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, ***ṭu¨** (> ***ṭi**) ~ ***s'u¨** (> ***s'i**) 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 roots such as ***ʔˤaˑˑˈsˈˤoˈ** 'stay' (> 'be') preserved in IE (*es-), HS, U and K, ***ʔitE** 'to eat' (IE, HS, M), ***baˤʔˈri** 'to hold, take' (all branches except U), ***wete** 'water' (all branches except K), ***nimʔ▽** 'name, to name' (IE, HS, U, A), as well as words connected with culture of the final paleolythic age (cf. AD NM), such as ***ka¨luˤu¨** 'woman of another moiety' > words for 'daughter-in-law', 'sister-in-law' and 'bride' in IE (Latin glo⁻s, Greek γα′λωϳ, Slavic *zъ˝lǐ / *zъ˝lьve), 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 aproximants	Nasals	Lateral sonants	Vibrants
Voiced	Voiceless	Emph.	Voiced	Voiceless				
b	p	p̣			w	m		
d	t	ṭ				n	l	
3	c	c̣	z	s				
3ˇ	cˇ	cˇ̣	zˇ	sˇ		nˋ = n̰	l̰	r
3′	c′	c′̣	z′	s′	y	n′	l′	r′
3ˆ	cˆ	cˆ̣	zˆ	sˆ				
g	k	ḳ				ŋ		
g	q	̣q	ɣ	χ				
				ħ (= ḥ)	ʕ			
	ʔ			h				

Symbols in the chart: affricates: ʒ = d͡z, c = t͡s, ʒˇ = d͡zˇ, cˇ = t͡sˇ; lateral obstruents: ʒ̂, ĉ, ĉ., ẑ, ŝ - lateralized ʒ, c, c., z, s; palatalized consonants: ʒ′, c′, c′., z′, s′, n′, l′, r′ = palatalized ʒ, c, c., z, s, n, l, r; l̤ and ǹ (= n̤) = cacuminal or retroflex n and l; uvular stops: g (voiced), q (voiceless), q̤ ("emphatic"); uvular fricatives: χ = Spanish j, ɣ = Arabic ·; epiglottal (pharyngeal) consonant: voiceless ħ (= h̤ = Arabic Ó), voiced ʕ (= Arabic Ø).

In proto-Nostratic, as it is reconstructed on the basis of exstant data, there are three series of stops and affricates: voiced (*d, *g, *ʒ, etc.), voiceless (*t, *k, *c, etc.), and "emphatic" (*t., *k., *c., 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 *t. = 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, "'" 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, "_U" means "before a labial vowel", "_E" 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 *n`, where a distinction is possible only if the phoneme is represented in Ostyak, so that in daughter languages where there are no *n|n`-roots common with Ostyak we cannot find formal froof of representation of N *n and N *n` separately, but only representation of unspecified *n|n`. In such cases we suppose (as a working hypotheses) that both phonemes (in the case described *n and *n`) are reflected in the same way, which is symbolized by "**". The letter "N" symbolizes an unspecified non-labial nasal consonant, "L" is an unspecified lateral sonorant. IE *G = *g|gʷ|gˆ, *Gʰ = *gʰ|gʰʷ|gˆʰ, *K = *k|kˆ|kʷ; M *G = *g|*g, *K = *k|q; *T = an unspecidied dental stop'; ⊥_/ 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).

N	S	Eg	B	K	IE	U	T	M	Tg	D
*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, *p⁻	*φ	*p	*p
*-p-	*p	f	*f	*p	*p, *b	*p, ?*w	*p, *∅	*b, *β >*ɣ	*p	*v
*p̣-	*p	p	*f	*p, *p̣	*p	*p	*h >*∅	*φ	*p	*p
*-p̣-	*p	p	*f	*p, *p̣	*p	*pp	*p, *∅	*b	*p	*pp
*d-	*d	d	*d	*d	*dʰ	*t	*ɟ	*d, _i/ *ʒ	*d	*t
*-d-	*d	d	*d	*d	*dʰ	*δ	*δ	*d	*d	*t̪/t̪,t̪,
*t-	*t	t	*t	*t	*d	*t	*t⁻	*d, _i/*ʒ	*d	*t
*-t-	*t	t	*t	*t	*d	*t	*t	*d	*d	*t̪/t̪,t̪,, *t/tt
*ţ-	*ṭ,*t	d	*ď	*ṭ	*t	*t	*t'	*t, –i/ *c	*t	*t
*-ţ-	*ṭ,*t	d,t	*ď,*t	*ṭ	*t	*tt	*t'	*t	*t	*tt/t
*g-	*g	g,ʒ	*g	*g	*gʰ, *gˆʰ, *gʷʰ	*k	*k⁻,*k'	*g,*g	*g	*k
*-g-	*g	g,ʒ	*g	*g	*gʰ, *gˆʰ, *gʷʰ	*ɣ	*g	*g,*g, *ɣ,*ɣ	*g	*:
*k-	*k	k,c	*k,*g?	*k	*g,*gˆ, *gʷ	*k	*k⁻	*k, *q	*k	*k
*-k-	*k	k,c		*k	*g,*gˆ, *gʷ	*k	*g,*k	*g,*g, *ɣ,*ɣ	*g	*k
*ķ-	*ḳ,*k	q,k	*ɣ	*ḳ	*k,*kˆ, *kʷ	*k	*k',*k⁻	*k,*q	*x	*k

*-ḳ-	*ḳ	q	*ɣ,*k	*ḳ	*k,*k̂, *kʷ	*kk	*k	*k,*q	*k	*kk
*g-	*ɣ	ʕ?		*ɣ	*x, *xʷ, [*x̂?]	*∅	*∅	*∅	*∅	*∅
*-g-	*ɣ	H		*ɣ	*X, ?*ɦ	*∅, ?*ɣ	*∅	*∅	*∅, ?*g	*∅
*q-	*χ	χ	*H	*q	*x, *xʷ, [*x̂?]	*∅	*∅	*∅	*∅	*∅
*-q-	*χ	χ	?*H	*q	*H	*∅	*∅	*∅,*g, ?*g	*∅, ?*g	*∅
*ạq-	*ḳ,*χ	q, χ	*ɣ	*ạq	*k,*k̂, *kʷ	*k	*kʻ,*k̄	*k,*q	*x	*k
*-ạq-	*ḳ	q	?*ɣ	*ạq	*k,*k̂, *kʷ	*k, *kk	*k	*k, *q	*k	*k, *kk
*ʒ-	*z	ʒ?		*ʒ`=*ʒ₁	*s	*s	*ɟ	*ʒ?	*ɟ	*c
*-ʒ-	*z	ʒ?		*ʒ`=*ʒ₁	*s	?*cʼ	*s??	*ʒ?	*ɟ?	*t̬?
*c-	*s	s?		*c`=*c₁	?*(s)K	*cʼ	*ɕ	*c?	*ɕ	*c
*-c-	*s	?c	?*s	*c`=*c₁	*s	*cʼ	*ɕ ?	*c?		
*c̣-	*c̣			*c̣`=*c̣₁	?*(s)K	*cʼ	*ɕ	*c	*ɕ	*c
*-c̣-	*c̣, *s	?ʒ		*c̣`=*c̣₁	*s	?*cʼ	*ɕ			*c
*s-	*š	s	*s	*s`=*s₁	*s	*s	*s	*s	*s	*c
*-s-	*š	s	?*s	*s`=*s₁	*s	*s	*s	*s	*S	*c
*z-	*z			*z`=*z₁	*H	*s	*ɟ	?*s	*s?	*c
¿*-z-		z?	?*z	*z`=*z₁	?*H	*s		?*s,*y		
*ʒʼ-	*z	z?		*ʒ	*s	*cʼ	*ɟ	*ʒ?	*ɟ?	
*-ʒʼ-	*z			*ʒ, *z	*s	*cʼ		*ʒ		*c
*cʼ-	*s	s?	*s	*c	*sK	*cʼ	*ɕ		*ɕ	*c
*-cʼ-	*s			*c	*s	*cʼ	*ɕ ?	*c?		*c
*cʼ̣-	*c̣			*c̣	*sK	*cʼ	*ɕ		*c	*c
*-cʼ̣-	*c̣, *s	?ʒ	?*s	*c̣	?*s	*cʼ(cʼ)	*ɕ	*c?		?*c(c)
*sʼ-	*š	s	*s	*s	*s	*sʼ	*s	*s	*s	*c
*-sʼ-	*š	s	*s	*s	*s	*sʼ	?*s	*s	*s	*c
*zʼ-	*z			*z	*H	*sʼ	*ɟ	?*s,*ʒ	*s	*c
*-zʼ-	*z	z?, sʼ?	*z?	*z	*H	*sʼ		*ʒ?		*c
*ǯ-	*δ	z?		*ǯ	*s	*cˇ,?*y ?*šˇ	*ɟ	*ʒ	*ɟ	*c,?*t̬
*-ǯ-	*δ	ʒ, d		*ǯ	*s,*d, *sd?	*δ	*δ	*ʒ,?*d	*ɟ?	
*čˇ-	*θ			*čˇ	*(s)t-	*cˇ	*ɕ ?*t̄	*c ??*d	*ɕ ?*ɟ	*c
*-čˇ-	*θ	?c	*s	*čˇ	*s	*cˇ	*ɕ	*c	*ɕ?	*c
*čˇ̣-	*θ̣			*čˇ̣	*(s)t	*cˇ	*ɕ	*c	*ɕ	*c
*-čˇ̣-	*θ̣	ʒ		*čˇ̣	*tʰ, *sT	*cˇ	*ɕ	*c	*ɕ	*c
*šˇ-	*š	s	?*s	*šˇ	*s	*šˇ	*s	*s	*s	*c
*-šˇ-	*š	s	*s	*šˇ	*s	*šˇ	*s	*s	*s	*c
*žˇ-	*š	? z		*z	*H	*šˇ	? *ɟ	*s	*ɟ	
*-žˇ-	*š,?*z	?? ʒ	*z	*ž,*z	*H	*šˇ			*ɟ ?	
*ǯˆ-	*ŝ	? s, sˇ		*ǯ`	*l, _Vl/*s-	*λ	? *ɟ	*ʒ	*ɟ ?	*c

¿*-ʒ^-			*s	*z`	*l	*z^	*l			*t̬,*t̬t̬
*c^-	*s^	sˇ	?*s	*c`	*s	*c'	*ɕ	*c	*ɕ	*c, ? *k'-
*-c^-	*s^	sˇ		?*c'	*s	*c'	*ɕ	*c	?*ɕ	*c
*c^._-	*s^.	? ʒ	?*z̬	*c`.	*s^	*c'	*ɕ	*c	*ɕ	*c, *k'-
*-c^._-	*s^.	ʒ		*c	*s	*c'	*ɕ	*c	*ɕ	*c
*s^-	*s^	sˇ	*sˇ ?	*s`=*s$_1$ (in *s`t, *s`i), *s	*s,*k̂s	*s^	*s	*s	*s	*c, ? *k'-
*-s^-	*s^	sˇ		?*sˇ	*s	*s^	*s	*s	*s, ?*l	*c
*z^-	*s^	sˇ	?*z	?*l	*l,?*s-	*l, *s^	*ɟ	*s	*s	*n
*-z^-	*s^	??n	?*s	*l	*l, _⊥/*s?	*z^	*l, ?*l'	*l	*l	*l, *l̬
*ɣ-	*ʕ	ʕ		*ɣ	*X	*∅	*∅	*∅	*∅	*∅
*-ɣ-	*ʕ	?ʕ	*H	?*ɣ, ⊥_/*χ	?*X, *∅	*∅, ?*ɣ	*∅	*∅, *g	*∅, *g	*∅, *:
*χ-	*ħ	ħ	*H	*χ	*X	*∅	*∅	*∅	*∅	*∅
*-χ-	*ħ	ħ	?*H|∅	*χ	*X	*∅	*∅,*:	*∅	*∅, *:	*∅
*ʕ-	*ʕ	ʕ	?*H|∅	*∅	*H	*∅	*∅	*∅	*∅	*∅
*-ʕ-	*ʕ	ʕ	?*H|∅	*∅	*H	*∅,*:	*∅,*:	*∅	*∅, *:	*∅
*ħ-	*ħ	ħ,h̲	*H	*∅	*H	?*∅	*∅	*∅		*∅
*-ħ-	*ħ	ħ	*H|∅	*∅	*H	*∅	*∅,*:	*∅		*∅
*h-	*h	?h		*∅	*X	*∅	*∅	*∅	*∅	*∅
*-h-	*h, *∅	h, ∅	?*∅	*∅	*X	*∅	*∅,*:	*∅	*∅	*∅,*:
*ʔ-	*ʔ	i̯, ʒ	*ʔ-,*H	*∅	*ʔ =*∅∅	*∅	*∅	*∅	*∅	*∅
*-ʔ-	*ʔ	ʒ,y,∅	*∅,*ʔ	*∅,*⊥.	*∅,*:	*ʔ =*∅	*∅,*:	*∅	*∅, *:	*∅
*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	**ɟ	*n	**n	*n
*-n-	*n	n	**n	*n	*n	*n	**n	*n	*n	*-n̪-, _t/*n
*n`-	*n	n	**n	**n	*n	*n`	*ɟ	*n	**n, ?? *ŋ	*n
*-n`-	*n	n	*n	**n	*n	*n`	*n	*n	**n	*-n`-, _t/*n, _#/*n̪
*n'-	*n	n		*n	*i̯	*n'	*ɟ(<*n')	*n, *ʒ	*n'	*n
*-n'-	*n			*n	*i̯,?*n	*n'	*n' >*y	?*n	*N	*-n'-
*ŋ-	*ʕ,*ɣ? _H/*n				*kn, *n	*n,?*–	*–,*ɟ	*–,*n	*ŋ	?*n
*-ŋ-	*n, *m	n	*n	*n	*n, *nGʰ	*ŋ	*ŋ	*ŋ, *ŋG, *ŋK, *ɣ,*ʁ	*ŋ	*ŋk
*l-	*l	? i̯	*l		*l	*l	*ɟ	*n	?*l	??*t

*-l-	*l	r ?,ʒ	*l	*l	*l	*l	*l	*l	*l	*l
*ḷ-	*l	?n	*l	*l	*l	*ḷ̬	*ɟ	*n	*l	*n
*-ḷ-	*l		*l	*l	*l	*ḷ̬	*l	*l	*l	*ḷ̬
*l'-	*l		? *l	*l	*l	*l'	? *ɟ	? *l, ? *n	? *l	*n', **n
*-l'-	*l	r, ʒ	*l	*l	*l	*l'	*l'	*l	*l	*ḷ̬
*r-	*r	r	?*r	*r	*r	*r	*ɟ	? *n	*l, *n	*n
*-r-	*r	r, ʒ	*r	*r	*r	*r	*r	*r	*r	*r̬; *r (< *r⊥)
*-r'-	*r	r, ʒ	*r	*r	*r	*r	*r'; _⊥/ *r	*r	*r	*r'
*w-	*w	w	*w	*w	*u̯	*w	*b,* °	*b	*b, ? * °	*v, _U/ *∅
*-w- /V_V	*w	w,ʒ	?*w	*w, *∅	*u̯	*w	*b,*∅	*β,*b	*b	*v
*-w- /⊥_V	*w, *∅	?? ∅, y	??*w	*w, *∅	*u̯,*∅	*w, *∅		*b,*∅	*b, *∅	*∅
*-w- /a,E _⊥	*w, *∅			*w	*u̯	*w,* °	?* °	*∅, * °	*∅	*∅, ? *v
*-w- /u_⊥	*:		?*:	*∅, *w	*:,*∅	*∅	*:	*∅		*:
*y-	*y	ï	?*y, *i	*∅, ?*y	*i̯,*ei̯	*y	*ɟ	*y	?*y	*y, * ¨
*-y- /V_V	*y	y, ∅	*y	*∅	*i̯/*i	*y	*y	*y	*y, *∅	*y, *∅
*-y- /⊥_V	*y,*∅	? ∅	*∅	*∅	*∅,* i̯̯, +*(s)-	*y,*⊥ʸ	*⊥ʸ, *∅,* ¨	*∅,*y	*∅, *⊥ʸ, *i̯	*∅, *⊥'
*-y- /V_⊥	*y,*∅	y, ∅	?*y	*∅	*i̯,*∅	*y,*⊥ʸ	*∅, *⊥ʸ	*∅,*y	*∅, *i̯	*∅,*y

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.

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

The formulae ***d-** > M *d, _i/ *ʒ, ***t-** > M *d, _i/ *ʒ and ***ṭ-** > M *t, _i/ *c reflect the pM affricatization *di > *ʒi and *ti > *ci.

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

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

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.

§ 2.2. 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 r-consonants. According to the extant data, D *-r'- (= *-r̤.- of the traditional notation, retroflex

vibrant/fricative) goes back to N ***-r′-** or ***-ry-**, D *-r̥- (= *-r̥- or *-t̥- 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.

§ **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- ↔ sˇ- in Manchu:

N *s′-, *sˇ- and *s- yield pTg *s- > LMc s-, as well as s- in other Tg languages (but h- in Lm). Cf. N ***s′ilKⱽ** 'to let out' > LMc silgi- ~ silki- 'get through (a narrow opeming), crawl (kriechen)', N ***s′u˙ˡ|z̧ˆu¨(-kE)¬*s′il|z̧ˆi(-kE)** 'mucus, slime, saliva' > LMc sileŋgi 'saliva, drivel', N ***s′|sˆeHmˌu˙ˡ** 'to swallow' > LMc simi- 'to suck, swallow', N ***s′amˌⱽˌge** 'hair, fine hair' > Tg *sen′ŋe(n) 'beard, fin, gill' > LMc seŋele 'cock's comb, gill', N ***sˇubyⱽ** 'spike, spear, to pierce' > LMc suyfun 'awl', N ***sˇu˙¨ŋU** 'snow' > LMc su(ŋ)- v. 'become covered with hoar-frost', N ***sˇiŋⱽrⱽ** 'mouse' > LMc siŋgeri id., N ***sˇehr′E** 'be awake, watch (over), feel, notice' > LMc sere- 'be awake, feel, guess, understand', N ***saPˌu˙ˡ-s′ⱽ** 'thorn, → 'needle' > ClMc sabsi- 'stitch', [N ***sonˌⱽqˌu˙ˡ** 'sinew, tendon' > Mc suna ~ su¯na 'rein of draught-dogs', N ***sa¨ru** 'sinew, fibre' > ClMc siren id., 'bow-string'.

But sometimes we find pTg *sˇ- > LMc sˇ-, as well as probably ҫ-, cˇ-. c′- in other Tg languages. It is likely to go back to *si̬- (that appears due to contraction of syllables: N *Si̬|u¨|E...ⱽ > *si̬ⱽ. For instance, N ***sˇu˙ˡwolⱽ** 'liquid, moisture' > LMc sˇula 'juice', N ***s′EXarⱽ** 'bright, daybreak' > *si̬ˌaₗ:ˌrⱽ > LMc sˇari 'white, clean', N ***sˇi̬ˌh˙ˡwⱽ** 'sun' > LMc sˇun 'sun' [÷ Pre-Classical Manchu siyun, Ul siu(n-), Nn siu~ ᵈ siu(n-), Neg siyun ~ siwun.

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

§ **2.2.4.** Another tentative hypothesis concerns a curious pD phoneme, reconstructed by G. Starostin as *k′- (GS RVP) and later as *ky- (GS 203-7). This is a phoneme that yields *c- (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 *c- precding *u, but later gave up this hypothesis (due to counter-examples). Pfeifer (Pf. 66) supposed that this is a cluster *ky- resulting from reduction of earlier *key- or *kiy-. In the light of external comparison I am inclined to suppose that pD *k′- goes back to N lateral obstruents (see N ***sˆEØrⱽ** 'reach, approach, enter', N ***cˆˌˌuˡPˌrⱽ** 'to twist, plait' and N ***cˆˌAØⱽ** 'perish').

§ **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 roots with an inherited internal palatal element (*y, *n′, *l′, and the like). My interpretation of his hyp. is as follows. The N emphatic consonants in root-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 *h′ that later yielded a movable *s-.

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 s&, S, ҫ: [is&ki // iSki // içki] (i.e. isˇki ~ im̟ki) 'two' (< iʰti), [iˢpax] 'silk thread', [iSt // içt] (i.e. isˇt ~ im̟t) 'dog', [pis&ti-//piSte-//piçti-] (i.e. pisˇti ← pim̟ti) 'write' (cf. Tn. SJ 77). In his formulation, the phoneme i had an allophone isˇ (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 *‘- in pIE and pN reconstructions. We shall call this *‘- "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 *‘- this laryngoid underwent palatalization (*‘- > *ˁʹ-) and became a movable *s- (e.g. NaIE *(s)ne⁻bʰ- 'narrow, thin' < *ˁʹne⁻i̯_b∇ < N **‘NAXiba**).

§ **2.2.7.** In several 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 labial vw. > U *w + non-labial vw.):

N ***goK̦∇sˆ|cˆ∇(-R∇)** '∈ canine' > FU (in ObU only) *ᵒʻwˋoksˆ∇r∇ 'fox'

N ***ØuK̦a** 'to see', 'eye' > FU (in Ugr only) *wok∇ (or *woɣ∇) v. 'see, look'

N ***Xomusʹ|sˆ∇** [or ***XomU(-sʹ|sˆ∇)**] 'articulation between a limb and the trunk', 'bones of the shoulder' > U *wamsʹ∇ 'rump, crupper, shoulder'

N ***Xᴦuˮ⁺ⁿC∇** 'thin, narrow' > FU *ᵒwinʹcʹa (~ *ᵒwicʹ∇?) 'thin, narrow'

N ***qur∇** 'to pierce, make a hole; hole, pit' > FU (att. in ObU) *ᵒwur∇y∇ > ObU *wu⁻rзy 'river-bed'

N ***Øulˁe** 'recent' (> new') > U {UEW} *wuzˆe '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).

§ **2.3. Structure of words (roots).** The roots (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	*u¨
	*e			*o	
		*a¨	*a		

The original vowels of the first syllable survive in proto-Uralic, partially in proto-Dravidian (where both ***a** and ***a¨** 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 (passive and non-active verbs) turned into Semitic *u ~ *u⁻ 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 *g, *k, *ḳ and * q̣, when followed by N *o, yield IE *gʷʰ, *gʷ and *kʷ; if followed by N *e and *a̤, 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 *ṳ 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).

§ 2.4.1. I can propose a tentative hypothesis about the prosodic origin of pT ascending diphthongs (in many, but not in all cases). These diphthongs are reconstructed on the ev. of the Chv reflexes of initial cnss.: Chv ј- ÷ NaT *k'- & *k'- < pT *k⁻i̯- & *k'i̯-; Chv c̆- ÷ NaT *t⁻- & *t'- < pT *t⁻i̯- & *t'i̯-; Chv s̆- ÷ 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 *ṭṳHr'∇ 'shin, knee' > pre-T *t'ṳˇːr' > **t'Eṳ⁻r' > pT *t'i̯ṳ⁻r' > NaT *t'ṳ⁻z and Chv чеӒр с̆ъr 'knee', N *ḲalØa 'to throw, leave' > pre-T *k'aˇːl > **k'Eal > T *k⁻i̯a⁻l- > NaT *k⁻al- and Chv jul-; N *s'ihar'u 'dirt' > pre-T *sEar' > T *si̯a⁻r' > NaT *sa⁻z 'swamp' and Chv s̆ur '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 (***ʔitE 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.

Where was the place of the (non-pronominal) subject? The only place left for it (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 ***ṭʼṳʔ ~ *sʼʼṳʔ** 'thou' in the direct case, other pronouns in oblique cases (***ноу∇** 'by me, my', ***k∇ ~ *g∇** 'thee, thy'), as well as pronominals [i.e. lexical words replacing the pronouns, e.g. ***ʔʼoʔkE** 'self' functioning as a lexical replacement for ***mi** 'I', whence IE *eĝ̂oH ~ *eĝ̂H-, Semitic *-a⁻ku, etc.], pers. pronouns of 1 pl. excl. (***n⁻∇, *gʼuʔ**), as well compound pronouns: ***mi ʔa** 'we', ***mi ʔa** 'ye' (with the plurality marker ***ʔa**),

[2] interrogative pronuns: ***Ḳo** 'who?', ***mi** 'what?', ***ya** 'which?'

[3] deictic particles ***ha, *hʼeʔ, *ʼhʼi, *ʼhʼu** 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] and ***ṭa̤** 'it' for inanimate, etc., for collectivity (***ʔaʼhʼa, *ḷA, * q̣∇**), collectivity-plurality (***n⁻ʼaʔ, * q̣∇**), plurality [***t∇** of plurality, ***ʼʔʼ∇sʼ∇** 'they' and ***yE** (= yʼiʔ?) 'these, they' for animate beings, as well as ***ʔ∇q∇** 'thing(s)' and prn.\n. of plurality], duality (***nʼiʔ** 'they [two]', ***ʼhʼæ ~ *ʼhʼu** for animate beings, ***yi** for inanimate objects), individualization (***ya** 'that which, belonging to' [← ***ya** 'which?'], ***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 pronuns, 3rd person from demonstratives), pronominal possessive suffixes, markers of the nominative case (e.g., IE nominative *-s [for nouns of the active gender] from the demonstrative active ***sE**), affixes of plural, dual and collectivity. In some languages (IE, S, K) the genitive case, too, is based on pronouns (e.g., IE *-oi‿os, K *-is < ***ya sE**, originally 'that which X').

C. Auxiliary words: [1] postpositions (in many cases functioning also as preverbs): ***nu** 'of, from', ***mA** - formative of marked accusative, ***ḳV** 'towards, to', ***sʹV** 'to, towards', ***t⌐aʺ⌐** 'from', as well as lexical words transformed into postpositions/preverbs: ***ʔin⌐ᴬ⌐** 'place' (→ 'in'), ***dιoyιa** 'place (within, below)' (→locative particle *da 'in'); [2] auxiliary words of other meanings: ***dˤi**, marker of imperfective (< an auxiliary verb?), ***nˈi** 'not', etc.

§ 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 of them 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 ***n⌐aʺ⌐** 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] 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 (bᴇ, postposition of the direct object, cf. Hrl. 35, 74-5) and Japanese (OJ wɔ > J o). On the analytical status of J o (< *ma), no (< N *nu) cf. Vrd.JG 278-82. The element ***nˈaʺ** 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 nꜣ abstract 'this' and 'these (things') ↔ pꜣ 'this' m. ↔ tꜣ 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⌐aˈwV** 'head' (preserved with this neaninf 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).

§ 5. Derivation. Less clear is the original status of the N etyma underlying derivational affixes of the daughter-languages. For some of them the analytic origin is obvious. Thus, the etymon ***mA** that underlies affixes of nomina actionis and nomina actoris 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; 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 *ko¨⌐r-be 'mail reindeer', Manchu oŋgo-ba 'forgetful') is interpreted as analytic on the evidence of its phonetic behaviour: the regular reflex of the 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 that differ by their final (usually third) consonant. These "root extensions" need serious investigation.

  § 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 a branch. (2) recently several scholars have expressed another 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 root, 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 *√tk̥ (cf. here s.v. N ***t̥oka** '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. "Eursiatic cognates" (Ruhlen OLs 16-7) and their cognates

in Hamito-Semitic, Kartvelian and Dravidian

Meaning	"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. *saˋn		*si-	
pronoun base (actually 'I','ego')	IE *egˆʰo-m 'ego'	S *-a̅ku 'I' & cognates in B, Eg.		
who?	IE *kʷo-, Ur. *ku, etc.	preserved in Om, Beja and Ch, but replaced by *m- 'whst?' elsewhere		replaced by *y∇ 'which?'
what?	Ur. *mi, Chv. mən, etc.	*m-	*min 'who? (< N *mi 'what?')	replaced by *y∇ 'which?'
which?	IE *yo-/*ye-, Ur. *yo-, etc.	? S *ʔayy-		*y∇
this	IE *kˆ-, 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. *a̤la 'do not'	S *ʔal 'do not'		*all- 'not to be'
plural	Ur., Turk. *-t, etc.	S *-a̅t pl.	*-ta	
two	IE *dwo̅-, etc.	S *tuʔm- 'twin'	*t̠qu-m- 'twin', t̠qu-cˇ-̣ '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 *ka̅p-, 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. **da̤l∇kʼ- > *ɟa̤la̤kʼ-	Glavda (Ch.) dla̅kʷaˋ	*bur-t̠ql̥-	
star	IE *Haste̅r	*ʕaθtar- 'Venus'		
moon	Korean tal (-l < *-r)	Ch. *√tr		
fish	Ur. *kala, Tung. *xol-sa, etc.	Ch. √klp	Svan ḳalmaχ	*kol(l)-
wolf	Ur. *loka 'fox', Mong. *noqa 'dog'		*leḳw- 'dog'	*nakka 'jackal, fox'
elder brother	Turkic *a̅ka, etc.	S *ʔaχ- < **ʔaq-		
edge	Ur. *ka̤cʼa̤, etc.	S *ḳic̣c̣-	Svan ḳa̤cχ	
wet	Ur. *nʼo̅re, etc.	?S *√nhr 'river'		*ni̅r 'water'
dark	[1] Ur. *pol∇, etc.			*pul(l)- 'brown'
	[2] FU *rṳm∇	Ch. *rim-	*rum-	
speak	IE *kel-, etc.	Arb qa̅la 'say', etc.	Sv. ̣qul- 'say'	
sleep	Ur *uni-, etc.	S *ʕu̅n-		
eat	IE *ed-, Mong.ide-	Cush. *ʔit-, Ron *ʔet-		
arrive	FUr. *tule	S *√dχl, Ch. *d̠∇l		
take, grasp	IE *kap-, etc.	Cush. *ḳab-		*kap(p)-
wash	Ht. arra-	Arab. √ʔry 'pour'		? *ur-
wash		S *rh̠sˆ̣	*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 *kʷasi̠o- 'basket' with U *kocʼa-~*kucʼa- 'drinking vessel' and OJ kasipa 'container for food\drink'. But the same root is

found in Semitic *'kaʔas- 'vessel', Berber *kʳuˉˡss- 'pot, drinking vessel', Cushitic (Xamir kuˇskuˇsaˉ 'Wasserkrug'), Kartvelian (Georgian k̟vac̟ia 'small earthern pot') and South Dravidian *kuⱼoc-a- (+ suffixes) 'potter' [see my entry #993 (**k̟o?Vc′V** '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 **t̟ʳuˉˡ** '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 **ʔaˆla**) 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 book is entitled "Indo-European and its Closest Relatives". To judge by these examples, Hamito-Semitic is closer to IE than IE itself! Greenberg's book actually proves that in this respect 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 vocubulary"). 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 **ʔeyʳoˡ** 'to 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 **ʔitE** 'to 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 roots and see what happened to them in HS and in Semitic: N **ʔeyʳoˡ** '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 *ʔaˉ in Arabic and *ʔaˉ in Hebrew (because Semitic verbal roots were devocalized and the intervocalic *-y- was lost); N **ʔitE** 'eat' survives in HS (namely in Cushitic and Chadic), but because of the devocalization of verbal roots it was lost in in Semitic (otherwise it would have been undistinguishable from other verbs with the same historical consonants, such as **ʔatⱼ?ⱼV** 'to 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.

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 clearly distinguishable from Proto-Slavic or Proto-Indo-Iranian). But among the 2800 N roots registered in this "Nostratic Dictionary" most (not less than 1600) do appear in HS. 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 os acceptable, I use the symbol " " before the siglum of the source. In quoting DQA, I use the abbreviation "incl." to indicate the acceptable comparisons between sub-branches. For instance, in the entry *ˡˈₗEˈgˈₗume 'wet\cold weather, dew' (→ 'wet snowfall' →'snow') I quote DQA #1232 in the following way: " DQA #1232 [A *l′uˉn′i; incl. Tg, Ko]", which means that I agree with the comparison (found in DQA) between NTg *luˉn′e 'wet snow' and pKo {S} *nuˉn 'snow', but not with another comparison in the same entry of DQA, namely that with the erroneously reconstructed M ïduˉn in HlM д*н(г) х*йгүн 'extreme cold' - in fact, literally 'full cold' with duˉ ŋ '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. I hope to include part of my critical remarks in my review articles or other papers.

§ **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 are very helpful in improving the quality and the exact formulation of the etymologies in the Nostratic Dictionary in preparation. 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 *NM* 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 NM.** The book "The Nostratic Marofamily and Linguistic Palaeontology" 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 **ʔitE** 'to 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 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. I am ready to accept the idea, but with one reservation: the concept "morphology" must include 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. bharā-mi - 2 sg. bharasi - 3 sg. bharati and Greek 1 sg. δι′δομι - 2 sg. δι′δο∅ - 3 sg. δι′δοσι. 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 yerexe-b 'I shall come', Kalmuck yoβ-na-β 'I go', Dagur icˇim-b′e¯ 'ich fahre, werde fahren', Moghol ra˙-na˙n-bi 'I come, am coming' with -b, -β, -b′e¯ and -bi < proto-Mongolic *bi 'I'; Kalmuck ɢarβ-cˇ, Buryat ɢarba-sˇ 'you [sg.] went out', Moghol ira˙n-cˇi 'you come', Dagur yawbei˛-sˇi 'you will go' with -cˇ, -sˇ, -cˇi and -sˇi < Mongolic *ci '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 *ḲERV̄ 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 *q̇"). In the entry in question the Kartvelian reflex is unknown, therefore we use *Ḳ. The unspecified R means "*r or *r′" (and not "all kinds of r-sounds", as Comrie erroneously believes). The distinction between the reflexes of *r or *r′ has

survived in Turkic and Dravidian only. If the word (as *Ḳ̣ERV) is not attested in Turkic and Dravidian, we have to use the capital letter *R. The symbol *E is used here instead of *eä̤ 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 *ä̤. Here I admit that it would have been more accurate to symbolize the reconstruction as *Ḳ̣eä̤RV (in order to rule out *i and *ṳ̈). The symbol *V (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 root.

§ **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' > сальнаɪ ульбка). But for any speaker of Russian (including those knowing French, like myself) сальный in both meanings is the same word. If in Russian we hear сальнаɪ ульбка (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 Russian (probably due to the translation of the English novel "Coolie" by the Indian writer Mulk Raj Anand), and in the famous song "От краɪ до краɪ" ("From border to border", by the poet Lebedev-Kumach) there are words: Пьёт ɪту песнɪ и рикши и кули, поеɪт ɪту песнɪ 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-o̤ 'I take', Armenian berem 'I carry, bring') and 'give birth to' (Gothic baɪran, English bear 'to give birth to', Albanian mberat 'pregnant'). It goes back to two or three different Nostratic words: (1) ***baˈʔˈri** 'to hold, take' [> Mongolian bari- 'hold'], (2) ***berEʔa** 'to give birth to; child' [> Dravidian *peṛ- v. 'beget, bear (a young)'], as well as possibly to (3) ***bä̤ˈrʔV** 'to give' [> Turkic beˉr- 'give', proto-Tamil *paric- 'gift']. In IE, due to the apophony, the vocalic distinction between roots with *a, *ä̤ 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 roots merged into one. In many Indo-European languages the root preserved the original meanings as polysemic variants (such as Gothic baɪran 'carry, bring, give birth', Old Irish breth 'fait de porter/emporter, fait de porter un enfant').

Dravidian *civ∇ŋki 'leopard' (or sim.) goes back to N **ʒˈiw∇mʟ∇ɟgE** 'leopard', but N **ʦˆˌɾˀⁱbⱽɣⱽ** 'hyena' merged with it (because in Dravidian in the word-initial position voiceless and emphatic affricates coalesced, and so did the intervocalic *-w- and *-b-), and as a result Dravidian *civ∇ŋki 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 desaster to etymological research. One would have to reject all Gothic reflexes of IE words unless they are found in other Germanic languages, or all Lithuanian reflexes of IE words and roots unless they are represented in Latvian and Prussian.

The Nostratic etymon *ka¨luǀu¨ '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 **ka¨luǀu¨** in the present dictionary).

The Nostratic word *qant̟∇ '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 then. But according to the laws of Nostratic comparative phonology (discovered by Illich-Svitych) the Semitic reflex has to be *χant̟-. 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 χanṭi 'front, front part of anything' (Jo. J 303). Both the sound and the meaning of the word corresponds 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 this extremely important 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? 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 is here rather high. With Turkic the situation is different: Turkic preserves much of the phonological information of proto-Altaic, so that proto-Turkic *toɫu 'hail' is a legitimate cognate of IE *del- 'rain, dew' and probably of FU *ta¨lwa¨ '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 cognate 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 *koˉṭ·ο/*koṭ·ṭ·- 'end, summit, top', IE *kʷe(:)d-/*kʷo(:)d- 'sharp point', Semitic *°√ḳ·twǀy ~ *ḳ·utt·- > Geez kˇʷəṭ·ṭ·, kˇʷəṭ·ṭ·a '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 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 teˇmpus, except for the final -us) are more important than those with little information (as French [tɑ˜] spelled as temps). Words with loss of phonological information may go back to different alternative etymons (as French [tɑ˜] 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 extremely 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), undergoes lexical attraction, analogy, etc. - cf. English off and of, life [lai·f] and live [liv], wife [wai·f] and woman ['wu-mən] / pl. women ['wɪ-mɪn], French homme and on, Hebrew 'leb̠ 'heart' and leˉ'b̠aˉb̠ 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 *qoruβǀɣu '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 *ɟiʟKɭ-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 **dik̠∇** 'edible cereals or fruit' (reflected in K *dik̠- and in HS *d∇k̠·-), we have to expect in Tungusian *ɟiʟKɭ-kte > *ɟikte. 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 '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 ῞υπνοǀ, Old Indian svapnah̠, etc. In proto-Italic we reconstruct *pes-ni-s (> Latin peˉnis '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 πε'οǀ 'penis'. If a proto-language lost phonemes in certain environments (e.g. 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: *K.awingV 'arm-pit' (IS I 344), *p'aliHma 'palm of hand' (IS III 93-5) and probably *purcˇV(gV) ~ *pu¨lcˇV(gV) 'flea' (IS II 99-100). In the present dictionary trisyllabic etyma are numerous (as can be seen by anyone who skims its entries). In my opinion, trisyllabic 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 *dᵣoTᵗgiHU 'fish'). 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 *-ᵣoTᵗ- of that root (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. ***ʔ**), ***ʕ** (incl. ***Ø**), ***b**, ***c** (incl. ***c′, *c⁻, *c', *C**), ***c.** (incl. ***c.″, *c⁻., *c.', C.**), ***cˇ, *cˇ., *cˆ, *cˆ., *g, *g, *ɣ** (incl. ***Γ**), ***h, *H** (incl. ***ʜ, *.**), ***ħ, *k** (incl. ***K, *K.**), ***k.** (incl. ***K.**), ***l** (incl. ***ḻ, *l⁻, *L**), ***l′, *m, *n** (incl. ***nˋ, *n⁻, *ɴ, *N**), ***n′, *ŋ, *p** (incl. ***p., *P, *P.**), ***q** (incl. ***Q**), ***.q, *r** (incl. ***R**), ***r′, *s** (incl. ***s′, *s⁻, *s′, *s', *S**), ***sˇ, *sˆ, *t** (incl. ***T**), ***t., *w, *χ** (incl. ***X**), ***y, *z** (incl. ***z′, *z⁻, *z', Z**), ***zˇ, *zˆ, *ʒ** (incl. ***ʒ′, *ʒ⁻, *ʒ', *ʒ**), ***ʒˇ, *ʒˆ**.

Among the word-medial consonants we distinguish between "weak consonants" (laryngeal, epiglottal, ***g**, ***q**, ***w** and ***y**) and "strong consonants" (all others).

The entries are arranged as follows:

[1] The basic arrangement: according the initial consonants and the first word-medial strong consonants. If there are no strong Inlaut-consonants, the weak consonants (according to their place in the alphabet) are taken into account instead.

[2] Within each group of entries with the same initial cns. and the same medial strong consonants the entries are arranged as follows:

 (a) first those without any additional consonants,
 (b) those with a weak cns. preceding the word-medial strong one,
 (c) those with a weak cns. following the word-medial strong one,
 (d) those with a third strong cns.

If in a N etymon in the medial position (Inlaut) there are two weak consonants and no strong one, the priority of one weak consonant over the other (as criteria for alphabetical oder) has been established according to the fate of these consonants in daughter languages. For instance, in the root ***k.AħwV** 'to seize, grasp, hold' the consonant *w has survived in Kartvelian and Dravidian, while the laryngeal was lost, therefore the entry was placed among the *k.-w-entries.

[3] Within each group of entries with identical consonants the entries are arranged according to the first vowel. Order of vowels: ***a** (incl. ***A**), ***a¨** (incl. ***æ**), ***e** (incl. ***E**), ***i, *o, *u** (incl. ***U, *X, *Y**), ***u¨, *V**.

The entries with consonants of doubtful identity (with the brackets ᵀ ᵗ) and of doubtful presence (those with the brackets ʟ ⌐) are treated as those with regular consonants. The optional second element of a N compound word [that within plain brackets ( )] is not taken into account in the order of the entries.

§ **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 (rather than to those quoted from other scholars).

If in same entry there are several cross-references to the same Nostratic root, 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 a reconstruction of descending proto-languages is followed by a name of a branch [e.g.: "S: CS", "FU (in FP)", "D (in SD)", etc.: FU (FP) *koˉčʼe v. 'crawl, clime, run', **D** (SD) *tot˳t˳° 'point, nipple'), it means that the word is attested in one branch of their family only, but it is reconstructed on the pS, pFU and pD level (using formulas of sound changes for the respective family 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 *ḳ˳aɗħ- ~ *ḳ˳uɗħ- 'thorn' is not reconstructed on a pHS or a pC level, but rather on the East Cushitic one.

§ 12. 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 banaˉ '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). It is the only reason to prefer š to the analytical symbol sˇ, 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 čˇ IPA uses either the digraph tš (which is misleading, because čˇ is one consonant rather a consonant cluster and because in many languages [such as Russian and Polish] there is phonemic opposition čˇ ↔ tsˇ) or the clumsy sign ǰ, instead of the generally understood čˇ (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:

	Fricatives		Affricates		
	Voiced	Voice-less	Voiced	Voice-less	Glottal-ized
Hissing	z	s	ʒ	c	c̣

Hissing-hushing (like in Kartvelian)	z`	s`	ʒ`	c`	c`.
Palatal	z′	s′	ʒ′	c′	c′.
Hushing	zˇ	sˇ	ʒˇ	cˇ	cˇ.
Lateral	zˆ	sˆ	ʒˆ	cˆ	cˇ.

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, ù, è, Z, à, À, J, S, ò, ì, Ó, ¬, 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 ʬ. But how shall one denote the corresponding voiced sibilant (as in Jibbali)? In my system ʬ is denoted as s', and its voiced counterpart as z', without necessity of any special explanation of the symbol z'. If necessary, the corresponding affricates will be naturally denoted as c' and ʒ'.

[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 sˇ, or an apico-alveolar s˙ (like in New Greek). What shall we do if one language has different dialectal variants? How shall we write the Arabic phoneme ô? In Cairo it is pronounced [g], in Bedouin and Iraqi Arabic [ò], in Urban Syro-Palestinian and Maghrebine Arabic [J], in Sudanese Arabic as palatal [ʃ], etc. (to use the IPA transcription). In this particular case of ô I have chosen to use a special super-dialectal transliteration symbol gˇ. 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], [S], [à], [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, ω, ʌ). *Feci quod potui, faciant meliora potentes.*

§ 12.1. 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.

§ 13. 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, Zaja.czkowski, Szapta⌐ (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 usual practice of quoting by the author's name and abbreviation of the periodical).

   § **14. 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), unlike other scholars (e.g. Pokorny) who 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 use the unconventional term Old Upper German ("altoberdeutsch"). Let us hope that these problems will not jeopardize the understanding of the etymology and 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).