Jenu-Kurumba 【Jēnu Ku<u>r</u>umba】 (a Dravidian dialect of the Nilgiri area)

Kdg = Kodagu; dls.: K = Karaḍa, MV = Merchara-Virajpet, Nl = Nalknad

Kkd = Kaikadi

Kn = Kannada; dls.: B = Barkur, Bd = Badaga, Bl = Bellary, Cr = Coorg Kn (Jenu Kuruba), G = Gowda, Gl = Gulbarga, Hl = Halakki, Hv = Havyaka, K = Kurumba, Km = Kumta, Nn = Nanjangud, O = Ola, R = Rabakavi, Sh = Sholiga, T = Tiptur

Krb = Kuruba (Betta-Kuruba)

Krmb = Alu-Kurumba, Jenu-Kurumba, and Pal-Kurumba (Dravidian dialects of the Nilgiri area)

Krg = Koraga; dls.: Krg M = Mudu, Krg O = Onti, T = Tappu

Ksb = Kasaba (= Kasava), a D language (or dial. of Irula)

Kt = Kota

McTm = Macro-Tamil (= Tamil and Malayalam)

Ml = Malayalam

OKn = Old Kannada

OTm = Old Tamil

PaK = Palu-Kurumba 【Pālu Kuṟumba】 (a Dravidian dialect of the Nilgiri area)

Td = Toda

Tm = Tamil; dls. and variants: ClTm = Classical (Literary) Tamil, B = Brahmani Tm, K = Kollimalai Tm, LP = Tm of Lower Perak, NA = North Arcot, T = Tirunelvali Tm, W = Western Tm

TmM = Tamil and Malayalam

Tu = Tulu; Tu Br = Brahmin dl. of Tulu

VI. 2. SCD = South-Central Dravidian (Telugu-Kui) {Km., Zv.}:

VI. 2.1: Telugu:

OTl - Old Telugu

Svr = Savara

Tl = Telugu; Tl Brh = Brahman Telugu; dls.: Tl G = Guntur dl., Tl Mrl = Merolu dl.

VI.2.2. GnD = Gondvana Dravidian {An.}:

Gnd = Gondi; dls.: Gnd A = Adilabad Gnd, Gnd B = Gnd of the Betul district, Gnd Ch = Chhindwara dl., Gnd ChM = Maria Gondi of the Chanda district, Gnd D = Gnd of Durg, Gnd DM = Gnd of the Dandami Marias, Gnd G = Gnd of Gadchiroli tahsil (Chanda district), Gnd HM = Gnd of the Hill Marias (incl. Gnd HMB, HMD, HMO, HMS = Gnd HM of the areas of Bogan Pallahor, Dhobi, Orcha village, and Sironcha tahsil respectively), Gnd K = Koya Gondi (= Koya, considered a separate lge by some scholars), Gnd KB = Gnd of Bhadrachalam and Rekapalli, Gnd KD = Gnd of the Koyas of Dorlas, Gnd KM = Gnd of the Koyas of Malkangiri tahsil, Gnd KS = Gnd of Koyas and Dorlas of South Bastar, Gnd MB = Gondi of the Bison Horn and Dandami Marias (Bastar), Gnd Mn = Gondi of Mandla, Gnd Mrd = Mardia Gondi, Gnd Mu = Gnd of the Murias of North Bastar, Gnd Nr = Gnd of the Murias of Naraipur (NW. Bastar), Gnd RCh = dialect of the Raj Gonds of Chanda district, Gnd RSr = dialect of the Raj Gonds of Sironcha tahsil; Gnd Sn = Gnd of Seoni, Gnd Y = Gnd of Yeotmal

Knd = Konda; dl.: Knd P = Pulgura dl.

KK = Kui-Kuwi lgs (Kui and Kuwi)

Kui; dl.: Kui KK = Kui of Kuttia Kandhs

Ku = Kuwi; dls.: Ku D = dl. of Ḍongriya Kondhs, Ku K = Kubi dialect (labelled so by DEDR XXX), Ku Kt = Kuttia Kuwi, Ku P = dl. of the Parja Kondhs of Bisamkatak, Ku Su = Sunkarametta, Ku T = Ṭēkriya Kondh

Mnd = Manda

Png = Pengo

VI.3. CD = Central Dravidian

Gdb = Gadba; dls.: Gdb K = Koṇḍēkor Gadba, Gdb O = Ollari, Gdb P = Gdb of Pottangi (Koraput distr.), Gdb Sl = Salur

Klm = Kolami

Nk = Naiki of Chanda (= Nk. [Ch.] of DEDR)

Nkr = Naikri (dialect of Kolami) (= Nk. of DEDR)

Prj = Parji

VI.4. NED = Northeastern Dravisian

Krx = Kurux, Kurukh, Oraon

Mlt = Malto

VI.5. Brh = Brahui

VII. E = *Elamite* (family)

AchEl = Achaemenian Elamite (from the 6th cent. B.C.)

El = Elamite (language)

MEl = Middle Elamite (13th through 11th cent. B.C.)

NEl = New Elamite (8th through 7th cent. B.C.)

OEl = Old Elamite (23rd through 14th cent. B.C.)

VIII. Gil = *Gilyak* (Nivkh); dls.: Gil A = Amur dl., Gil ES = East Sakhalin dl.

IX. CK = *Chukchee-Kamchadal*, Kamchukchee

IX.1. ChK = Chukchee-Koryak (proto-Chukchee)

Chk = Chukchee

Aly = Alyutor, алюторский

Kor = Koryak (Chawchuwen Koryak, "нымыланский")

Pln = Palana Koryak

IX.2. Im = Itelmen (= Kamchadal); dls.: Im W, Im E, Im S

X. EA = *Eskimo-Aleut*

X.1. Esk = Eskimo; dls.: Esk I = Inupiaq-Inuit, Esk Y = Yupik; subdialects: Esk AY = Alaskan Yupik (incl.: PY = Pacific Y [PY K = Konyag, PY Ch = Chugach], CAY = Central AY [CAY BB = Bristol Bay, CAY K = Kuskokwim, CAY Nun = Ninivak Island, CAY Nl = Nelson Island, CAY HBC = Hooper Bay & Chevak, CAY Y = Yukon, CAY NS = Norton Sound with CAY NSU = Unalik]), SbY = Siberian Yupik (incl.: SbY Na = Naukan, CSbY = Central SbY [= Chaplino], SbY Sr = Sireniki), Esk WlI = Wales I, Esk AI = North Alaska I, Esk CI = Canada I (incl. Esk MkI = Mackenzie Bay I), Esk LI = Labrador I, Esk WGI = West Greenlandic Esk I.

X.2. Ale = Aleut, dls.: Ale E = Eastern Ale, Ale W =Western Ale (incl. Ale A = Atkan, Ale Au = Attuan