ISSN - 0132-1447

GEORGIAN NATIONAL ACADEMY OF SCIENCES

საქართველოს მეცნიერებათა ეროვნული აკადემია

Bulletin

მოამბე

ტ.I (=ტ.175), № 3
ივლისი - აგვისტო - სექტემბერი
2007
New Series
vol. I (=vol. 175), no. 3
JULY - AUGUST - SEPTEMBER

GEORGIAN ACADEMY PRESS
TBILISI

«საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე»

«BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES»

Scientific Journal Issued Quarterly in English & Georgian Languages

Founded in 1940

Bulletin of the Georgian National Academy of Sciences presents the Proceedings of the Georgian Academy in different fields of Science and Technology: Mathematical & Physical Sciences, Geology & Chemical Sciences, Biological & Medical Sciences, Engineering & Applied Sciences, Agricultural & Environmental Sciences, Humanities & Social Sciences, etc.

TABLE OF CONTENTS

შინაარსი

Biological Sciences
ბიოლოგიური მეცნიერებანი

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Mathematics

Ultra-Newtonian Gravitation

Vladimir Boltyanski*

* *CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS A.C.*

(Presented by Academy Member G. Kharatishvili)

ABSTRACT. In the article a theory of the gravitation is given that contradicts the Einsteinian general relativity theory. The presented theory is based on the *postulate of flowing space* that was formulated by the author. The postulate means that the light velocities in direction to the central mass and in the opposite direction have the values which do not coincide. From the postulate of flowing space we deduce (very easily) Schwarzschild's metric. Furthermore, the gravitational redshift formula is proved that is postulated (without any proof) in the general relativity theory. Moreover, it is shown that a gravitational blueshift is possible. Some experiments are described two of which can discover the blueshift, whereas the third one can resolve the discussion between the Einsteinian gravitation law and the postulate of flowing space. Finally, the gravitational redshift of distant objects of the universe is explained without Hubble's hypothesis on "expanding universe". At the end of the article it is shown that the presented theory can be justified by consideration of a flux of some particles (in contrast to the general relativity theory that has, because of influence of Hilbert, purely geometrical, non-physical character). © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *gravitation, Einsteinian law, redshift, universe, Hubble hypothesis, light velocity.*

1. Introduction.

In 1915 Einstein published an article on the *general relativity theory.* It contained a geometrical picture of the gravitation more exact than Newtonian gravitation law.

The following visual explanation ("mental experiment" by Feynman's expression) shows the initial Einsteinian idea. If we put a heavy ball M on a strongly tight piece of a cloth, then we obtain a deepening on the cloth. A light marble placed on the cloth rolls to M. Even if we push the marble perpendicularly to the direction going to M, then it moves around M until (because of the friction and decrease of the velocity) it falls to M. Abstracting from the friction, we may say that the curvature of the cloth creates the rotation of the marble around M. At the end of XIX century physicists said on an elastic medium ("ether") filling the empty space in which the light spreads as the sound spreads in the air. Einstein supposed that the mass M bends the medium, and this explains the moving of the planets (although it is unclear how the mass can bend the surrounding medium).

Einstein turned to the well-known mathematician Hilbert with the question, how is it possible to describe mathematically the curvature of the space? Hilbert sent to Einstein several letters in which he explained what is Riemann metric of a curved space, what are Christoffel's symbols, and what is the curvature tensor. Considering the Einsteinian physical idea, Hilbert understood that in this case it is necessary to use a convoluted curvature tensor $R_{\alpha\beta}$ named the *Ricci tensor.* The vanishing of that tensor gives the mathematical description of the Einsteinian idea.

In a time Einstein published the paper on the general relativity theory, under only his name, without mentioning of Hilbert. Now *Einsteinian law of gravitation* is expressed in a fine geometrical form: *In the empty space the Ricci tensor $R_{\alpha\beta}$ is identically equal to zero.* Empty space means that it does not contain any substance and any field except the gravitational one.

Mathematically elegant description of the general relativity theory is given in Dirac's book [5], where at first a compact description of Riemannian theory of curvature is given and then the Einsteinian law of gravitation with some conclusions is explained. Note only that instead of indication of the names of greatest mathematicians Gauss, Riemann, Hilbert who support the Einsteinian law the author writes that "some results are applied which were obtained mainly in the last century".

The components of the Ricci tensor $R_{\alpha\beta}$ are expressed by Christoffel's symbols $\Gamma^{\alpha}_{\mu\nu}$ and their first derivatives. Furthermore, the Christoffel's symbols are expressed by the first derivatives of the components of the metric tensor $g_{\alpha\beta}$ that defines the invariant spacetime interval $ds^2 = g_{\alpha\beta}dx^{\alpha}dx^{\beta}$. Here and in the sequel, as in the Riemann geometry, if the same index is twice encountered (once above and once below), then the summation over that index is made. The Greek indices take the values 0, 1, 2, 3 and the Latin indices take the values 1, 2, 3.

The equalities $R_{\alpha\beta} = 0$ form a complicated nonlinear system of differential equations of second order with partial derivatives. For a curved (non-planar) space the only solution of that system was obtained by the German physicist Schvarzschild. His solution is related to the case of a *static, spherically symmetric* gravitational field. The field is static if g_{00} does not depend on the time and, moreover, $g_{0m} = 0$ (see the deduction of the Schvarzschild's solution in section 18 of the book [5]).

The equalities $g_{0m} = 0$ enclosed in the definition of a static field mean that the value of the light velocity in two opposite directions coincide. Thus, to the Einsteinian postulate that the mass bends the enclosing space it is necessary to add the postulate which affirms that the values of the light velocities "there" and "back" coincide. Note that in all physical experiments the average velocity "there" and "back" is considered (or, as the physicists say, the light velocity "along the closed way").

Schvarzschild's metric implied some conclusions (the rotation of Mercouri's perihelion, the bend of light trajectories near the Sun) which are well confirmed by some observations. Thus, the Einsteinian gravitational law became famous.

2. The structure of the article.

In this article we outline an alternative theory named here ultra-Newtonian one. Some of its results are published in [1-4].

Consider a static Galilean system $\mathcal{G}$ that is referred to an orthonormal coordinate system x^1, x^2, x^3. Assume that at the origin of the system $\mathcal{G}$ a non-rotating, spherically symmetric mass M rests. By $t = x^0$ denote the Galilean time.

If under the influence of the Newtonian potential $\varphi = \dfrac{GM}{r}$ created by M a mass point moves from the infinity, where it had the vanishing initial velocity, then the mass point moves along the ray going to M. In the distance r from the origin it has the velocity $\bar{v}$ with $\|\bar{v}\| = \sqrt{2\varphi} = \sqrt{\dfrac{2GM}{r}}$. The corresponding acceleration is equal to $\|\bar{a}\| = \dfrac{d\|\bar{v}\|}{dr} = \dfrac{d\|\bar{v}\|}{dt}\cdot\|\bar{v}\| = \dfrac{2GM}{r^2}$ in the total conformity with the Newtonian gravitational law. Thus, if a neighborhood of the mass M is filled by mass points of negligible mass, then the whole space enclosing M as if "flows" to M, being subjected to the free fall. This is just the intuitive picture that replaces in our theory the Einsteinian intuitive picture on the mass M bending a strongly tight piece of a cloth.

Thus, instead of the Einsteinian postulate which affirms that the mass bends the enclosing space we introduce the postulate which affirms that the space enclosing the mass M as if "flows" to M with a velocity depending on the distance r from M. A small volume of the "flowing space" (a freely falling laboratory) is an inertial system in which the laws of the partial relativity theory hold. The exact statement of the "flowing space" postulate with the deduction of Schvarzschild's metric in the Galilean system are contained in section 3.

In section 4 Ricci's tensor in the system $\mathcal{G}$ is calculated. We show that $R_{00} = 0$, $R_{0p} = 0$, but $R_{pq} \neq 0$, contradicting the Einsteinian gravitation law.

In section 5 we describe the irradiation of an atom in the Galilean system $\mathcal{G}$ and the *variation of the light velocity* along the trajectory (again in the system $\mathcal{G}$). For example, if an atom rests in the surface of a star, then it is not immovable in the freely falling inertial laboratory, but is immovable in the corresponding Galilean system $\mathcal{G}$ that is not inertial. Hence the irradiation of that atom is subordinated to the laws which are distinct from ones of the relativity theory. Irradiation law in the Galilean system proved in section 5 implies, firstly, a mathematical deduction of the formula for the gravitational redshift (that is *postulated without any proof* in the general relativity theory) and, secondly, the principial

possibility of the gravitational *blueshift*. We describe some experiments which allow to discover that gravitational blueshift.

Note that the experiment of Paund-Rebka (1959) and the experiment of Paund-Snider (1965) which are based on using of Mossbauer's effect discover gravitational redshift under the influence of gravitational field of the Earth. It is possible that the converted experiments (which measure gravitational shift not in the direction to the center of the Earth, but in the opposite direction) also will discover a gravitational blueshift.

In section 6 an experiment is described the result of which (if positive) can resolve the discussion between the Einsteinian gravitational law and the postulate of "flowing space" in behalf of the last one.

Section 7 contains the description of Lorentz' transformations in the Galilean system. Nevertheless, the difference from their description in an inertial system is too small for an experimental verification (in a laboratory on the Earth).

In section 8 we show that the redshift of the distant objects of the universe can be explained in the frames of the presented ultra-Newtonian gravitational theory *without* using Hubble's hypothesis on "expanding universe".

Finally, in section 9 we show that, in contrast to the general relativity theory that has (because of the influence of Hilbert) pure geometrical, non-physical character, the presented ultra-Newtonian gravitation theory can be justified with consideration of a flux of some particles.

Thus, the laws of relativity theory in the Galilean, non-inertial system form the main contents of the article.

3. Locational time.

Instead of the Einsteinian postulate which affirms that "the mass bends the space" we introduce the following

Postulate of the flowing space. *The enclosing space of the mass M "flows" to M with a velocity* $\overline{v} = \frac{d\overline{x}}{dt}$ *whose value depends on the distance of the point* $\overline{x}$ *from M and is equal to zero at the infinity. The corresponding acceleration* $\overline{a} = \frac{d\overline{v}}{dt}$ *has a scalar potential* φ, *i.e.,* $\overline{a} = \operatorname{grad}\varphi$. *The potential* φ *satisfies the Laplace equation*

$$\left(\frac{\partial^2}{\left(\partial x^1\right)^2} + \frac{\partial^2}{\left(\partial x^2\right)^2} + \frac{\partial^2}{\left(\partial x^3\right)^2}\right)\varphi = 0. \tag{1}$$

Theorem 1. *The velocity of the "flowing space" has at the point* $x = (x^1, x^2, x^3)$ *the value* $\sqrt{2\varphi}$, *where* $\varphi = \frac{GM}{r}$ *is the Newtonian gravitational potential that is created by the mass M and is equal to zero at the infinity, i.e.,*

$$\overline{v} = -\sqrt{\frac{2GM}{r}}\,\overline{e}; \tag{2}$$

here r is distance of the point $\overline{x}$ *from the origin and* $\overline{e} = \frac{1}{r}\overline{x} = \left(\frac{x^1}{r}, \frac{x^2}{r}, \frac{x^3}{r}\right)$ *is the unit vector directed from M.*

Proof. Equation (1) has a unique spherically symmetric solution (up to a constant summand and a constant coefficient):

$$\varphi = \frac{GM}{r} + \text{const}.$$

Let $v(r)$ and $a(r)$ be the scalar functions with $\overline{v} = v(r)\overline{e}$ and $\overline{a} = a(r)\overline{e}$. Then $v(r) = \frac{dr}{dt}$, $a(r) = \frac{dv(r)}{dt} = \frac{d\varphi}{dr}$ and hence

$$d\left(v(r)^2\right) = 2v(r)\,dv(r) = 2v(r)\,a(r)\,dt = 2a(r)\,dr = 2d\varphi.$$

This implies

$$v(r)^2 = \frac{2GM}{r} + C.$$

The integration constant C is equal to zero, since the velocity $\upsilon(r)$ tends to zero when $r \to \infty$. Thus, $\upsilon(r)^2 = \dfrac{2GM}{r}$ and hence

$$\upsilon(r) = -\sqrt{\frac{2GM}{r}} \tag{3}$$

(the sign minus is taken, since, by the postulate of flowing space, the velocity is directed to the central mass M). ■

Make an important note. Let $\bar{x}$ be a point in the distance r from the origin. Denote by $\mathcal{G}_{\bar{x}} = \{dx^1_{\mathcal{G}}, dx^2_{\mathcal{G}}, dx^3_{\mathcal{G}}\}$ the Galilean system obtained from $\mathcal{G}$ by the translation of the origin to the point $\bar{x}$ and by $\mathcal{F}_{\bar{x}} = \{dx^1_{\mathcal{F}}, dx^2_{\mathcal{F}}, dx^3_{\mathcal{F}}\}$ the system that is fixed in the flowing space, has the origin at the point $\bar{x}$ (at the moment $t = 0$), and has the same direction of the axes as the system $\mathcal{G}_{\bar{x}}$. Using the notation $d\bar{x}_{\mathcal{G}} = \{dx^1_{\mathcal{G}}, dx^2_{\mathcal{G}}, dx^3_{\mathcal{G}}\}$ and $d\bar{x}_{\mathcal{F}} = \{dx^1_{\mathcal{F}}, dx^2_{\mathcal{F}}, dx^3_{\mathcal{F}}\}$, the passage from $\mathcal{G}_{\bar{x}}$ to $\mathcal{F}_{\bar{x}}$ is described by the formulas

$$d\bar{x}_{\mathcal{F}} = d\bar{x}_{\mathcal{G}} - \bar{\upsilon}dt_{\mathcal{G}}, \quad dt_{\mathcal{F}} = dt_{\mathcal{G}}. \tag{4}$$

If a mass point is immovable in the flowing space, i.e., has the velocity 0 with respect to $\mathcal{F}_{\bar{x}}$, then its velocity with respect to the Galilean system $\mathcal{G}_{\bar{x}}$ is equal to $\bar{\upsilon}$. Even if a mass point has a velocity $\bar{u}$ with respect to $\mathcal{F}_{\bar{x}}$, then its velocity in the Galilean system $\mathcal{G}_{\bar{x}}$ is equal to $\bar{u} + \bar{\upsilon}$.

Since the local system $\mathcal{F}_{\bar{x}}$ is inertial and the passage (4) is distinct from the Lorentz transformations, the Galilean system $\mathcal{G}_{\bar{x}}$ *is not inertial.* Consequently the value of the light velocity in the Galilean system is distinct from c.

A photon has (in the flowing space) its own velocity $\bar{c}$ of the value c. With respect to the Galilean system the photon has the velocity $\bar{c} + \bar{\upsilon}$, where $\bar{\upsilon}$ is described by formula (2). In particular, the photon going to M has the superlight velocity $c_1 = c + \upsilon$ and the photon going from M has the sublight velocity $c_2 = c - \upsilon$.

In the Galilean system $\mathcal{G}_{\bar{x}}$ all photons passing at a moment through the point $\bar{x}$ will be in a time $dt_{\mathcal{G}}$ situated in the surface of the sphere of radius $cdt_{\mathcal{G}}$ centered at the point $\bar{x} + \bar{\upsilon}dt_{\mathcal{G}}$ (Fig. 1). That sphere consists of all points $\bar{x} + d\bar{x}_{\mathcal{G}}$ which are situated in the distance $cdt_{\mathcal{G}}$ from the point $\bar{x} + \bar{\upsilon}dt_{\mathcal{G}}$, i.e., satisfy the condition $\left\|d\bar{x}_{\mathcal{G}} - \bar{\upsilon}dt_{\mathcal{G}}\right\| = cdt_{\mathcal{G}}$. In other words, for a given $dt_{\mathcal{G}}$ that sphere has in the Galilean system $\mathcal{G}_{\bar{x}}$ the equation $ds^2_{\mathcal{G}} = 0$, where $ds^2_{\mathcal{G}} = c^2(dt_{\mathcal{G}})^2 - \left\|d\bar{x}_{\mathcal{G}} - \bar{\upsilon}dt_{\mathcal{G}}\right\|^2$, i.e.,

$$ds^2_{\mathcal{G}} = (c^2 - \|\bar{\upsilon}\|^2)(dt_{\mathcal{G}})^2 + 2\langle\bar{\upsilon}, d\bar{x}_{\mathcal{G}}\rangle dt_{\mathcal{G}} - \left\|d\bar{x}_{\mathcal{G}}\right\|^2. \tag{5}$$

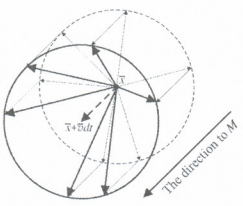


Fig. 1.

We call (5) the *main quadratic polynomial.* In the space of variables dx^0, dx^1, dx^2, dx^3 it defines a pseudo-Euclidean metric that is analogous to the pseudo-Euclidean metric of Minkowski. By virtue of (4), the main quadratic polynomial (5) can be rewritten in the system $\mathcal{F}_{\bar{x}}$ as $c^2(dt_{\mathcal{F}})^2 - \left\|dx_{\mathcal{F}}\right\|^2$, i.e., the main quadratic polynomial (5) represents the spacetime interval that is written however not in the inertial system $\mathcal{F}_{\bar{x}}$, but in the Galilean system $\mathcal{G}_{\bar{x}}$.

In the sequel, up to section 7, we will conduct reasoning only in the Galilean system $\mathcal{G}_{\bar{x}}$. Therefore we will omit (up to section 7) the index $\mathcal{G}$, i.e., we will use in $\mathcal{G}_{\bar{x}}$ the notation $d\bar{x} = (dx^1, dx^2, dx^3)$ instead of $d\bar{x}_{\mathcal{G}} = \{dx^1_{\mathcal{G}}, dx^2_{\mathcal{G}}, dx^3_{\mathcal{G}}\}$.

Let now $dt = dt_{(+)}$ be the Galilean time of spreading of a light signal from the point $\bar{x}$ to $\bar{x} + d\bar{x}$, and $dt_{(-)}$ be the Galilean time of spreading of a light signal from $\bar{x} + d\bar{x}$ back to $\bar{x}$. Then $dt_* = \frac{1}{2}\left(dt_{(+)} + dt_{(-)}\right)$ is the *locational* time of spreading of a light signal between $\bar{x}$ and $\bar{x} + d\bar{x}$ (the light signal emanates from $\bar{x}$ and then, "reflecting" at the point $\bar{x} + d\bar{x}$, returns to $\bar{x}$, Fig. 2). Dividing the distance $\|d\bar{x}\|$ between the points $\bar{x}$ and $\bar{x} + d\bar{x}$ by the locational time t_* we obtain the average light velocity "there" and "back", i.e., the velocity "along the closed way".

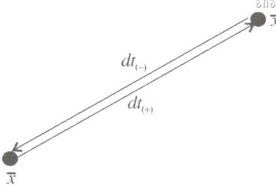


Fig. 2.

We observe the events in the locational time. Indeed, if we have only *one* clock (at the point $\bar{x}$), then we can find only the average time "there" - "back". Even if we wish to measure the light velocity *in one direction,* then it is necessary to use *two* clocks (at the points $\bar{x}$ and $\bar{x} + d\bar{x}$). In this case the clocks have to be *synchronized.* But to synchronize the clocks we need the light velocity in one direction. We obtain a vicious circle.

Theorem 2. *Under the passage from the Galilean time dt to the locational time dt_* the metric ds^2 described by the main quadratic polynomial turns to Schwarzschild's metric.*

Proof. By (5), the Galilean time $dt = dt_{(+)}$ of spreading of the signal from $\bar{x}$ to $\bar{x} + d\bar{x}$ is the *positive* solution of the equation $ds^2 = 0$. Furthermore, $dt_{(-)}$ is the positive solution of the equation obtained from the equation $ds^2 = 0$ by the substitution $-d\bar{x}$ instead of $d\bar{x}$. In other words, $-dt_{(-)}$ is the *negative* solution of the equation $ds^2 = 0$. Thus,

$$dt - dt_* = dt_{(+)} - dt_* = dt_{(+)} - \frac{dt_{(+)} + dt_{(-)}}{2} = \frac{dt_{(+)} + (-dt_{(-)})}{2}.$$

Applying Viet's Theorem on the solutions of a quadratic equation and using the expression (2) for the velocity *, we obtain from the equation $ds^2 = 0$ the equality

$$dt - dt_* = -\frac{\langle \bar{v}, d\bar{x} \rangle}{c^2 - \|\bar{v}\|^2},$$

and hence

$$dt_* = dt + \frac{\langle \bar{v}, d\bar{x} \rangle}{c^2 - \|\bar{v}\|^2}.$$

Now, extracting the perfect square in the right hand side of (5), we rewrite the main quadratic polynomial as

$$ds^2 = \left(c^2 - \|\bar{v}\|^2\right)\left(dt + \frac{\langle \bar{v}, d\bar{x} \rangle}{c^2 - \|\bar{v}\|^2}\right)^2 - \frac{\left(\langle \bar{v}, d\bar{x} \rangle\right)^2}{c^2 - \|\bar{v}\|^2} - \|d\bar{x}\|^2.$$

i.e.,

$$ds^2 = \left(c^2 - \|\bar{v}\|^2\right)dt_*^2 - \frac{\left(\langle \bar{v}, d\bar{x} \rangle\right)^2}{c^2 - \|\bar{v}\|^2} - \|d\bar{x}\|^2.$$

Using (2), we obtain

$$ds^2 = \left(c^2 - \frac{2GM}{r}\right)dt_*\, dt_* - \left(\|d\bar{x}\|^2 + \frac{\frac{2GM}{r^3}}{c^2 - \frac{2GM}{r}}\langle \bar{x}, d\bar{x} \rangle^2\right),$$

i.e., in coordinate form,

$$ds^2 = \left(c^2 - \frac{2GM}{r}\right) dt_*^{\,0}\, dt_*^{\,0} + \left(h_{pq} - \frac{\dfrac{2GM}{r^3} x_p x_q}{c^2 - \dfrac{2GM}{r}}\right) dx^p\, dx^q,$$

where $h_{pq} = -1$ for $p = q$ and $h_{pq} = 0$ for $p \neq q$. But this is just Schwarzschild's metric (see section 18 in [5], where Schwarzschild's metric is written in spherical polar coordinates). ■

4. Ricci's tensor in Galilean system

Certainly, for Schwarzshild's metric Ricci's tensor $R_{\alpha\beta}$ vanishes (that is clear, since Schwarzshild's metric is just obtained by the resolution of the system $R_{\alpha\beta} = 0$ for $\alpha, \beta = 0,1,2,3$). Thus, it seems that the space has the "Einsteinian curvature". Nevertheless, that curvature arises *artificially,* because of the Einsteinian principle on coincidence of the light velocity values "there" and "back" (see section 18 in [4], where for deduction of Schwarzshild's metric the equalities $g_{0m} = 0$ are used), i.e., because of the passage to the locational time. The following lemma shows that for metric (5), in contrast to Schwarzshild's one, some components of Ricci's tensor are distinct from zero (in this connection, see section 6 below, where the difference between the both metrics is considered).

Lemma 1. *Let G be the Galilean coordinate system and at its origin a non-rotating, spherically symmetric mass M of radius r_0 rests. Then the components of Ricci's tensor for the metric defined by the main quadratic polynomial* (5) *have in the system G the following values (for the surrounding space of the mass M):*

$$R_{oo} = 0,\ R_{op} = R_{po} = 0,\ R_{pq} = \frac{10(GM)^2}{c^4 r^6}(x_p\, x_q + h_{pq}).$$

Proof. The metric (5) can be rewritten (for $r \geq r_0$) in the form

$$ds^2 = \left(c^2 - \frac{2GM}{r}\right)(dx^0\, dx^0) + 2\sqrt{\frac{2GM}{r^3}}\, h_{pq}\, dx^q (dx^p\, dx^0) + h_{pq}\, (dx^p\, dx^q),$$

i.e., in the form $ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta$, where

$$g_{00} = c^2 - \frac{2GM}{r},\ \ g_{0p} = g_{p0} = \sqrt{\frac{2GM}{r^3}}\, h_{pq}\, x^q,\ \ g_{pq} = h_{pq}.$$

Now, using the formula $\Gamma_{\mu\nu\sigma} = \frac{1}{2}\left(g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\sigma\nu,\mu}\right)$ for Christoffel's symbol of the first kind and the standard formula $r_{,p} = -\frac{x_p}{r}$, we obtain

$$\Gamma_{000} = 0,\ \Gamma_{00p} = \Gamma_{0p0} = 0,\ \Gamma_{p0q} = \Gamma_{pq0} = \frac{1}{2}(g_{p0,q} - g_{0q,p}) = 0,\ \Gamma_{pqk} = 0,$$

$$\Gamma_{0pq} = \frac{1}{2}(g_{0p,q} + g_{0q,p}) = -\sqrt{\frac{9GM}{2r^7}}\, x_p x_q + \sqrt{\frac{2GM}{r^3}}\, h_{pq}.$$

Furthermore, we use the matrix $(g^{\alpha\beta})$ reciprocal to the matrix $(g_{\alpha\beta})$

$$g^{00} = \frac{1}{c^2},\ \ g^{0p} = g^{p0} = \frac{1}{c^2}\sqrt{\frac{3GM}{r^3}}\, x^p,\ \ g^{pq} = -h^{pq} - \frac{2GM}{r^3} x^p\, x^q.$$

where $h^{\alpha\beta} = 0$ for $\alpha \neq \beta$ and $h^{\alpha\alpha} = -1$. (We note that instead of the computing of the components $g^{\alpha\beta}$ it is sufficient to check that $g_{\alpha\mu} g^{\mu\beta}$ coincides with Kronecker's symbol δ^α_β, i.e., $\delta^\alpha_\beta = 1$ for $\alpha = \beta$ and $\delta^\alpha_\beta = 0$ otherwise.)

With the help of the matrix $(g^{\alpha\beta})$, it is possible to determine Christoffel's symbols of the second kind $\Gamma^{\sigma}_{\nu\mu} = g^{\sigma\alpha}\Gamma_{\alpha\nu\mu}$. We obtain

$$\Gamma^0_{00}=0,\ \Gamma^p_{00}=0,\ \Gamma^0_{p0}=0,\ \Gamma^0_{0p}=0,\ \Gamma^q_{p0}=0,\ \Gamma^q_{0p}=0$$

$$\Gamma^0_{pk}=\frac{1}{c^2}\Gamma_{0pk},\ \Gamma^q_{pk}=\frac{1}{c^2}\sqrt{\frac{2GM}{r^3}x^q}\,\Gamma_{0pk}.$$

Finally, using the formula

$$R_{\alpha\beta}=\Gamma^{\lambda}_{\alpha\lambda,\beta}-\Gamma^{\lambda}_{\alpha\beta,\lambda}+\Gamma^{\rho}_{\alpha\lambda}\Gamma^{\lambda}_{\rho\beta}-\Gamma^{\rho}_{\alpha\beta}\Gamma^{\lambda}_{\rho\lambda},$$

we obtain (with the help of some identical, but enough complicated calculation) the components of Ricci's tensor $R_{\alpha\beta}$ indicated in the lemma. ■

Thus, metric (5) does not satisfy the Einsteinian law of gravitation. Nevertheless, it satisfies the Einsteinian law approximately, up to infinitesimal of the fourth order with respect to $\frac{\|\bar{v}\|}{c}$. Indeed, since the inequality $\frac{|x|}{r}<1$ holds and, by (3), $\left(\frac{\|\bar{v}(r)\|}{c}\right)^4=\left(\frac{2GM}{rc_2}\right)^2$ we obtain

$$R_{pq}=\frac{5}{2r^2}\cdot\left(\frac{2GM}{rc_2}\right)^2\cdot\left(\frac{x_p}{r}\frac{x_q}{r}+\frac{h_{pq}}{r^2}\right)=O\left(\frac{\|\bar{v}(r)\|}{c}\right)^4.$$

5. Gravitational redshift and blueshift.

Consider the radiation of an atom resting at a point $\bar{x}_0$ in the surface of a star that has mass M and radius r_0. In the Einsteinian general relativity theory it is postulated that at the radiation moment the *gravitational redshift* arises, i.e., an increase of the wave length:

$$\frac{\Delta\lambda}{\lambda}\approx\frac{GM}{r_0c^2},$$

but *during the spreading along the light trajectory the wave length does not change.* We name the affirmation written with italic by Dirac's postulate (see section 17 in [5]). In particular, for the Sun $\frac{\Delta\lambda}{\lambda}\approx 2.1\cdot10^{-6}$; experiments confirm this value.

We give another model. There is no gravitational redshift at the radiation moment, but along the light trajectory in the Galilean system the wave length *does not conserve a constant value.* This is possible, since in the Galilean system the light velocity value is not constant.

The following theorem gives the law of changing of the wave length along the light trajectory in the Galilean system. In its statement (and in the sequel) the formulas are written up to infinitesimal of higher order with respect to $\frac{v^2}{c^2}$.

Theorem 3. *Assume that a photon is radiated by an atom that is immovable in the Galilean system (for example, by an atom resting in the surface of a star that has mass M and radius r_0). Furthermore, assume that at the radiation moment the photon has the wave length λ_0 and it has the wave length λ at a point $\bar{x}$ in distance r from the origin, i.e., from the center of M (the wave length is considered in the Galilean system). Then*

$$\frac{\Delta\lambda}{\lambda_0} = -\frac{\Delta\varphi}{c^2} = \frac{1}{2}(\varphi(\bar{x}_0) - \varphi(\bar{x})),$$

where $\Delta\lambda = \lambda - \lambda_0$ *and* $\varphi = \dfrac{GM}{r}$ *is the Newtonian gravitational potential.*

Proof. If an atom rests in the surface of a star that has mass M and radius r_0, then it is immovable in the corresponding Galilean system.

The mass of the photon is equal to $m = \dfrac{\hbar\nu}{c^2}$, where ν is its frequency. The work required for the transference of the photon from the point $\bar{x}$ situated in the distance r to a point $\bar{x} + d\bar{x}$ in the distance $r + dr$ from the origin is equal to $-\dfrac{\hbar\nu}{c^2}d\varphi$. If $dr > 0$, then the potential energy of the photon increases and hence its kinetic energy $\hbar\nu$ decreases. Therefore $\hbar d\nu = \dfrac{\hbar\nu}{c^2}d\varphi$, i.e., $\dfrac{d\nu}{\nu} = \dfrac{d\varphi}{c^2}$. Integrating, we obtain $\Delta \ln \nu = \dfrac{\Delta\varphi}{c^2}$. If the gravitation is not large, then $\dfrac{\Delta\nu}{\nu}$ is small, and we have $\Delta \ln \nu = \ln\dfrac{\nu + \Delta\nu}{\nu} = \dfrac{\Delta\nu}{\nu}$. Thus, $\dfrac{\Delta\nu}{\nu} = \dfrac{\Delta\varphi}{c^2}$. Furthermore, since $\lambda d\nu = -\nu d\lambda$, we conclude that $\dfrac{\Delta\lambda}{\lambda} = \dfrac{\Delta\varphi}{c^2}$. ■

If a photon recedes, then $\Delta\lambda > 0$, since $r > r_0$, i.e., we obtain a gravitational redshift (Fig. 3). In particular, if $r \to \infty$, Theorem 3 implies $\dfrac{\Delta\lambda}{\lambda_0} = \dfrac{GM}{r_0 c^2}$, i.e., we obtain the formula that is postulated in the general relativity theory.

If a photon moves from a point $\bar{x}_0$, where it had the wave length λ_0, to a mass M with an observer in its surface, then by Theorem 3 the wave length perceptible by the observer is *lesser* than λ_0 (Fig 4), i.e., by Theorem 3, a gravitational *blueshift* appears. We describe two experiments which can confirm that blueshift.

Fig. 3.

Experiment 1. Assume that at a point $\bar{x}_0$ in distance $r_0 \approx 30\,000$km from the center of the Earth (say, on a satellite, Fig. 5) there is a laser that sends a light signal of the wave length λ_0 to the Earth. Denoting the mass and the radius of the Earth by M and r, respectively, and the wave length perceptible on the surface of the Earth by λ, we conclude from Theorem 3 that the signal is perceptible on the Earth with the *gravitational blueshift*

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\lambda - \lambda_0}{\lambda_0} = \frac{1}{c^2}\left(\frac{GM}{r_0} - \frac{GM}{r}\right) \approx -5.5 \cdot 10^{-10}.$$

To exclude the gravitation action of the Sun, the experiment has to be made at the sunset time, when the trajectory going from the laser is perpendicular to the direction to the Sun (Fig. 5).

Fig. 4.

Fig. 5.

Experiment 2. Consider a rocket in the half distance between the Earth and the Sun (Fig. 6). Denote by M the mass of the Sun and by r0 the distance from the Earth to the Sun.

Hence the rocket is situated in the distance $r = \frac{1}{2} r_0$ from the Sun. Theorem 3 implies that the light signal sent from the Earth will be perceptible on the rocket with the *gravitational blueshift*

$$\frac{\Delta\lambda}{\lambda_0} = \frac{1}{c^2}(\varphi(\bar{x}_0)\varphi(\bar{x}_0)) = \frac{1}{c^2}\left(\frac{GM}{r_0} - \frac{GM}{r_0}\right) \approx -9{,}9 \cdot 10^{-7}.$$

Fig. 6.

6. Comparison of the postulates.

Consider a conclusion from the postulate of flowing space directly obtained with the help of the main quadratic polynomial (5) *without* the passage to the locational time, i.e., *without* the passage to Schwarzschild's metric.

Let a spherically symmetric mass M be situated at the origin of the Galilean system. We fix a basic distance L and consider (in a laboratory in a distance r from M) two segments be and $be1$ of the length $\frac{L}{2}$ the first of which is directed to M and the second one is perpendicular to it (Fig. 7). At the points e and e_1 there are the mirrors which reflect the light going from b back to the point b. At the point b there is a semitransparent mirror. The glass plate p is included to make the length "there" and "back" the same.

The light going to M has the velocity $c_1 = c + \upsilon$, whereas the light going from M has the velocity $c_2 = c - \upsilon$. Therefore the light signal which is emanated from b, then is reflected at e and returns to b, uses for that way the time

$$\Delta t_M = \frac{L}{2(c+\upsilon)} + \frac{L}{2(c-\upsilon)} = \frac{L}{c}\left(1 + \frac{\upsilon^2}{c^2}\right).$$

Furthermore, by (5), the light velocity in the perpendicular direction (along the segments be_1 and e_1b) has the value $c_\perp = \sqrt{c^2 - \upsilon^2} = c\left(1 - \frac{\upsilon^2}{2c^2}\right)$. Therefore, the light signal which is emanated from b, then is reflected at e_1 and returns to b, uses for that way the time

$$\Delta t_\perp = \frac{L}{c_\perp} = \frac{L}{c}\left(1 + \frac{\upsilon^2}{2c^2}\right).$$

Thus, if the signals are emanated from b simultaneously, then the second signal returns to the point b *earlier* than the first one, and the lag time is equal to

$$\Delta t = \Delta t_M - \Delta t_\perp = \frac{L}{2c} \cdot \frac{\upsilon^2}{c^2}.$$

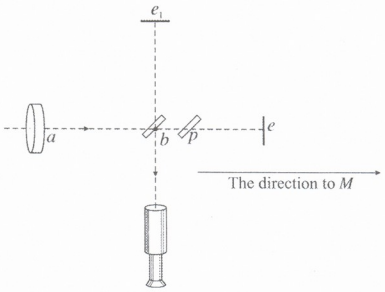


Fig. 7.

By Theorem 1 we obtain $\frac{\upsilon^2}{2c^2} = \frac{2GM}{rc^2}$, and therefore

$$\Delta t = \Delta t_M - \Delta t_\perp = \frac{L}{c} \cdot \frac{GM}{rc^2}.$$

In particular, assume that the Sun is situated at the origin of the Galilean system and the laboratory is situated on the Earth (we note that the Earth with a great exactness may be considered as a Galilean system, since its distance from the Sun is approximately constant). Since the period T of the spectral line of sodium is equal to $1.965 \cdot 10^{-15}$sec, we obtain from (5) that the equality $\Delta t = \frac{1}{2}T$ holds when $\Delta x = L = 29.7$ m. Thus for that basic distance L the lag time Δt is equal to the *semiperiod* of the spectral line of the sodium. This allows to make the following experiment.

Experiment 3. In Michelson-Morley's experiment (see. Exercise 33 to Chapter I in [6]) instead the segments *be* and be_1 the reflection of the light at several mirrors is considered in such a manner that the light passed a way of the length L = 22m. Let us repeat that experiment, taking L = 29.7 m and directing the interferometer to the Sun (since the interferometer has to be situated in a horizontal plane, it is necessary to make the experiment at the sunset time, Fig. 8). Then one of two light signals will have the lag time that is equal to semiperiod of the spectral line of sodium, and the field of vision in the eyepiece of the telescope will be *dark*. Even if the interferometer is rotated in 45° (such that the bisector of the angle between *be* and be_1 will be directed to the Sun, Fig. 9), then $\Delta t = 0$, and the field of vision in the eyepiece of the telescope will be *light*.

Fig. 8.

This shows the difference between the values of the light velocity (average, i.e., along a closed way) in the direction to the Sun and in the perpendicular direction.

The modern technology allows to realize experiment 3 in a simpler manner. Instead of the segment *be* take a spool with light-guide (and the same spool instead of the segment be_1). To exclude the gravitational influence of the Earth both the spools should be situated in such a manner that each loop of the light-guide will be situated in the horizontal plane. Now the length of light trajectory in Experiment 3 should be multiplied by $\frac{2}{\pi}$, since $\frac{1}{2\pi}\int_0^{2\pi}|\cos\alpha|\,d\alpha = \frac{2}{\pi}$. Thus, the length of the light-guide (in each one of the spools) should be replaced by $L \cdot \frac{\pi}{2} = 29.7 \cdot 1.57$ m $= 46.6$ m (certainly, if we use the same spectral line of sodium).

Fig. 9.

We note that, by Lemma 1, metric (5) does not satisfy the Einsteinian law of gravitation. Moreover, it does not satisfy the Einsteinian postulate on coincidence of the values of light velocity "there" and "back", i.e., metric (5) does not satisfy the main postulates of the general relativity theory. The positive result of Experiment 3 will mean that the postulate of flowing space describes the real processes more exactly than Schwarzschild's metric (that satisfies the main postulates of the Einsteinian general relativity theory). Thus, the controversy between the Einsteinian general relativity theory and the presented here ultra-Newton theory will be solved in behalf of the last one.

Furthermore, Laplace's equation is *linear* and hence, summing the potentials of the type $\frac{GM}{r}$ for arbitrary distribution of masses, we again obtain a solution of Laplace's equation. At the same time, Ricci's tensor is *non-lineal.* It is possible that a double star gives another counterexample to the Einsteinian gravitation law.

7. Lorentz' transformations in Galilean system

Let $\mathcal{G}$ be a Galilean coordinate system and at its origin a non-rotating, spherically symmetric mass M rests. We consider the orthonormal coordinate systems $\mathcal{G}_{\bar{x}} = \left\{dx^1_{\mathcal{G}}, dx^2_{\mathcal{G}}, dx^3_{\mathcal{G}}\right\}$ and $\mathcal{F}_{\bar{x}} = \left\{dx^1_{\mathcal{F}}, dx^2_{\mathcal{F}}, dx^3_{\mathcal{F}}\right\}$ introduced after the proof of Theorem 1.

Let now $\mathcal{F}'_{\bar{x}} = \left\{dx'^1_{\mathcal{F}}, dx'^2_{\mathcal{F}}, dx'^3_{\mathcal{F}}\right\}$ be the system that moves uniformly with respect to the flowing system $\mathcal{F}_{\bar{x}}$ with the velocity $\bar{u}$ that is directed to the origin and has the value $\|\bar{u}\| = u$. We assume that the system $\mathcal{F}'_{\bar{x}}$ has the same directions of axes as $\mathcal{F}_{\bar{x}}$ and at the moment $t = 0$ both the systems have the coinciding origins. Moreover, assume that the axis x^1 is directed to the origin and the axes x^2, x^3 have perpendicular directions.

Since the systems $\mathcal{F}_{\bar{x}}$ and $\mathcal{F}'_{\bar{x}}$ are inertial, the passage from $\mathcal{F}_{\bar{x}}$ to $\mathcal{F}'_{\bar{x}}$ is described by Lorentz' transformations:

$$dx^1_{\mathcal{F}'} = \frac{1}{W}(dx^1_{\mathcal{F}} - u dt_{\mathcal{F}}), \quad dt_{\mathcal{F}'} = \frac{1}{W}\left(-\frac{u}{c^2}dx^1_{\mathcal{F}} + dt_{\mathcal{F}}\right), \tag{7}$$

where $W = \sqrt{1-\frac{u^2}{c^2}}$.

Now we pass to the Galilean system $\mathcal{G}_{\bar{x}}$. Denote by $(dx'^1_{\mathcal{G}}, dt'_{\mathcal{G}})$ coordinates of the point $(dx^1_{\mathcal{F}'}, dt_{\mathcal{F}'})$ and by $(dx^1_{\mathcal{G}}, dt_{\mathcal{G}})$ coordinates of the point $(dx^1_{\mathcal{F}}, dt_{\mathcal{F}})$ in the Galilean system $\mathcal{G}_{\bar{x}}$. Applying (4) to Lorentz' transformations (7), we obtain

$$dx'^1_{\mathcal{G}} - \upsilon dt'_{\mathcal{G}} = \frac{1}{W}((dx^1_{\mathcal{G}} - \upsilon dt_{\mathcal{G}}) - u dt_{\mathcal{G}}), \quad dt'_{\mathcal{G}} = \frac{1}{W}\left(-\frac{u}{c^2}(dx^1_{\mathcal{G}} - \upsilon dt_{\mathcal{G}}) + dt_{\mathcal{G}}\right).$$

Making identical algebraic transformations, we obtain the following lemma.

Lemma 2. *Let $\mathcal{F}'_{\bar{x}}$ be the system that moves uniformly with the velocity $\bar{u}$ with respect to the system $\mathcal{F}_{\bar{x}}$. If $\bar{u}$ is parallel to $\bar{\upsilon}$, then the corresponding Lorentz' transformations have in the Galilean systems $\mathcal{G}_{\bar{x}}$ the following form:*

$$dx'^1_{\mathcal{G}} = \frac{1}{W}\left(\left(1-\frac{u\upsilon}{c^2}\right)dx^1_{\mathcal{G}} - u\left(1-\frac{\upsilon^2}{c^2}\right)dt_{\mathcal{G}}\right),$$

$$dt'_{\mathcal{G}} = \frac{1}{W}\left(-\frac{u}{c^2}dx^1_{\mathcal{G}} + \left(1+\frac{u\upsilon}{c^2}\right)dt_{\mathcal{G}}\right).$$

Even if the vector $\bar{u}$ *is perpendicular to* $\bar{\upsilon}$, *say, is directed along the axis* x^2, *then Lorentz' transformations have the form analogous to* (7):

$$dx'^2_{\mathcal{G}} = \frac{1}{W}(dx^2_{\mathcal{G}} - udt_{\mathcal{G}}), \qquad dt'_{\mathcal{G}} = \frac{1}{W}\left(-\frac{u}{c^2}dx^2_{\mathcal{G}} + dt_{\mathcal{G}}\right). \blacksquare$$

The above lemma means that in the Galilean system the main quadratic polynomial is invariant under Lorentz' transformations indicated in the lemma (this affirmation can be verified directly, too). Moreover, Lemma 2 implies that if $\bar{u}$ is vertical or horizontal, then, respectively,

$$\frac{dx^1_{\mathcal{G}}}{dt_{\mathcal{G}}} = u\cdot\frac{1-\frac{\upsilon^2}{c^2}}{1-\frac{u\upsilon}{c^2}} \approx u\cdot\left(1-\frac{\upsilon^2}{c^2}+\frac{u\upsilon}{c^2}\right), \qquad \frac{dx^2_{\mathcal{G}}}{dt_{\mathcal{G}}} = u.$$

This is an essential difference between the ultra-Newtonian gravitation theory and the Einsteinian one. Nevertheless, that difference is too small for an experimental verification.

8. Redshift of distant objects of the universe

The observation shows that the radiation of the distant objects of the universe has the gravitational redshift proportional to their removal. The accepted cause is Hubble's hypothesis on the expansion of the universe, as the result of which Doppler's effect of distant objects arises. In contrast to that accepted point of view, in the presented ultra-Newtonian gravitation theory the following affirmation holds:

Theorem 4. *In the frames of the ultra-Newtonian gravitation theory it is possible to construct without Hubble's hypothesis a model of the universe in which the gravitational redshift is proportional to the distance of the objects.*

A sketch of a proof is given in [4]. For completeness of the article we give here a modified proof.

Proof. Imagine the "universe" as a ball of radius R^* centered at the origin of the Galilean system. Assume that the ball B_r of radius r contains the spherically symmetric mass $M(r) = \frac{2}{\pi}M^* \arctan\left(\frac{\pi}{2}\cdot\frac{kr^2}{R^*-r}\right)$, where $M^* = \frac{c^2R^*}{G}$.

Then the density of the masses tends to zero when $r \to R$.

Let $\bar{x}$ be a point in distance r from the origin. According to the potential theory, the gravitational action of the exterior part of the ball B_r is equal to zero at the point $\bar{x}$. Hence at $\bar{x}$ the potential being made by the whole "universe" *coincides* with the gravitational action of the ball B_r, i.e., $\varphi(\bar{x}) = \frac{GM(r)}{r}$. At the origin $\varphi = 0$. By Theorem 3, for the light going from $\bar{x}$ to the origin we have $\frac{\Delta\lambda}{\lambda} \approx kr$ if r is not very great; indeed,

$$\arctan\left(\frac{\pi}{2}\cdot\frac{kr^2}{R^*-r}\right) \approx \frac{\pi}{2}\cdot\frac{kr^2}{R^*-r} \approx \frac{\pi}{2}\cdot\frac{kr^2}{R^*}$$

and therefore

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\varphi(\bar{x})}{c^2} = \frac{GM(r)}{rc^2} \approx \frac{G}{rc^2}\cdot M^*\cdot\frac{kr^2}{R^*} = kr.$$

Moreover, $\frac{\Delta\lambda}{\lambda} < 1$ for all $r < R^*$. For the points distinct from the origin the picture is analogous.

We note that in the described model of the "universe" the gravitational shift becomes to be *blue* when the point $\bar{x}$ approaches to the boundary of the "universe", i.e., when $r \to R^*$. ■

9. Flux of tachyons

Naturally the question arrises how the postulate of the "flowing space" can be justified? We suggest the following hypothetic explanation. The mass M emanates a spherically symmetric flux of which have neither mass nor energy, but the value of the velocity of tachyons is essentially greater than the light velocity c (in the flowing space).

Interacting with a mass point (or with a photon), the tachyon as if "drags up" it, i.e., the mass point (or a photon) is a little displaced to the mass M. The displacement obtained because of the "tachyon wind" causes a velocity of a mass point (or a photon) in direction to M (Fig. 10), i.e., the space as if "flows" in direction to the mass M. There is a more detailed explanation of the action of the tachyons in articles [2, 3].

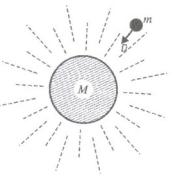


Fig. 10.

Thus, in contrast to the General Relativity Theory, it is possible to explain the ultra-Newtonian gravitation with the help of a flux of some particles which are named here *tachyons.* And if the photons create the time, then the tachyons create the flowing space.

მათემატიკური ფიზიკა

ნიუტონის ზეგრავიტაცია

ვ. ბოლტიანსკი*

* *მათემატიკის კელეიით ცენტრი*

(წარმოდგენილია აკადემიკოს ვ. ხარატიშვილის მიერ)

ნაშრომში მოცემულია გრავიტაციის თეორია, რომელიც ეწინააღმდეგება აინშტაინის ფარდობითობის ზოგად თეორიას. წარმოდგენილი თეორია დაფუძნებულია ავტორის მიერ ჩამოყალიბებულ დინებადი სივრცის პოსტულატზე. ეს პოსტულატი გულისხმობს, რომ სინათლის სიჩქარეს ცენტრის მიმართულებით და მის საწინააღმდეგოდ ერთმანეთისგან განსხვავებული მნიშვნელობები აქვს. დინებადი სივრცის პრინციპიდან გამომდინარე მიღებულია შვარცშილდის მეტრიკა. გარდა ამისა, გამოყვანილია გრავიტაციული წოთელი წანაცვლების ფორმულა, რომელიც ფარდობითობის ზოგად თეორიაში პოსტულირებულია დამტკიცების გარეშე. ნაჩვენებია აგრეთვე, რომ შესაძლებელია აისფერი წანაცვლება. აღწერილია რამდენიმე ექსპერიმენტი, რომელთაგან ორში შესაძლებელია დამზერილ იქნას აისფერი წანაცვლება, მაშინ როდესაც მესამე ექსპერიმენტს შეუძლია გადაჭრას წინააღმდეგობა აინშტაინის მიზიდულობის კანონსა და "დინებადი სივრცის"

პოსტულატს შორის. სამყაროს შორეული ობიექტების წითელი წანაცვლება ახსნილია "გაფართოებადი სამყაროს" შესახებ ჰაბლის ჰიპოთეზაზე დაყრდნობის გარეშე. სტატიის ბოლოს ნაჩვენებია, რომ წარმოდგენილი თეორია შეიძლება დადასტურდეს, თუ განვიხილავთ ზოგიერთი ნაწილაკის ნაკადს (საწინააღმდეგოდ ფარდობითობის ზოგადი თეორიისა, რომელსაც ჰილბერტის გაულენით გააჩნია წმინდა გეომეტრიული, არაფიზიკური აზრი).

REFERENCES

1. *V.G. Boltyanski* (1985), Relativistic fantasy. Obuchenieto po Mat. no. 2, 17-26 (in Bulgarian).
2. *V.G. Boltyanski* (1989), In: Dynamics of Non-Homogenous Systems. Trudy Vsesoyuzn. Inst. Sistem. Issled. 14, 57-67. Moscow (in Russian).
3. *V.G. Boltyanski* (1995), Journal of Applied Analysis, 1, no. 2, 105-119.
4. *V.G. Boltyanski* (2005), Bulletin of the Georgian Academy of Sciences 172, no. 1, 35-38.
5. *P.A.M. Dirac* (1975), General Theory of Relativity. Florida State University. John Wiley & Sons, New York - London - Sydney - Toronto.
6. *E.F. Taylor, J.A. Wheeler* (1966), Spacetime Physics. San Francisco – London.

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Mathematics

On Stable Quaternionic Polynomials

George Khimshiashvili*

* *A. Razmadze Mathematical Institute, Tbilisi*

(Presented by Academy Member N. Vakhania)

ABSTRACT. We present several results on the location and structure of the zero-set of a quaternionic polynomial. Our main result provides an effectively verifiable criterion of stability of such polynomials. We also explain how one can find the number of components of the zero-set having negative real parts. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *unilateral quaternionic polynomial, stable polynomial, Euler characteristic, mapping degree, quadratic form, signature.*

1. We deal with the zero-sets of certain polynomials of one variable over the algebra of quaternions **H** [1]. Consider a so-called *unilateral quaternionic polynomial* of algebraic degree n having the form

$$P(q) = a_n q^n + a_{n-1} q^{n-1} + \ldots + a_1 q + a_0,\ a_0, \ldots, a_n \in \mathbf{H}.$$

Such polynomials naturally form a left **H**-module. It is also useful to consider products of such polynomials assuming that the variable q commutes with coefficients. If the coefficient by monomial of the highest degree is equal to one then P is called a monic (unilateral) quaternionic polynomial of degree n.

Obviously, while investigating the structure of the zero-set $Z(P)$ of P, without reall loss of generality one can assume that P is monic and we will do so in the sequel.

As was proved by S.Eilenberg and I .Niven [2], such a polynomial always has a root in **H** (see also [3], [4]). At the same time, it is well known that the zero-set of such a polynomial can be infinite. For example, the zero-set of polynomial $P(q) = q^2 + 1$ consists of all purely imaginary quaternions of modulus one which form a unit two-dimensional sphere in the hyperplane $\mathbf{H}_0 = \{\mathrm{Re}\ q = 0\}$. It is also easy to produce examples like $q^3 - q^2 + q - 1 = (q - 1)(q^2 + 1)$ where the zero-set contains isolated points as well as infinite components.

Recently, a comprehensive description of zero-sets of unilateral quaternionic polynomials was achieved in [3], [4], [5]. In particular, it was proved in [3], [4] that the zero-set $Z(P)$of such a polynomial P consists of points and two-dimensional metric spheres. Moreover, the results of [3] and [5] imply that the Euler characteristic of the structural sheaf of the zero-set is equal to the algebraic degree of polynomial. In other words, if one takes into account the multiplicities of components of the zero-set, then the number of points plus the doubled number of spheres is equal to n [5].

The latter result gives a natural analog of the "Fundamental Theorem of Algebra" for quaternionic polynomials. It should be added that the methods of [4] and [5] enable one to effectively calculate the number of spherical components of $Z(P)$ for any concrete polynomial P. Since the topological Euler characteristic of $Z(P)$ can be also calculated by the Bruce formula (cf., e.g., [5]), in this way one can find the number of isolated zeroes, which provides substantial information about the geometric structure of $Z(P)$.

Similar but less precise results have been obtained in a recent paper [6]. More precisely, it is shown in [6] that the number of isolated zeroes plus the doubled number of spherical zeroes does not exceed the algebraic degree of P, but

there are no results about the sum of multiplicities of components analogous to the aforementioned quaternionic version of the Fundamental Theorem of Algebra obtained in [5]. Moreover, the problem of finding the number of spherical components of $Z(P)$ was not considered in [6].

2. Having the above results, one may investigate further problems concerned with the zero-sets of quaternionic polynomials. Several interesting problems of such kind were suggested in a recent paper [7] in relation with some problems of mathematical physics. According to [7], to establish the stability of solutions to certain quaternionic differential equations, it is important to have methods of checking that all roots of a given unilateral quaternionic polynomial have negative real parts.

This is a natural quaternionic analog of the classical Maxwell problem about the stable complex polynomials [8]. By analogy with the complex case let us say that a quaternionic polynomial is *stable* if all of its roots have negative real parts. In other words, a polynomial is stable if all of its roots lie in the left half-space defined by the hyperplane $\mathbf{H}_0$.

We aim at establishing an effective criterion of stability of quaternionic polynomial P which is in the spirit of the classical Stodola theorem [8]. In general case, one may wish to find the number of components of $Z(P)$ which lie on the left of $\mathbf{H}_0$. We shall show that this more general problem can also be solved effectively using the signature formulae for topological invariants presented in [5]. It should be added that our considerations and results make an essential use of the results obtained in [3-5].

3. We proceed with presenting the main result. Let P be a unilateral quaternionic polynomial as above. Introduce a new unilateral polynomial P^* which is obtained from P by changing each of its coefficients with its conjugate. In other words, we put $P^*(q) = \sum (a_s)^* q^s$, where the asterisk denotes quaternionic conjugation which acts by changing the sign of the imaginary part of the quaternion. Next, we put $N(P) = PP^*$. Obviously, $N(P)$ is monic and the algebraic degree of N(P) is 2n. The following fact, which was established in [3], can be verified by direct calculation.

Lemma 1. *All coefficients of polynomial $N(P)$ are real numbers.*

Notice now that, given a real polynomial R, one can consider the set $ZZ(R)$ of pairwise sums of roots of R where the number of appearances of each root of R *is equal to its multiplicity. Obviously, there exists a uniquely defined real monic polynomial $Q(R)$* such that its zero-set coincides with $ZZ(R)$.

Lemma 2. *Coefficients of polynomial $Q(R)$ are algebraically expressible through coefficients of R.*

Indeed, the coefficients of $Q(R)$ are symmetric functions of the roots of R, so by the fundamental theorem on symmetric polynomials they can be algebraically expressed through the elementary symmetric functions of the roots which by the Viete theorem are expressible through the coefficients of R.

We are now in a position to formulate the main result.

Theorem 1. *A unilateral quaternionic polynomial P is stable if and only if all the coefficients of polynomials $N(P)$ and $Q(N(P))$ are positive.*

The proof relies on results of [4] and [8] and goes in two steps. First, one derives from the results of [4] that the set of the real parts of the roots of P is always finite and coincides with the set of the real parts of the roots of $N(P)$. Next, one uses the results on stable real polynomials to show that the stability of $N(P)$ is equivalent to the positivity of coefficients of $N(P)$ and $Q(N(P))$. Details will be presented elsewhere.

Notice that this criterion is indeed effective because the coefficients of $N(P)$ and $Q(N(P))$ can be algebraically computed from the coefficients of P. Moreover, it works even in the case if the zero-set of P is infinite. In this way we obtain a complete solution of the stability problem for unilateral quaternionic polynomials.

2. Let us now present a more general result concerned with the problem of finding the number of components of Z(P) having negative real parts. Notice that this problem is meaningful even in the case where $Z(P)$ is infinite because all roots lying in one spherical component have the same real parts.

In order to treat this problem consider P as a polynomial endomorphism of four-dimensional real vector space. Then, as was explained in [4], [5], one can use the Bruce formula [4] to algorithmically compute the geometric Euler characteristic $g(P)$ of $Z(P)$ which is equal to the number of geometrically distinct isolated roots plus the doubled number of the geometrically distinct spherical roots. Notice that, unlike the aforementioned sheaf-theoretic Euler characteristic of $Z(P)$, $g(P)$ need not be equal to n.

Analogously, using the results on the computation of the Euler characteristic of semi-algebraic set presented in [4] one can algorithmically compute the (geometric) Euler characteristic $g_-(P)$ of the set of roots lying in the left half-space $\mathbf{H}_-$. Taking into account that the final formula involves only the signatures of quadratic forms effectively constructible from the coefficients of P [4], we arrive at the second main result.

Theorem 2. *The Euler characteristic $g_-(P)$ of the set of the roots of P lying in the left half space $\mathbf{H}_-$ can be effectively calculated from the coefficients of P using a finite number of algebraic and logical operations.*

Obviously, if $g_-(P) = 0$, then polynomial P is stable. Thus Theorem 2 is indeed more general than Theorem 1.

The author acknowledges financial support by INTAS grant 06-1000017-9093.

მათმატიკა

მდგრადი კვატერნიონული პოლინომების შესახებ

გ. ხიმშიაშვილი*

* *ა. რაზმაძის მათმატიკის ინსტიტუტი, თბილისი*

(წარმოდგენილია აკადემიკოს ნ. ვახანიას მიერ)

ნაშრომში მოცემულია კვატერნიონული პოლინომის მდგრადობის ეფექტური კრიტერიუმი. მოყვანილია აგრეთვე უფრო ზოგადი შედეგი კვატერნიონული პოლინომის ფესვთა სიმრავლის სტრუქტურის შესახებ.

REFERENCES

1. *S. Lang* (1965), Algebra. N.Y.
2. *S. Eilenberg, I. Niven* (1942), Bull. Amer. Math. Soc. 28, 123-130.
3. *N. Topuridze* (2003), Georgian Math. Journal 10, 745-762.
4. *G. Khimshiashvili* (2001), Proc. A.Razmadze Math. Inst. 125, 1-121.
5. *G. Khimshiashvili* (2002), Bull. Georgian Acad. Sci. 165, No.3, 465-468.
6. *A. Pogorui, M. Shapiro* (2004), Complex Variables 49, No.6, 379-389.
7. *G. Ducati, S. Di Leo* (2001), J. Math Physics 42, 2036-2065.
8. *M. Postnikov*. (1982), Stable Polynomials. (Russian) Nauka, Moscow.
9. *P. Griffiths, J. Harris* (1982), Principles of Algebraic Geometry. N.Y.

Received January, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Mathematics

Numerical Analysis of the Stress Distribution by the Boundary Elements Method in Continuous Body with a Hole

Natela Zirakashvili*

* *I. Javakhishvili Tbilisi State University*

(Presented by Academy Member R. Bantsuri)

ABSTRACT. The stress-deformed state of tunnels has been studied. Numerical solutions of the corresponding boundary value problem are obtained by the boundary elements method. The corresponding curves are constructed.

Key words: *boundary elements, fictitious load.*

The stress-deformed state of an underground tunnel is studied, particularly, the main stresses initiated on the tunnel walls are investigated, depending on the depth of its location and on the distance of overground structures from the tunnel.

We consider a boundary value problem as a mathematical model of the mentioned practical problem for a semi-infinite space with a hole. We take a homogeneous isotropic body in the plane deformed state. Its interior boundary surface is stress-free. Nonzero normal loading on one part of the plane boundary surface and zero normal loading on the other are given. Zero tangent loading is given on the whole boundary surface.

Since the body is in plane deformation state, we consider a two-dimensional boundary value problem on the half-space with a hole (Fig. 1).

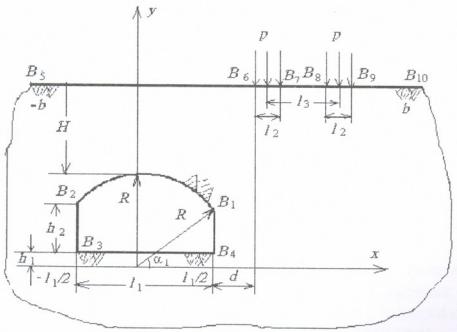


Fig. 1. Half-space with a hole.

Setting of the problem. Let us have a system of equilibrium equations with respect to unknown D, K, u, v on the infinite domain presented in Fig. 1.

$$\begin{aligned} &\frac{\partial D}{\partial r}-\frac{1}{r}\frac{\partial K}{\partial \alpha}=0, \quad \frac{1}{r}\frac{\partial (ru)}{\partial r}+\frac{1}{r}\frac{\partial v}{\partial \alpha}=\frac{D}{\lambda+2\mu}, \\ &\frac{1}{r}\frac{\partial D}{\partial \alpha}+\frac{\partial K}{\partial r}=0, \quad \frac{1}{r}\frac{\partial (rv)}{\partial r}-\frac{1}{r}\frac{\partial u}{\partial \alpha}=\frac{K}{\mu} \end{aligned} \tag{1}$$

Let us find the solution satisfying the following boundary conditions:

$$\begin{aligned} &r=R,\ \alpha_1<\alpha<\pi-\alpha_1:\ \sigma_{rr}=0,\ \sigma_{r\alpha}=0, \\ &x=\pm l_1/2,\ h_1<y<h_1+h_2:\ \sigma_{xx}=0,\ \sigma_{xy}=0\Leftrightarrow \sigma_{rr}=0,\sigma_{r\alpha}=0, \\ &-l_1/2<x<l_1/2,\ y=h_1:\ \sigma_{yy}=0,\ \sigma_{xy}=0\Leftrightarrow \sigma_{rr}=0,\sigma_{r\alpha}=0, \\ &-b<x<l_1/2+d,\ y=R+H:\ \sigma_{yy}=0,\ \sigma_{xy}=0\Leftrightarrow \sigma_{rr}=0,\sigma_{r\alpha}=0 \\ &l_1/2+d<x<l_1/2+d+l_2,\ y=R+H:\ \sigma_{yy}=-p,\sigma_{xy}=0\Leftrightarrow \\ &\qquad\Leftrightarrow \sigma_{rr}=-p\sin^2\alpha_1,\sigma_{r\alpha}=-p\sin\alpha\cos\alpha, \\ &l_1/2+d+l_3<x<l_1/2+d+l_2+l_3,\ y=R+H:\ \sigma_{yy}=-p,\sigma_{xy}=0 \\ &\qquad\Leftrightarrow \sigma_{rr}=-p\sin^2\alpha_1,\sigma_{r\alpha}=-p\sin\alpha\cos\alpha, \\ &l_1/2+d+l_2<x<l_1/2+d+l_3,\ y=R+H:\sigma_{yy}=0,\sigma_{xy}=0\Leftrightarrow\sigma_{rr} \end{aligned} \tag{2}$$

where u, v are the components of the displacement vector, $\sigma_{\alpha\alpha}$, σ_{rr}, $\sigma_{r\alpha}$ are the components of the stress tensor in the polar coordinate system, σ_{xx}, σ_{yy}, σ_{xy} are those in Cartesian coordinates. $D/(\lambda+2\mu)$ is a divergence of the displacement vector, K/μ is a rotor of the displacement vector, $\lambda=E\nu/((1+\nu)(1-2\nu))$, $\mu=E/(2(1+\nu))$ are known constants, E – Young modulus, ν - Poisson coefficient. To obtain a numerical solution of the problem we use one of the methods of boundary elements, in particular, the method of fictitious loadings [1, 2].

Numerical procedure. To conduct a numerical procedure of solving the boundary value problem (1), (2) let us divide the boundaries B_1B_2, B_2B_3, B_3B_4, B_4B_5, B_5B_6, B_6B_7, B_7B_8, B_8B_9, B_9B_{10} into elements N_1, N_2, N_3, N_2, N_4, N_5, N_6, N_5, N_7. If we take into account the expressions of stresses written by means of boundary coefficients of the stress influence [2], and satisfy the boundary conditions (2), one can obtain the following system of $2N$ equations with $2N$ unknowns ($N=N_1+2N_2+N_3+N_4+2N_5+N_6+N_7$):

$$\left.\begin{aligned} &\sum_{j=M_1+1}^{M_2}\left(A_{ss}^{ij}P_s^j+A_{sn}^{ij}P_n^j\right)+\sum_{j=M_3+1}^{M_4}\left(A_{ss}^{ij}P_s^j+A_{sn}^{ij}P_n^j\right)=0 \\ &\sum_{j=M_1+1}^{M_2}\left(A_{ns}^{ij}P_s^j+A_{nn}^{ij}P_n^j\right)+\sum_{j=M_3+1}^{M_4}\left(A_{ns}^{ij}P_s^j+A_{nn}^{ij}P_n^j\right)=-p \end{aligned}\right\}, i=M_1+1,\ldots,M_2,M_3+1,\ldots,M_4,$$

$$\left.\begin{aligned} &\sum_{j=1}^{M_1}\left(A_{ss}^{ij}P_s^j+A_{sn}^{ij}P_n^j\right)+\sum_{j=M_2+1}^{M_3}\left(A_{ss}^{ij}P_s^j+A_{sn}^{ij}P_n^j\right)+\sum_{j=M_4+1}^{M_4+N}\left(A_{ss}^{ij}P_s^j+A_{sn}^{ij}P_n^j\right)=0 \\ &\sum_{j=1}^{M_1}\left(A_{ns}^{ij}P_s^j+A_{nn}^{ij}P_n^j\right)+\sum_{j=M_2+1}^{M_3}\left(A_{ns}^{ij}P_s^j+A_{nn}^{ij}P_n^j\right)+\sum_{j=M_4+1}^{M_4+N}\left(A_{ns}^{ij}P_s^j+A_{nn}^{ij}P_n^j\right)=0 \end{aligned}\right\}, i=1,\ldots,M_1, \tag{3}$$

$$M_2+1,\ldots,M_3,M_4+1,\ldots,M_4+N_7,$$

where $M_1=N_1+2N_2+N_3+N_4$, $M_2=M_1+N_5$, $M_3=M_2+N_6$, $M_4=M_3+N_5$; A^{ij}_{ss}, A^{ij}_{sn}, A^{ij}_{ns}, A^{ij}_{nn} are boundary coefficients of the influence of stresses for the problem under consideration. P^j_s and P^j_n are fictitious unknown values. They have no physical sense and are introduced only as a tool for solving the particular problem, but with their linear combination we can express regular tangent and normal stresses [1, 2] which are used to satisfy the boundary conditions.

After solving the system of equations (3) by an arbitrary method, displacements or stresses may be expressed by another linear combination of fictitious loadings P^j_s and P^j_n (j=1, …, N) at any point of the body.

Numerical realization of boundary-value problem (1), (2) is made. Numerical values of tangent stress s_t are obtained on contour $B_1B_2B_3B_4$ (Fig.1) and the corresponding curves are constructed for the following data: R=12.5, ν=0.33, E=2.5×10^3, ρ=-500 ton/m^2, h_1=2 m, h_2=5.015 m, I_1=20.7 m, I_2 = 2 m, I_3 = 6 m, H = 5, 10, 20, 40 m and d = 10, 3, -10.35, -14.35 m. In Figs. 2 and 3 the diagrams of the distribution of stress s_t on the interior boundary contour are presented.

Fig. 2a) shows diagrams of the stress distribution on the interior contour for four values of d at H=10m. H is the depth of the location of tunnel and d is distance between overground constructions and tunnel wall (see Fig.1). As expected, the farther overground the constructions from the tunnel are, the less is the stress σ_t on the walls of the tunnel.

Fig. 2b) presents the diagrams of the stress distribution on the interior contour for four values of d at H=40 m. In the case when overground constructions are far from the wall of the tunnel the effect is the same as in the previous case. When constructions are quite near to the tunnel (located directly over it), then the stress σ_t increases much more than in the previous case.

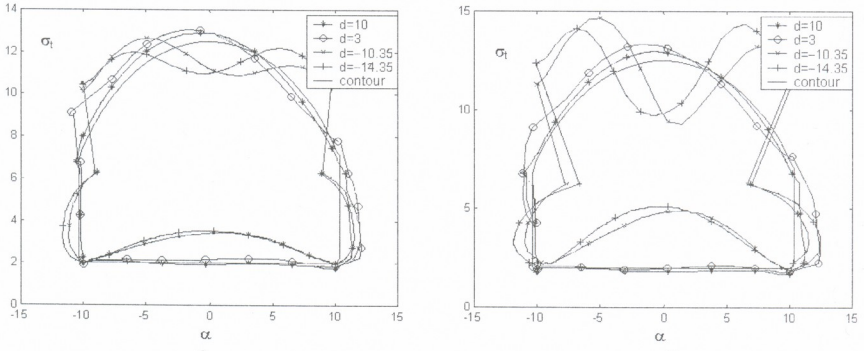


Fig. 2. Distribution of the stress σ_t on the hole outline when loading ρ=-500t/m^2 ; a) H=10m, b) H=40m.

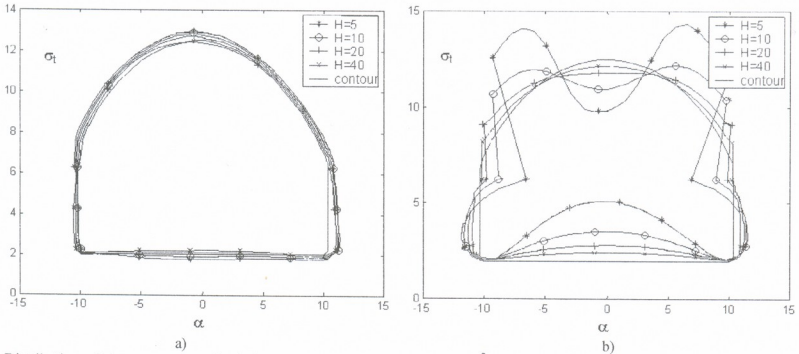


Fig. 3. Distribution of the stress σ_t on the hole outline, when loading ρ=-500t/m^2 ; a) d=10m, b) d=-14.35m.

In Fig. 3a) diagrams of the stress distribution on the interior contour are given for four values of *H*, when d=10 m. As seen from the obtained data, when overground constructions are far from the tunnel, then stress σ_t is small and varies little at different depths of the tunnel. Fig.3b) shows distribution of the stress σ_t on the interior contour for four values of *H* when d=14.35 m. In this case overground constructions are located directly over the tunnel. The obtained data show that the deeper the tunnel is located, the less the tangent stress σ_t is.

მათმატიკა

სასაზღვრო ელემენტთა მეთოდის გამოყენებით ხვრელიან უსასრულო სხეულში ძაბვების განაწილების რიცხვითი ანალიზი

ნ. ზირაქაშვილი*

* *ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

(წარმოდგენილია აკადემიკოს რ. ბანცურის მიერ)

ნაშრომში შესწავლილია გვირაბის დაძაბულ-დეფორმირებული მდგომარეობა. მიღებულია სათანადო სასაზღვრო ამოცანის რიცხვითი ამონახსნები სასაზღვრო ელემენტთა მეთოდით. აგებულია შესაბამისი გრაფიკები.

REFERENCES

1. *S.L. Crouch, A.M. Starfield* (1983), Boundary element methods in solid mechanics. London-Boston-Sydney.
2. *N. Zirakashvili* (2006), Proc. A. Razmadze Mathematical Institute. **141**, 139-147.

Received December, 2006

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Theory of Elasticity

Efficient Solution of BVPs of Thermoelasticity for Half-Plane

Lamara Bitsadze*

* *I. Javakhishvili Tbilisi State University*

(Presented by Academy Member R. Bantsuri)

ABSTRACT. In the present paper explicit solutions of first and second boundary value problems (BVP) of thermoelasticity are constructed for the two-dimensional equations of thermoelastic transversally isotropic half-plane. For their solutions we used the potential method and constructed special fundamental matrices, which reduced the first and second BVPs to a Fredholm integral equation of the second kind. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *potential method, integral equations, thermoelasticity, fundamental matrix.*

Let D be the upper half-plane with the boundary $S(S: x_3 = 0)$, and the normal is (0,1).

We say that a body is subject to a plane deformation parallel to the plane Ox_1x_3 if the second component of the displacement vector $u(u_1, u_2, u_3)$ equals zero and the components u_1, u_3 depend only on x_1, x_3. In this case the basic two-dimensional equations of thermoelasticity for the transversally isotropic body can be written as follows [1]

$$C(\partial x)u = B\operatorname{grad} u_4, \tag{1}$$

$$\left(a_4 \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_3^2}\right) u_4 = 0, \tag{2}$$

where $C(\partial x) = \left\| C_{pq}(\partial x) \right\|_{2\times 2}$, $B = \left\| B_{pq} \right\|_{2\times 2}$, $C_{11}(\partial x) = c_{11}\dfrac{\partial^2}{\partial x_1^2} + c_{44}\dfrac{\partial^2}{\partial x_3^2}$,

$C_{21}(\partial x) = C_{12}(\partial x) = (c_{13} + c_{44})\dfrac{\partial^2}{\partial x_1 \partial x_3}$, $C_{22}(\partial x) = c_{44}\dfrac{\partial^2}{\partial x_1^2} + c_{33}\dfrac{\partial^2}{\partial x_3^2}$, $B_{11} = \beta$, $B_{22} = \beta'$, $B_{12} = B_{12} = 0$,

$a_4 = \dfrac{k}{k'}$, $\beta = c_{13}\alpha' + 2\alpha(c_{11} - c_{66})$, $\beta' = c_{33}\alpha' + 2\alpha c_{13}$, α, α' are coefficients of temperature extension, k, k' are coefficients of thermal conductivity, c_{11}, c_{44}, c_{13}, c_{33} are Hooke's coefficients. $u = u(u_1, u_3)$ is a displacement vector, u_4 is the temperature of body.

Definition. The function $f(x)$ defined in D is called regular, if it has integrable in D continuous second derivatives and $f(x)$ itself and its first derivatives are continuously extendable at every point of S and the conditions of infinite are added $f(x) \in O(1)$, $\frac{\partial u}{\partial x_k} = O\left(|x|^{-2}\right)$, $k = 1,3$, where $|x|^2 = x_1^2 + x_3^2$.

For the equation (1) -(2) we pose the following BVPs. Find a regular solution $u(x)$, $u_4(x)$, of the equations (1) -(2), if on the boundary S one of the following conditions are given:

Problem 1. $u^+ = f(x_1)$, $u_4^+ = f_4(x_1)$.

Problem 2. $\tau_{13}^+ = c_{44}\left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right) = f_1(x_1)$, $\tau_{33}^+ = c_{13}\frac{\partial u_1}{\partial x_1} + c_{33}\frac{\partial u_3}{\partial x_3} - \beta' u_4 = f_3(x_1)$, $\frac{\partial u_4}{\partial x_3} = f_4(x_1)$.

From the equation (2) we find u_4 and the solution of the equation (1) will be presented in the form $u(x) = V(x) + u_0(x)$, where $V(x)$ is a solution of homogeneous equation $C(\partial x)V = 0$, and $u_0(x)$ is a particular solution of equation $C(\partial x)u_0(x) = B\mathrm{grad}u_4$.

1. Solution to the first BVP for half-plane. A solution to the equation (2) in the domain D is

$$u_4(x) = \frac{1}{\pi}\mathrm{Im}\int_s \frac{f_4(t)}{t - z_4}dt, \tag{3}$$

where $z_4 = x_1 + i\sqrt{a_4}x_3$, $f_4 \in H$.

One particular solution $u_0(x)$ to equation (1) is the following

$$u_0(x) = \frac{1}{\pi}\mathrm{Im}\sum_{k=2}^{4}\begin{Vmatrix} A_k & 0 \\ 0 & B_k \end{Vmatrix}\int_s \mathrm{grad}\sigma_k \ln\sigma_k f_4(t)dt, \tag{4}$$

where $\sigma_k = t - (x_1 + \alpha_k x_3)$, $\alpha_k = i\sqrt{a_k}$, $A_k = (-1)^k\left[A_4(c_{44} - c_{33}a_k)\sqrt{a_2a_3a_k^{-1}} + B_4\sqrt{a_4}(c_{13} + c_{44})\right]d$,

$B_k = (-1)^k\left[-A_4(c_{44} + c_{13})\sqrt{a_2a_3a_k^{-1}} + B_4\sqrt{a_4}(c_{44} - c_{33}a_2a_3a_k^{-1})\right]d$, $d^{-1} = \left(\sqrt{a_2} - \sqrt{a_3}\right)(c_{44} + c_{33}\sqrt{a_2a_3})$,

$d_4^{-1} = c_{33}c_{44}(a_4 - a_2)(a_4 - a_3)$, $A_4 = \left[\beta(c_{44} - c_{33}a_4) + \beta' a_4(c_{13} + c_{44})\right]d_4$, $B_4 = \left[-\beta(c_{44} + c_{13}) + \beta'(c_{11} - c_{44}a_4)\right]d_4$,

a_k $(k = 2,3)$ are the positive roots of a characteristic equation

$$c_{33}c_{44}a^2 - \left(c_{11}c_{33} - c_{13}^2 - 2c_{13}c_{44}\right)a + c_{11}c_{44} = 0.$$

It is easy to show that $u_0(x) = 0$, when $x_3 = 0$.

A solution to the equation $C(\partial x)V = 0$, when $V^+ = f(t)$, will be sought in the domain D in terms of the double layer potential

$$V(x) = \frac{1}{\pi}\mathrm{Im}\sum_{k=2}^{3}\left\|N_{pq}^{(k)}\right\|_{2\times 2}\int_s \frac{g(t)dt}{t - z_k}, \tag{5}$$

where

$$N_{11}^{(k)} = (-1)^k d(c_{33}a_k - c_{44})\sqrt{a_2a_3a_k^{-1}},\ N_{21}^{(k)} = i(-1)^k d(c_{13} + c_{44}),\ N_{12}^{(k)} = \sqrt{a_2a_3}N_{21}^{(k)},$$

$$N_{22}^{(k)} = (-1)^k d(c_{44}a_k - c_{11})\sqrt{a_k^{-1}},\ k = 2,3.$$

$g(t)$ is an unknown real vector-function. To determine it we obtain the following integral equation

$$g(t_0)+\frac{1}{\pi}\operatorname{Im}\sum_{k=2}^{3}\left\|N_{pq}^{(k)}\right\|_{2\times 2}\int_s\frac{g(t)dt}{t-t_0}=f(t_0). \tag{6}$$

Taking into account the fact that $\sum_{k=2}^{3}N_{11}^{(k)}=\sum_{k=2}^{3}N_{11}^{(k)}=1$, $\sum_{k=2}^{3}N_{12}^{(k)}=0$. From the equation (6) we have $g(t_0)=f(t_0)$ and (5) takes the form

$$V(x)=\frac{1}{\pi}\operatorname{Im}\sum_{k=2}^{3}\left\|N_{pq}^{(k)}\right\|_{2\times 2}\int_s\frac{f(t)dt}{t-z_k}.$$

Thus we have obtained the Poisson type formula for the solution of the first BVP for the half-plane.

Note that $f\in C^{1,\alpha}(S)$ and satisfies the condition $f(t)=C+\frac{\alpha}{|t|^{1+\beta}}$ at infinity, where C and α are constant vectors and $\beta>0$.

2. Solution to the second BVP for half-plane. The solution to the equation (2) has the form

$$u_4(x)=\frac{1}{\pi\sqrt{a_4}}\operatorname{Re}\int_s\ln(t-z_4)f_4(t)\,dt.$$

A particular solution $u_0(x)$ to the equation (1), when $\tau_{13}(u_0)=0$, $\tau_{33}(u_0)=0$ on S, is the following vector

$$u_0(x)=\frac{1}{\pi}\operatorname{Re}\sum_{k=2}^{4}\begin{Vmatrix}A_k & 0\\ 0 & B_k\end{Vmatrix}\int_s\operatorname{grad}\sigma_k^2\ln\sigma_k f_4(t)\,dt,$$

where $A_k=(-1)^k c_{44}(c_{13}+c_{33}a_k)(A_4+B_4)\left(\sqrt{a_2a_3a_k^{-1}}-\sqrt{a_4}\right)n$,

$B_k=(-1)^k c_{44}(c_{11}+c_{13}a_k)(A_4+B_4)\left(\sqrt{a_2a_3a_k^{-1}}-\sqrt{a_4}\right)a_k^{-1}n$,

$n^{-1}=\left(\sqrt{a_2}-\sqrt{a_3}\right)\left(c_{11}c_{33}-c_{13}^2\right)$, $s^{-1}=2c_{33}c_{44}\sqrt{a_4}(a_4-a_2)(a_4-a_3)$, $A_4=\left[\beta(c_{44}-c_{33}a_4)+\beta' a_4(c_{13}+c_{44})\right]s$,

$B_4=\left[-\beta(c_{44}+c_{13})+\beta'(c_{11}-c_{44}a_4)\right]s$.

Now let us consider the second BVP for the equation $C(\partial x)V=0$. We look for the solution as a single layer potential of the second kind

$$V(x)=\frac{1}{\pi}\operatorname{Re}\sum_{k=2}^{4}\left\|L_{pq}^{(k)}\right\|_{2\times 2}\int_s\ln(t-z_k)h(t)dt, \tag{7}$$

where h is an unknown real vector; the coefficients $L_{pq}^{(k)}$ can be written as follows:

$$L_{11}^{(k)}=(-1)^k(c_{13}+c_{33}a_k)n,\ L_{12}^{(k)}=(-1)^k i(c_{13}+c_{33}a_k)\sqrt{a_2a_3a_k^{-1}}n,$$

$$L_{21}^{(k)}=(-1)^k i(c_{11}+c_{13}a_k)\sqrt{a_k^{-1}}n,\ L_{22}^{(k)}=(-1)^{k+1}(c_{11}+c_{13}a_k)a_k^{-1}\sqrt{a_2a_3}\,n.$$

Taking into account the boundary condition $\tau_{13}^{+}(V)=f_1(x_1),\ \tau_{33}^{+}(V)=c_{13}\frac{\partial V_1}{\partial x_1}+c_{33}\frac{\partial V_3}{\partial x_3}=f_3(x_1),\ x_1\in S$, after direct calculation we find $h(t)=f(t)$.

Therefore, we have the following Poisson type formula for the solution of the second BVP

$$V(x)=\frac{1}{\pi}\operatorname{Re}\sum_{k=2}^{4}\left\|L_{pq}^{(k)}\right\|_{2\times 2}\int_{s}\ln(t-z_k)f(t)dt.$$

For the regularity of the solution $V(x)$ it is sufficient that $\int_s f(t)dt=0$, $f\in C^{0,\alpha}(S)$, $\alpha>0$, and $f(t)=O\left(\left|t^{-1-\beta}\right|\right)$, $\beta>0$, for large $|t|$.

This work was supported by I. Javakhishvili Tbilisi State University.

დრეკადობის თეორია

თერმოდრეკადობის განტოლებათა სისტემის ძირითადი სასაზღვრო ამოცანების ეფექტური ამოხსნა ნახევარსიბრტყის შემთხვევაში

ლ. ბიწაძე*

* *ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

(წარმოდგენილია აკადემიკოს რ. ბანცურის მიერ)

ნაშრომში განხილულია თერმოდრეკადობის განტოლებათა სისტემა ტრანსვერსულად იზოტროპული სხეულისათვის ორი განზომილების შემთხვევაში. პირველი და მეორე ძირითადი სასაზღვრო ამოცანებისათვის მიღებულია პუასონის ტიპის ფორმულები ნახევარსიბრტყის შემთხვევაში.

REFERENCES

1. *W. Nowachi* (1962), Thermoelasticity. Moscow (Russian).

Received January, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Astronomy

On the Probability of Collisions of Interplanetary Space Vehicles with Meteors

Rolan I. Kiladze*

* *Academy Member, E. Kharadze National Astrophysical Observatory of Georgia, Tbilisi*

ABSTRACT. On the average once in 4 years uncontrolled geostationary satellites suffer small sudden changes of speed, which is connected with their collisions with fine space debris. Most of these events are caused by collisions with meteoric bodies. Such collisions threaten space vehicles as well, sent to planets of the solar system to study their physical nature.

The present paper is devoted to determining the degree of risk of collision of a space vehicle with meteors at different possible variants of its interplanetary orbit. The study allows to select - out of several possible variants – the least hazardous trajectory of interplanetary flight of a space vehicle from the point of view of meteoric danger.

Keywords: *interplanetary voyage, meteor, danger, probability.*

1. Introduction

Since the 1990s, in connection with the increased number of space vehicles (SV) investigating interplanetary space, the need has arisen of a careful calculation of their orbits from the viewpoint of minimization of expenditure of power resources, as well of the selection of the minimally hazardous orbits in space.

In the 21st century interplanetary flights of SV with people on board are also planned and practically preparations are under way for expeditions to the nearest planets. Therefore, careful choice of trajectories of flights is not only important but necessary as well.

Our long-term work with uncontrolled geostationary satellites shows that from time to time (on the average, once in 4 years) they suffer small - within the limits of several mm/s - sudden changes of speed, which is connected with their collisions with fine space debris. Most of these events are caused by collisions with meteoric bodies [1-5].

Obviously, SVs, sent to planets of the solar system to study their physical nature, are threatened with such collisions, entailing disaster.

The present study is devoted to determining the degree of risk of collision with meteors for different possible variants of SV interplanetary orbit.

As is known, the bulk of meteoric bodies moves in the form of streams created by comets. These bodies move along the orbit of the comet generating them, at speeds of some dozens of km/s, and at a relatively short (within one million kilometers) distance from it. Sporadic meteors are smaller in number and they move chaotically.

The task of the present work is to choose from a set of possible heliocentric orbits of a SV the optimum, allowing to avoid the crossing of the orbits of comets (the orbits being the sites of congestion of meteoric bodies) by it and ensuring maximal remoteness from them.

It is necessary to note that the details of the distribution of meteoric bodies in streams are more or less known

only for those few objects that cross the Earth's orbit during their annual revolution round the Sun.

Owing to this, it is necessary to assume a uniform distribution of meteoric bodies along orbits. For different streams the identicalness of their spatial density and the laws of its decrease with distance from the orbit of a comet are also postulated.

2. Calculation of the risk-factor of collision of a space vehicle with meteors

We shall further assume that the spatial density of meteoric bodies changes by the Gauss formula according to the distance from the orbit of the comet that has generated meteoric stream:

$$\rho_\Delta = \rho_0 e^{-\left(\frac{\Delta}{\Delta_0}\right)^2}, \tag{1}$$

where ρ_D is the spatial density of meteoric bodies at distance Δ from the comet's orbit, r_0 - spatial density of meteoric bodies on the orbit (constant for all streams). The parameter Δ_0 is accepted to equal 0.01 AU, i.e. 1.5×10^6 km.

To calculate the value of Δ let us introduce a rectangular system of coordinates connected with the orbit of the comet. The origin of the coordinate system coincides with the Sun, X-axis shall be directed to the perihelion of the orbit and Y-axis - to the plane of this orbit, in the direction of the movement of the comet.

The position of a SV at some moment t in the selected system of coordinates will be designated through x_0, y_0, z_0.

The problem of finding the value of Δ is thus reduced to finding the minimal distance between the SV and the orbit of the comet, i.e. to the search of the minimum of the function:

$$\Delta^2 = (x - x_0)^2 + (y - y_0)^2 + z_0^2, \tag{2}$$

where the points on the orbit of the comet are designated through the x and y coordinates.

From theoretical astronomy it is known that:

$$x = \frac{p\cos v}{1 + e\cos v}, \qquad y = \frac{p\sin v}{1 + e\cos v}, \tag{3}$$

where p is the parameter of the orbit, e - its eccentricity, v - true anomaly.

Equating to zero the result of differentiation (2) with respect to v and substituting the values x and y, determined from (3), we receive:

$$(ex_0\cos v_m + x_0 + ep)\sin v_m = y_0[e\cos^2 v_m + (1+e^2)\cos v_m + e], \tag{4}$$

where the value of the true anomaly for a point on the comet's orbit, being at the minimal distance from the SV, is designated through v_m.

Inputting a new variable

$$q = tg\frac{v_m}{2}, \tag{5}$$

the equation (4) can be given in a simpler form:

$$y_0(1-e)^2 q^4 + 2(ep + x_0 - ex_0)q^3 + 2(ep + x_0 + ex_0)q - y_0(1+e)^2 = 0. \tag{6}$$

The fourth degree of the equation (6) relative to q reflects the fact that generally from one point on an ellipse it is possible to drop four perpendiculars, satisfying the condition of the extremum (4).

3. Solution of the equation for true anomaly

The equation (6) is solved most conveniently by way of consecutive approximations.

It is possible to find geometrically, the first approximation to the root of this equation, corresponding to minimal Δ, by crossing the ellipse of the comet's orbit with the bisector of an angle formed by straight lines, connecting the SV with the focuses of the ellipse.

These lines (the sides of an angle) form some angles with X-axis, whose values we shall designate through α and β. They are determined by the following equations:

$$tg\alpha = \frac{y_0}{x_0}, \quad tg\beta = \frac{(1-e^2)y_0}{(1-e^2)x_0 + 2ep}, \tag{7}$$

whence the factor k of the inclination of the bisector is equal to:

$$k = tg\frac{\alpha+\beta}{2}. \tag{8}$$

Hence, the equation for the point of intersection of the bisector with the comet's orbit is as follows:

$$\frac{p\sin v_0}{1+e\cos v_0} - y_0 = k\left(\frac{p\cos v_0}{1+e\cos v_0} - x_0\right), \tag{9}$$

where the preliminary value of v_m is designated through v_0.

Substituting in (9) the value of v_0, determined by means of (5), for the initial (zero) approximation of q we obtain the equation:

$$[(1-e)(y_0 - kx_0) - kp]q_0^2 - 2pq_0 + (1+e)(y_0 - kx_0) + kp = 0, \tag{10}$$

easily solved relative to q_0.

Because the eccentricities of the comets' orbits are close to unity, the member of the fourth degree in the equation (6) is small, allowing to consider a third degree equation, instead of (6):

$$q^3 + 3Aq - 2B = 0, \tag{11}$$

where

$$A = \frac{ep + (1+e)x_0}{3[ep + (1-e)x_0]}, \quad B = \frac{y_0[(1+e)^2 - (1-e)^2 q^4]}{4[ep + (1-e)x_0]}. \tag{12}$$

The value of B is specified by consecutive approximations.

If the inequality takes place:

$$B^2 + A^3 > 0, \tag{13}$$

then the equation (11) has one real root, according to the Kardano formula, equal to

$$q = \left(B + \sqrt{B^2 + A^3}\right)^{1/3} + \left(B - \sqrt{B^2 + A^3}\right)^{1/3} \tag{14}$$

If the inequality (13) does not take place, the equation (11) has three real roots. In that case we shall use iterations in the process of approximations:

$$q_{k+1} = (2B - 3Aq_k)^{1/3}. \tag{15}$$

Here it is useful to apply the following routine allowing to essentially improve the convergence of the iteration process. If we have three consecutive approximations q_1, q_2 and q_3, the following expression may be derived:

$$q_4 = \frac{q_1 q_3 - q_2^2}{q_1 + q_3 - 2q_2}, \tag{16}$$

then q_4 will essentially prove more precise than the previous approximations. With a view to accelerating the convergence of the process of iterations the formula (16) can be applied repeatedly.

4. Transition from the system of coordinates connected with the SV orbit to the system of the comet's orbit

In the system of coordinates connected with an SV orbit (if X-axis is directed to its perihelion) the coordinates of the SV x_1, y_1, z_1 are expressed similarly to (3):

$$x_l = \frac{p_l \cos v_l}{1 + e_l \cos v_l}, \quad y_l = \frac{p_l \sin v_l}{1 + e_l \cos v_l}, \quad z_l = 0. \tag{17}$$

The index l corresponds to the system of coordinates related to the SV orbit and the value i_l for each separately taken moment t_l is determined through the solution of Kepler's equation:

$$tg\frac{v_L}{2} = \sqrt{\frac{1+e_l}{1-e_l}}\, tg\frac{E_L}{2}, \tag{18}$$

where

$$E_l - e_l \sin E_l = M_l. \tag{19}$$

The duration of the flight of the SV is equal to

$$\Delta T = a_L^{3/2}(M_L - M_0), \tag{20}$$

where the initial and final values of the mean anomaly of the SV are designated through M_0 and M_L respectively; a_L is a semimajor axis of the SV orbit.

To calculate the values of v_l by means of the expressions (19) and (20), we shall divide the interval DT into L equal parts.

Transition from the coordinates x_l, y_l, z_l to the coordinates x_0, y_0, z_0 in the system connected with the orbit of the comet is effected by the formulas of spherical trigonometry:

$$x_0 = x_l(\cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i) - y_l(\sin\omega\cos\Omega + \cos\omega\sin\Omega\cos i),$$

$$y_0 = x_l(\cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i) - y_l(\sin\omega\sin\Omega - \cos\omega\cos\Omega\cos i), \tag{21}$$

$$z_0 = \sin i(x_l \sin\omega + y_l \cos\omega),$$

where the inclination, argument of perihelion and the longitude of the Node of the SV orbit relative to the comet's orbit are designated through i, ω and Ω respectively. These values are calculated by means of the expressions:

$$\cos i = \cos i_n \cos i_l + \sin i_n \sin i_l \cos(\Omega_n - \Omega_l),$$

$$\omega = \omega_l - arctg\frac{\sin i_n \sin(\Omega_n - \Omega_l)}{-\cos i_n \sin i_l + \sin i_n \cos i_l \cos(\Omega_n - \Omega_l)}, \tag{22}$$

$$\Omega = arctg\frac{\sin i_l \sin(\Omega_n - \Omega_l)}{\sin i_n \cos i_l - \cos i_n \sin i_l \cos(\Omega_n - \Omega_l)} - \omega_n.$$

where the index n designates the elements of an orbit of n^{th} comet relative to the ecliptic.

Finally, the non-normalized probability of collision of the SV with a meteor for the entire time of flight will be:

$$P_{LN} = \frac{\Delta T}{L+1}\sum_{i=0}^{L}\sum_{k=1}^{N} e^{-\left(\frac{\Delta_{ik}}{\Delta_0}\right)}, \tag{23}$$

where N designates the total number of the orbits of the comets used at calculations.

By the software realized on Fortran, in the case of using elements of the orbits of 500 comets and dividing the DT interval into 1000 times, it takes PC Pentium – 4, to calculate the probability of collision of an SV with meteors, 4s of machine time.

Thus, the present study allows choosing the safest, out of several possible variants of the trajectory of interplanetary flight of SV, from the point of view of collisions with meteors.

ასტრონომია

მეტეორებთან საპლანეტთაშორისო ხომალდის შეჯახების ალბათობის შესახებ

რ. კილაძე*

* *აკადემიის წევრი, ე. ხარაძის სახელობის ნაციონალური ასტროფიზიკური ობსერვატორია*

უმართავი გეოსტაციონარული თანამგზავრები საშუალოდ ოთხ წელიწადში ერთხელ მყისიერად იცვლიან სიჩქარეს, რაც გამოწვეულია კოსმოსურ ნაგავთან მათი შეჯახებებით. ასეთი შემთხვევების უდიდესი ნაწილის მიზეზია შეჯახებები მეტეორებთან.

მსგავს შეჯახებებს განიცდიან მზის სისტემის პლანეტებისკენ მათი ფიზიკური ბუნების შესასწავლად გაგზავნილი კოსმოსური ხომალდებიც.

წინამდებარე ნაშრომი მიზნად ისახავს კოსმოსური ხომალდების მეტეორებთან შეჯახების რისკის განსაზღვრას საპლანეტთაშორისო ორბიტის სხვადასხვა ვარიანტისათვის. ნაშრომი საშუალებას იძლევა, საპლანეტთაშორისო ხომალდის ტრაექტორიის რამდენიმე შესაძლო ვარიანტიდან არჩეული იქნეს მეტეორული საფრთხის თვალსაზრისით ყველაზე უხიფათო.

REFERENCES

1. *R. I. Kiladze* (2001), O novoi teorii dvizheniya geostatsionarnykh sputnikov. In: Nasireddin Tusi i Sovremennaya Astronomia, Shemakhinskaya AO (About a new theory of motion of geostationary satellites, In: Nasireddin Tusi and Modern Astronomy), 113-121 (in Russian).
2. *R. I. Kiladze, A. S. Sochilina, K.V. Grigoriev, A. N. Vershkov* (1997), On the investigation of long-term orbital evolution of geostationary satellites, Proc. 12th Intern. Sympos. Space Flight Dynamics, 53-57.
3. *R. I. Kiladze, A. S. Sochilina, K.V. Grigoriev, A. N. Vershkov* (1999), On occasional changes of velocities of geostationary satellites, US Nav. Obs., 20-23 Oct., 1998. Washington, DC, 339-351.
4. *R. I. Kiladze, A. S. Sochilina, F. R. Hoots, R. France, A. N. Vershkov* (2001), On statistics of changes in rates of drift among uncontrolled geostationary objects, Proc. III European Conf. of Space Debris, ESOC, 367-372.
5. *A. S. Sochilina, R. I. Kiladze* (2004), O dinamicheskom metode kontrolya zasoreniya geostatsionarnoi orbity. Trudy nauchnogo seminara "Ekologia i kosmos" 13 Fevralya 2004 g., St.-Petersburg (About a dynamic method of monitoring the contamination of a geostationary orbit. Proc. Scientific Seminar «Ecology and Space» on February, 13th, 2004), 19, (in Russian).

Received August, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Physics

On Vortex Model of Planet Formation in Keplerian Disks

Alexander G. Tevzadze*, George D. Chagelishvili, Jumber G. Lominadze§**

* *E. Kharadze Georgian National Astrophysical Observatory, Tbilisi*
** *Academy Member, E. Kharadze Georgian National Astrophysical Observatory, Tbilisi*
§ *Academy Member, E. Kharadze Georgian National Astrophysical Observatory, Tbilisi*

ABSTRACT. Invoking a hypothesis that long-lived vortices in the protoplanetary nebula play an important role in the three-phase model of planet formation, we analyze the first phase and present results on the stability and nonlinear development of vortices against the background of shearing Keplerian flows. We discuss the conditions under which vortical perturbations evolve into long-lived self-sustained structures and describe the properties of these equilibrium vortices. The properties of equilibrium vortices appear to be independent of the initial conditions and depend only on the local disk parameters. In particular, we find that the ratio of the vortex size to the local disk scale height increases with the decrease of the sound speed, reaching values well above the unity. The process of spiral density wave generation by the vortex leads to the formation of spiral shocks attached to the vortex. These shocks may have important consequences on the long term vortex evolution and possibly on the global disk dynamics. Our study strengthens the arguments in favor of anticyclonic vortices as the candidates for the promotion of planet formation. Hydrodynamic shocks that are an intrinsic property of persistent vortices in compressible Keplerian flows are an important contributor to the overall balance. These shocks support vortices against viscous dissipation by generating local potential vorticity and should be responsible for the eventual fate of the self-sustained anticyclonic vortices. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *planet formation, protoplanetary disks.*

1. Introduction

The modern trend of investigations on the formation of planetary systems may be attributed to Laplace [1], who introduced the nebular hypothesis to rationalize the distribution and the motion of planets and their satellites in the solar system. The conjecture is based on the fact that the solar system is formed from a rotating flattened gas cloud. In fact the astronomical nebular theory advanced by Laplace is a root of many astrophysical investigations as much as disk-like rotating bodies represent widely spread and important structural elements of our Universe. Among them are galaxies, quasars, accretion disks surrounding compact stellar objects (white dwarfs, neutron stars and black holes) in binary systems, protostellar and protoplanetary disks. The formation of protoplanetary disks proceeds by the following scenario: Stars condense from an interstellar medium consisting mostly of gas with an admixture of solid particles called interstellar dust. Both observational and numerical studies suggest that, as the central star contracts, it leaves around material that contains a sufficiently great share of the initial angular momentum of the whole system. In this nebula, the centrifugal force balances the stellar gravity in the radial direction and a protoplanetary disk is formed. Such disks represent initial material for planet formation.

It is just over ten years since the first planet outside our solar system was detected. Since then, much work has focused on understanding the formation of planetary systems in general and our solar system in particular. It is investigation of the dynamics of coherent vortices in protostellar and protoplanetary disks that has received increasing attention from the end of the

last century [2]. However, it was von Weiszacker who proposed the hypothesis that long-lived vortices in protoplanetary nebula can play an important role in planet formation [3]. Indeed, it has been shown that vortices, if sustained long enough, lead to particle aggregation in their core and to the formation of protoplanets [4-10]. Invoking a three-phase model for the planetary formation [11-12], one can consider the process of vortex formation as the first phase. The second phase is characterized by the accumulation of solids in the center of vortices and by the growth of a planetary core, and the third phase by the accretion of gas onto the core. Our research is related to the first phase of the planet formation – development of long-lived vortices.

The vortex scenario for planet formation encounters an apparent obstacle: any structure in a protoplanetary/Keplerian disk is subject to strong shearing that may eventually lead to its decay. The only mechanism for sustaining a stable vortex in such flows is nonlinearity. Hence, vortices that may start the process of planet formation should exceed a critical threshold in their amplitude. Direct numerical simulations are therefore an important tool in these studies. Several works have been devoted to the analysis of the possibility of forming and maintaining coherent vortex structures in the strongly sheared flow pertaining to a Keplerian disk, both in barotropic configurations, where the perturbations of potential vorticity are conserved [6,7,13- 15], and in baroclinic situations, where potential vorticity can be generated [11-12]. In the incompressible case it has been shown that coherent vortex structures can indeed survive nonlinearly (with conserved potential vorticity). In this case anticyclonic vortices can survive longer than cyclonic ones [13] and give rise to the Rossby waves in the system [15]. The effects of compressibility have not yet been fully analyzed and require rigorous study. Two-dimensional time-dependent numerical simulations of vortices in viscous compressible Keplerian disks can be found in [6]. Vorticity waves are considered as one of the constituents of (anticyclonic) vortex dynamics, but without description of the wave properties and any analysis of their genesis and dynamics (the subject of the study was the stability and lifetime of vortices). Detailed analytical and numerical study of the dynamics of perturbations (vortex/aperiodic mode, Rossby and spiral-density waves) in 2D compressible disks with a Keplerian law of rotation has been performed in [16]. The study has shown that small amplitude coherent circular vortex structures are capable of generating density-spiral waves linearly. The main features of this generation phenomenon have been studied, using global direct numerical simulations.

Here we present the results of a numerical study of the nonlinear dynamics of vortices and spiral-density waves in compressible Keplerian disk flows. We show that nonlinear anticyclonic vortices undergo direct nonlinear adjustment to the long-lived self-sustained coherent structure. Vortices generate density-spiral waves under the influence of Keplerian shear. We show the nonlinear development of spiral-density waves that result in the formation of spiral shocks with a steady spatial pattern. These shocks may increase the stability of anticyclonic vortices by slowing down their decay and may also affect the global disk dynamics.

2. Physical model and numerical setup

The following are the basic equations governing the dynamics of disk flows rotating around the gravitating center in cylindrical co-ordinates:

$$\frac{\partial \rho}{\partial t}+\frac{1}{r}\frac{\partial}{\partial r}\left(r\rho V_r\right)+\frac{1}{r}\frac{\partial}{\partial \varphi}\left(\rho V_\varphi\right)=0, \tag{1}$$

$$\frac{\partial V_r}{\partial t}+(V\nabla)V_r-\frac{V_\varphi^{\,2}}{r}=-\frac{1}{\rho}\frac{\partial P}{\partial r}-\frac{\partial \Phi}{\partial r}, \tag{2}$$

$$\frac{\partial V_\varphi}{\partial t}+(V\nabla)V_\varphi+\frac{V_\varphi V_r}{r}=-\frac{1}{\rho r}\frac{\partial P}{\partial \varphi}, \tag{3}$$

$$\left(\frac{\partial}{\partial t}+(V\nabla)V\right)P=\gamma\frac{P}{\rho}\left(\frac{\partial}{\partial t}+(V\nabla)V\right)\rho, \tag{4}$$

where

$$(V\nabla)=V_r\frac{\partial}{\partial r}+\frac{V_\varphi}{r}\frac{\partial}{\partial \varphi}.$$

Our equilibrium state corresponds to the Keplerian disk flow with $V_0=(0,r\Omega(r))$ and $\Omega(r)\sim r^{-3/2}$. The central gravitational potential balances the centrifugal force: $\Phi(r)\sim r^{-1}$. Equilibrium pressure and density are set to be constants.

Initial conditions for our simulations are composed by Keplerian equilibrium flow and local vortical perturbations (denoted with prime) with the following geometry:

$$V_x'(0)=\pm\frac{\varepsilon(y-y_0)}{q}\exp\left[-\frac{(x-x_0)^2}{a^2}-\frac{(y-y_0)^2}{q^2a^2}\right], \tag{5}$$

$$V_y'(0)=\mp\frac{\varepsilon(x-x_0)}{q}\exp\left[-\frac{(x-x_0)^2}{a^2}-\frac{(y-y_0)^2}{q^2a^2}\right], \tag{6}$$

$$u'(0)\ @0. \qquad (7)$$

Here ε defines the amplitude of the initial perturbation and its sign determines the vortex polarity (positive in the case of anticyclonic vortex). The parameters a and q describe, respectively, the size (in the radial direction) and aspect ratio of an elliptic vortex. A circular vortex corresponds to $q = 1$, and $q > 1$ refers to a vortex elongated in the azimuthal direction. (x_0, y_0) so that it corresponds to the radial location $r_0 = 1$.

We perform direct numerical simulations (DNS) based on the nonlinear set of Eqs. (1-4) using initial conditions corresponding to the sum of equilibrium Keplerian flow and a local vortex perturbation shown in Eqs. (5-7). Numerical simulations are based on the DNS code PLUTO [17] with implemented FARGO scheme [18].

3. Nonlinear development of vortices

The evolution of the perturbations depends on three main parameters: the amplitude and size of the perturbation (respectively ε and a) and the sound speed in the disk, Cs. Our first aim is to determine the region, in the parameter space described by ε, a and Cs, in which the evolution of the initial perturbation leads to a stable, long-lived equilibrium configuration. As we shall see, before reaching this final state the system undergoes, in the course of several disk revolutions, a transition phase that we call *nonlinear adjustment*. Our second aim is to provide a detailed description of the equilibrium vortex configuration. With these purposes we have performed runs with different values of the three parameters and different numerical setups, using Eqs. 5-7 with $q = 1$ (circular vortices). In particular, by increasing the value of the sound speed ($Cs = 0.001,\ 0.01,\ 0.1$), we have explored the behavior of the vortex changing a at fixed e and changing e at fixed a. Additional calculations have been performed for the purpose of exploring in more detail particular regions of the parameter space. For instance, additional values of Cs have been used for a better understanding of the scaling behaviors of some of the vortex properties. Moreover, in order to investigate whether the general behavior is changed by varying the shape and structure of the initial perturbation, we have performed computations with different values of the ellipticity parameter.

We found that there are two threshold parameters for the vortex amplitude that control the fate of initially imposed vertical perturbation in Keplerian flow. The first threshold parameter $\varepsilon = 0.1$. When the vortex amplitude exceeds the first nonlinear threshold, a two-stage process occurs. First, the vortex is sheared into a narrow vortex layer, which then undergoes local instabilities. We then observe the formation of small-scale weak anticyclonic vortices at different azimuthal locations.

Vortices with $\varepsilon > \varepsilon^{**}$ experience direct adjustment from the initial to the final self-sustained structure, *i.e.*, a strong anticyclonic vortex is developed. The final equilibrium configuration appears to be a non-linear attractor reached by the system if the initial amplitude exceeds ε^{**} and if the initial spatial scale falls in a range discussed in the next subsection. Indeed, the same nonlinear state is developed from all initial vortices satisfying these conditions, independently of the details of the initial potential vorticity distribution (exponential or algebraic, circular or elliptic). Fig. 1 shows the nonlinear vortex configuration developed from the initially imposed anticyclonic vortex perturbations with amplitude exceeding the second nonlinear threshold.

The result of nonlinear adjustment strongly depends also on the initial vortex size. In order to quantify this dependence, we have carried out computations with vortices of different initial size a. Our numerical results show that vortices undergoing direct nonlinear adjustment have initial size in the range $a_{min} < a < a_{max}$.

When the spatial scale of the initial vortex exceeds a_{max}, the evolution is quite complex. We observe a radial

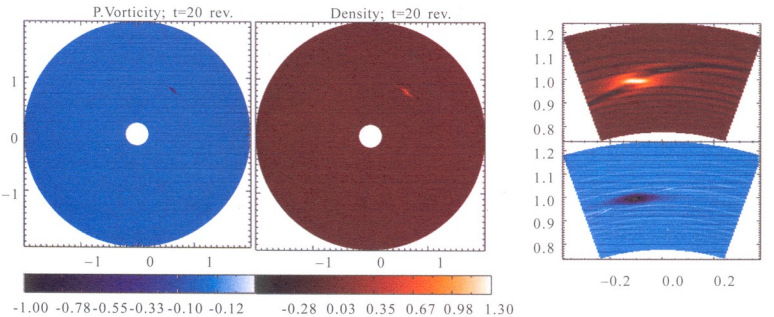


Fig. 1. Result of the nonlinear adjustment of the initially imposed anticyclonic vortex 20 local disk revolutions. Surfaces of the potential vorticity (left) and density (center) are shown in blue and red color, respectively. The right panel zooms on the vortex structure. Density maximum in the center of the vortex corresponds to the potential vorticity minimum.

transfer of potential vorticity in both directions, caused by the combined action of shocks (radiated from the initially imposed supersonic vortex) and of flow curvature (inducing Rossby wave variations). When the size of the vortex is smaller than a_{min} and the initial vortex amplitude exceeds the second nonlinear threshold, we observe nonlinear adjustment to a final configuration with sizes of the order of the initial value.

One of the main goals of the present study is to describe the stability and structure of long-lived self-sustained vortices in Keplerian disks. For this purpose, we selected cases undergoing direct adjustment to a single vortex, and we followed their long-time behavior. Fig. 1 shows radial profiles of potential vorticity and density at the center of a vortex.

Interestingly, the self-sustained vortex seems to be temporarily able to oppose viscous dissipation, exhibiting for some time an increase of the maximum potential vorticity. This effect implies production of potential vorticity. On the other hand, potential vorticity is a nonlinearly conserved quantity in barotropic flows. Thus, the spiral shock waves are the sources of the coherent generation of potential vorticity necessary to support or even enhance the vortex.

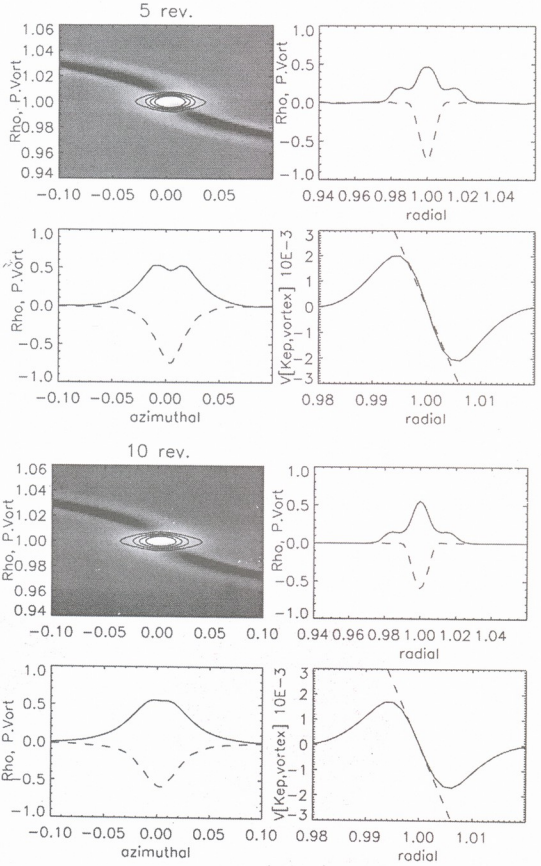


Fig. 2. The structure of the nonlinear long-lived anticyclonic vortex. The two panels whose vortex structure after 5 and 10 local disk revolutions, respectively. Graphs show grayscale visualisation of the density over potential vorticity contours (top left), together with radial and azimuthal cuts in density and potential vorticity. Radial velocity profiles of the background Keplerian flow and local nonlinear vortex are shown in bottom right graphs. Surfaces of the density perturbations reveal high density cores in the center of the vortex together with the well profound density-spiral waves beaming from the vortex center.

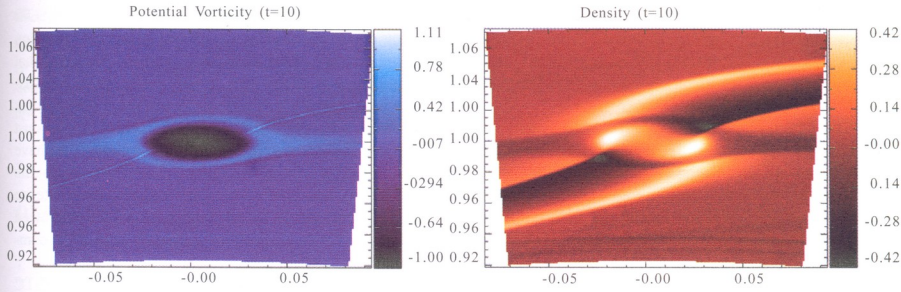


Fig. 3. Potential vorticity (left) and density (right) of the long-lived anticyclonic vortex after 10 local disk revolutions. Anticyclonic potential vorticity balances increased density within the vortex. Shock waves are revealed in jump lines beaming out of the central circulation. Density jumps associated with this steady pattern of shock waves are also readily seen.

The nonlinear dynamics of cyclonic vortices shows significant differences from anticyclonic ones. In the limit of small amplitudes, cyclonic vortices decay on the shearing timescale, related to the geometrical stretching induced by the background differential rotation.

Vortices in shear flows generate spiral density waves by a linear mechanism first described in [19] and further investigated numerically in Keplerian flows [16]. In the present context we study the characteristics of this process when nonlinear forces are also active. As clearly seen from Fig. 2 the dynamics of potential vorticity is accompanied by wave emission, not only during the transition time, but also afterwards, when the nonlinear self-sustained vortex is fully developed. This enforces the fact that nonlinearity does not suppress the wave emission. Moreover, after the adjustment, waves emitted by the coherent vortex structure appear to develop into spiral shocks. Coherent emission of acoustic waves leads to the density accumulation in compressible motions and to the development of shocks. The shock front naturally matches the ray trajectory of the generated waves, since these waves are responsible for the sustained pattern of dissipating shock waves.

Traces of shock generation by vortices may be found in [20]. In our simulations we verify the existence of a steady pattern of spiral shocks produced by a single self-sustained vortex. Moreover, shock waves appear to be an inherent property of vortices in sheared compressible flows, but they can be observed only at fairly high resolutions using shock capturing schemes. Our high resolution calculations allow to study these shock waves in fine details. Fig. 2 shows the distributions of potential vorticity and density for the vortex and the attached shock waves. One can distinctly recognize a wave-crest of the density-spiral wave developing into a double shock configuration, with the shock ahead of the vortex facing the outward region. A couple of much weaker shocks, parallel to the strong ones, appear to be present, although they remain barely visible. These shocks strongly affect the density structure of the developed vortex configuration, resulting in a splitting of the vortex core. Eventually, however, the shearing background leads to the merging of the cores, but the shocks persist.

Spiral shocks induced by a wake of planets are believed, in some situations, to be responsible for planet migration (see [21] and references therein). In our computations no radial variation of the vortex position has been observed. As seen in previous studies, spiral shocks affect dust accretion rates on the vortex core and thus promote the formation of a planetesimal. In this sense, they increase the importance of anticyclonic vortices in planetary formation. The presence of shocks has consequences for the vortex evolution. Nevertheless, the final fate of these structures cannot be easily foreseen and requires much longer simulations. We can here only sketch some possible scenarios. One possibility is the exhaustion of matter in the vortex bearing ring and the formation of an isolated planetesimal. Arguments that spiral shocks may lead to the gap formation can be found in [22]. On the other hand, shocks heat the ring at the radius where the vortex is sustained, which in turn may trigger the linear Rossby wave instability due to the unusual entropy gradient in the disk matter [23]. The instability will induce radial mixing and possibly the destruction of the coherent vortices. On the other hand, spiral shocks can be themselves unstable in three dimensions [24]. Hence, as we said, longer and possibly three-dimensional simulations can clarify this issue.

4. Summary

We have shown the possible existence of anticyclonic vortices with sizes exceeding the Keplerian disk height scales. We have followed the evolution of such vortices for 200 local revolutions, showing their persistence and stability.

We have found that the development of long-lived self-sustained nonlinear anticyclonic vortex configura-

tion occurs only when the amplitudes of the initially imposed vortex perturbations exceed some threshold value. We have interpreted the latter process as the nonlinear vortex adjustment and studied the parameters that can describe this process.

The structure of the developed long-lived vortex does not depend on the initial vortex configuration, provided it exceeds the second threshold amplitude and its size does not exceed a limiting value. In this sense we found a nonlinear attractor that is the final configuration of a wide range of initial vortical perturbations.

Vortices generate density-spiral waves that rapidly develop into shocks. As a result, a long-lived nonlinearly balanced vortex is accompanied by two spiral compressible shock waves facing both radial directions.

We analyzed cyclonic vortices at nonlinear amplitudes. It seems that the linear decay due to the shearing deformation is accelerated by nonlinear effects. Our study contributes to the scenario of planetary formation inside the core of the long-lived vortices. We found that protoplanetary disks with lower sound speed can sustain vortices with a higher ratio of vortex size to disk thickness and create more favorable conditions for dust trapping and mass accumulation. In this context, we have also found a steady increase of density inside the nonlinearly balanced vortex, partly, due to the existence of persistent, steady, spiral shock waves that we showed to be an intrinsic property of stationary vortices in compressible Keplerian fows.

Acknowledgements. This work is supported by ISTC grant G-1217. A.G.T. would like to acknowledge the hospitality of Osservatorio Asrtonomico di Torino. Numerical calculations have been partly performed in CINECA (Bologna, Italy) thanks to the support by INAF.

ფიზიკა

კეპლერულ დისკებში პლანეტების წარმოშობის გრიგალური მოდელის შესახებ

ა. თევზაძე*, გ. ჩაგელიშვილი**, ჯ. ლომინაძე§

* *ე. ხარაძის საქართველოს ეროვნული ასტროფიზიკური ობსერვატორია, თბილისი*
** *აკადემიის წევრი, ე. ხარაძის საქართველოს ეროვნული ასტროფიზიკური ობსერვატორია, თბილისი*
§ *აკადემიის წევრი, ე. ხარაძის საქართველოს ეროვნული ასტროფიზიკური ობსერვატორია, თბილისი*

პლანეტების წარმოშობის სამფაზოვანი მოდელის მიხედვით პროტოპლანეტარულ ნისლოვანებაში ხანგრძლივად არსებულ გრიგალებს შეუძლიათ მნიშვნელოვანი როლი ითამაშონ პლანეტების ფორმირების პროცესში. ვიყენებთ რა ამ ჰიპოთეზას, ჩვენ ვაანალიზებთ პლანეტების ჩამოყალიბების პირველ ფაზას და წარმოვადგენთ არაერთგვაროვნად მბრუნავ კეპლერულ დისკებში გრიგალების მდგრადობისა და არაწრფივი დინამიკის კვლევის შედეგებს. ჩვენ ვმსჯელობთ იმ პირობებზე, რომლის დროსაც გრიგალური ტიპის შეშფოთებები ევოლუციის შედეგად აჟალებენ სიცოცხლისუნარიან თვითშენარჩუნებად კოჰერენტულ სტრუქტურებს და აღვწერთ ამ წონასწორული გრიგალების თვისებებს. ამ არაწრფივი წონასწორული გრიგალების თვისებები არ არის დამოკიდებული საწყის პირობებზე და განისაზღვრება მხოლოდ დისკის ლოკალური პარამეტრებით. კერძოდ, კვლევებმა აჩვენეს, რომ გრიგალის ზომის შეფარდება დისკის მახასიათებელ ლოკალურ სისქესთან იზრდება ბგერის სიჩქარის კლებასთან ერთად და აღწევს ერთზე მნიშვნელოვნად უფრო მაღალ სიდიდეებს ცივ პროტოპლანეტარულ დისკებში. გრიგალების მიერ სპირალურ-გრავიტაციული ტალღების გენერაციის პროცესი იწვევს გრიგალიდან გამომავალი დარტყმითი ტალღების გენერაციას. ასეთი ტიპის დარტყმითმა ტალღებმა შეიძლება გამოიწვიონ მნიშვნელოვანი შედეგები როგორც გრიგალის ხანგრძლივი ევოლუციის პროცესში, ისე დისკის გლობალურ დინამიკაში. კეპლერულ დისკებში სიცოცხლისუნარიანი გრიგალების თანმდევი ჰიდროდინამიკური დარტყმითი ტალღები არიან ჯამური სურათის

მნიშვნელოვანი მონაწილენი. დარტყმით ტალღებს შეუძლიათ ლოკალური პოტენციური ცირკულაციის გენერაცია, გრიგალების კვება დისიპაციური დანაკარგების ასანაზღაურებლად და მათი საბოლოო ბედის განსაზღვრა. ჩვენი კვლევები ამყარებენ არგუმენტებს იმ მოდელის სასარგებლოდ, რომლის მიხედვითაც ჰიდროდინამიკური გრიგალები წარმოადგენენ პლანეტების ფორმირების პროცესის ხელშემწყობ ფაქტორს, და პლანეტების ფორმირება იწყება ხანგრძლივად არსებული თვითშენარჩუნებადი გრიგალების ბირთვებში.

REFERENCES

1. *P. S. Laplace* (1796), Exposition du Système du Monde (Courcier: Paris).
2. *F. C. Adams and R. Watkins* (1995), Astrophys. J. **451**, 314.
3. *C. F. von Weiszacker* (1944), Z. Astrophys., **22**, 319.
4. *P. Barge and J. Sommeria* (1995), Astron. and Astrophys. **295**, L1.
5. *P. Tanga, A. Babiano, B. Dubrulle and A. Provenzale* (1996), Icarus, 121, 158.
6. *P. Godon and M. Livio* (1999), Astrophys. J. **521**, 319.
7. *P. Godon and M. Livio* (1999), Astrophys. J. **523**, 350.
8. *P. H. Chavanis* (2000), Astron. and Astrophys. **356**, 1089.
9. *C. de la Fuente Marcos and P. Barge* (2001), MNRAS, **323**, 601.
10. *A. Johansen, A. C. Andersen and A. Brandenburg* (2004), Astron. and Astrophys. **417**, 361.
11. *H. H. Klahr* (2004), Astrophys. J. **606**, 1070.
12. *H. H. Klahr and P. Bodenheimer* (2003), Astrophys. J. **582**, 869.
13. *A. Bracco, P. H. Chavanis, A. Provenzale and E. A. Spiegel* (1999), Physics of Fluids, **11**, 2280.
14. *S. S. Davis* (2002), Astrophys. J. **576**, 450.
15. *S. S. Davis, D. P. Sheehan and J. N. Cuzzi* (2000), Astrophys. J. **545**, 494.
16. *G. Bodo, G. D. Chagelishvili, G. Murante, A. G. Tevzadze, P. Rossi and A. Ferrari* (2005), Astron. and Astrophys. **437**, 9.
17. *A. Mignone, G. Bodo, S. Massaglia, T. Matsakos, O. Tesileanu, C. Zanni and A. Ferrari* (2007), Astrophys. J. Supp. **170**, 228.
18. *F. Masset* (2000), Astron. and Astrophys. Supp. **141**, 165.
19. *G. D. Chagelishvili, A. G. Tevzadze, G. Bodo, S. S. Moiseev* (1997), Phys. Rev. Letters, **79**, 3178.
20. *B. M. Johnson and C. F. Gammie* (2005), Astrophys. J. **635**, 149.
21. *J. C. B. Papaloizou and C. Terquem* (2006), Reports of Progress in Physics, **69**, 119.
22. *C. Schafer, R. Speith, R. Gunther and W. Kley* (2004), Astronomische Nachrichten Supplement, **325**, 85.
23. *R. V. E. Lovelace, H. Li, S. A. Colgate and A. F. Nelson* (1999), Astrophys. J. **513**, 805.
24. *S. H. Lubow and G. I. Ogilvie* (1998), Astrophys. J. **504**, 983.

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Mechanics

Engineering Methods of Modeling of Oscillation Processes and Rational Design of Mechanical Systems

Robert Adamia*

* *Academy Member, R. Dvali Institute of Mechanics of Machines, Tbilisi*

ABSTRACT. Methodology of optimization synthesis of linear multimass mechanical systems oscillation processes and rational design of structural schemes is offered, allowing to choose rational relationship of machine transmission elastic-mass (inertial and stiffness) parameters ensuring their steady functioning with minimum dynamic coefficient at transition regimes. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *oscillation beatings, dimensionless generalized parameters, dynamic loads, vibration processes*

The basic scientific-technical task of modern mechanical engineering is connected with the creation of high productive and durable machines and constructions with low metal-intensivity.

The productivity of the machines and mechanisms depends upon the speed of their working parts and solving actual task of minimizing duration of transitional regimes and similar characteristics of technological cycles.

Unfortunately calculating and designing principal link parameters are mostly carried out on the basis of energy-force parameters and safety margin value without sound theoretical grounding and without taking into account dynamic characteristics of machines and peculiarity of transitional regimes.

If safety margin is taken more than is necessary it causes manufacturing of too heavy, massive machines and hence unjustified expenditure of metals. If safety margin is less than necessary, it leads to machine failure.

Modern industry (especially heavy engineering industry) is at present subjected to great losses due to fatigues failure of important parts of machines. The same is the case with mobile means of civil and defence purpose (units of transmission and lorries, trucks, tanks, vessel and aircraft engines of rotor type etc.).

The main reasons of failure are mistakes made at the initial stage of designing machines and incorrect selection of design parameters (mass and rigidity of links and quantity of their relationship), neglecting dynamic vibrating processes.

According to statistics, heavy engineering industry (particularly metallurgy) is subjected to considerable damage because of machine failure.

According to experimental data, 70% of failure is caused by internal resonance regimes in the mechanical and electromechanical systems, which has the character of oscillation beatings of elastic force amplitudes, and even at small technological loading fast fatigues breakage of important parts (especially transmission) is caused.

It is known that oscillation beatings arise at the approach of lowest values of natural oscillation frequencies, and the nearer these frequencies, the more is the negative effect. Coincidence of these natural frequencies causes the fastest breakage of a construction.

The picture of oscillation beating at nearness of frequency of natural oscillations ($\beta \approx \beta_2$) in torsion mechanical systems with two degrees of freedom is expressed in Fig. 1.

Unfortunately, this negative effect of oscillation beating is unknown to the majority of engineer-designers and they neglect this effect during designing.

According to the result of investigations, oscillation frequency is functionally connected with the values of the system constructive parameters, especially with the ratio of elastic and mass parameters.

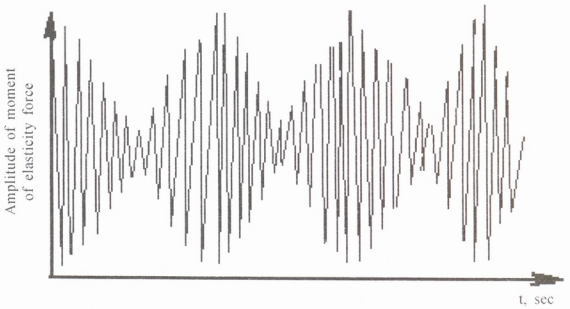


Fig. 1.

To determine natural oscillation frequencies, it is necessary to solve a characteristic equation of differential equations.

At high order equation (at design of real machine-aggregates), it is necessary to use computing technique. In this way a designer can determine the value of natural oscillations of the designed machines, but he cannot solve the problem of optimization synthesis.

Difficulties with these problems can easily be solved by an engineer method of optimizing synthesis of dynamic processes suggested by the author. The method is not to be solved with high order characteristic equation, and using easily attainable mathematics body, designers are able to determine systems generalized dimensionless parameters, proportional to the mechanical system and to construction parameters of the examined system.

To solve the problem it is necessary to consider high order differential equation [1,2]

$$\left.\begin{aligned} &X^{(2n)} + a_0 X^{[2(n-1)]} + \ldots + a_{n-2}\ddot{X} + a_{n-1}X = 0 \\ &(n = 1,2,3,\ldots), \end{aligned}\right\} \quad (1)$$

where $a_1, a_2, \ldots a_{n-1}$ are determined by the value of system parameters.

If argument t is substituted by value $t = \tau / a_0$, the following equation will be obtained:

$$X^{(2n)} + X^{[2(n-1)]} + C_1 X^{[2(n-2)]} + \ldots + C_{n-1}X = 0, \quad (2)$$

where C_k are generalized dimensionless parameters:

$$\left.\begin{aligned} &C_1 = \frac{a_1}{a_0^2}; C_2 = \frac{a_2}{a_0^3}; \ldots C_\cdot = \frac{a_{n-1}}{a_0^n}; \\ &(n = 1,2,3,\ldots). \end{aligned}\right\} \quad (3)$$

It is determined that the C_k parameter can alter in the limits:

$$0 \le C_k \le \frac{n-1}{2n}(k = 1,2\ldots n-1). \quad (4)$$

Passing on from equation (1) to equation (2) an important practical significance for simplified the problem of optimizing synthesis of mechanical systems.

In equation (1), the range of alteration of $a_1, a_2, \ldots, a_3$ determined by construction parameters, is very wide and indefinite. That is why, using this equation to solve the problem of optimizing synthesis, the designer would have to calculate infinite variants by using computer, which is uneconomic and unjustified.

Passing to generalized dimensionless parameters, C_k significantly narrows the range of alteration of differential equation coefficients and contracts the volume of computing.

For example, in the case of a differential equation of the eighth order, generalized parameters have the following maximum values:

$$C_1 = 3/8, C_2 = 1/16, C_3 = 1/256.$$

The values of coefficients a_i are determined by elastic and inertia parameters of the system.

For example, in the case of a three mass torsional system, values of a_0 and a_1 are calculated according to the formulas:

$$\left.\begin{aligned} a_0 &= C_{12}\frac{\theta_1 + \theta_2}{\theta_1\theta_2} + C_{23}\frac{\theta_2 + \theta_3}{\theta_2\theta_3}; \\ a_1 &= C_{12}\cdot C_{23}\frac{\theta_1 + \theta_2 + \theta_3}{\theta_1\theta_2\theta_3}, \end{aligned}\right\} \quad (5)$$

where C_{12} and C_{23} are rigidities of corresponding elastic shafts and – $\theta_{(i=1,2,3)}$ inertia moment of corresponding masses.

Substituting (5) values in equation (3) generalized dimensionless parameters can be determined, which, in turn, define the degree of nearness of natural oscillation frequencies, and hence, actual possibilities of the origi-

nation of oscillation beatings.

It follows from equation (4) that the value of generalized parameters for any degree of freedom will not exceed – 0.5.

Thus, if the degree of freedom n=100, which means that we are dealing with a differential equation of 200 degree, then:

$$\max C = \frac{n-1}{2n} = \frac{100-1}{200} \approx 0.5.$$

It follows from equation (4) that in the case of a differential equation of the fourth order ($n = 2$), variation of the generalized parameters takes place in the range

$$0 \le C_1 \le 0.25. \quad (6)$$

In the case of differential equation of the sixth order ($n = 3$)

$$0 \le C_1 \le 0.33, \quad 0 \le C_2 \le 1/27. \quad (7)$$

For equation of the eighth order

$$0 \le C_1 \le 3/8, 0 \le C_2 \le 1/16, 0 \le C_3 \le 1/256. \quad (8)$$

As a result of numerous experiments and a lot of computing, it has been ascertained that the higher the value of C_i, (which in turn attests to a high degree of nearness of low frequencies of natural frequency of oscillations of the system) depending on the ratio of construction parameters of mechanical systems of any degree of freedom, the sharper the oscillation beating in mechanical systems.

The maximum value of C_{1max}=0.25 – for a differential equation of the fourth order, C_{1max}=1/3 – for a differential equation of sixth order, causes the pure oscillation beating with the highest value of elastic force amplitudes in mechanical systems. In this case the smallest natural frequencies of the system become equal to each other ($\beta_1 \approx \beta_2$), that is, internal resonance has originated.

To avoid that phenomenon, the designer should try to remove the smallest frequencies of natural oscillation maximally from each other, that is, to withdraw the system from the zone of maximum values of generalized parameters.

The investigation has determined that the upper nonoptimal range for C_1 value exists for the mechanical systems with the degree of freedom two (n=2) under the range characterized by

$$C_1 = (0.18 \div 0.25). \quad (9)$$

For the system with the degree of freedom three (n=3)

$$C_1 = (0.25 \div 1/3). \quad (10)$$

For the system with the degree of freedom four (n=4)

$$C_1 = (0.33 \div 0.375). \quad (11)$$

In general, maximum permissible value of the generalized parameter C_1 should be 20% less than its maximum value, determined according to expression (4). Otherwise, at the initial stage of designing a designer must change the construction parameters ratio, so that the newly selected parameter values would ensure the withdrawing of C_1 from the range of nonoptimal values (9,11).

The author has determined that there exist the second (lower) range of nonoptimal values of parameters C_1 and for the systems with the degree of freedom two it is defined with the range

$$0 \le C_1 \le 0.04. \quad (12)$$

For the systems with the degree of freedom three (n=3)

$$0 \le C_1 \le 0.07. \quad (13)$$

For the systems with the degree of freedom four (n=4)

$$0 \le C_1 \le 0.09. \quad (14)$$

It is found that mechanical systems, whose ratio of construction parameters stipulates the lower nonoptimal value range of the generalized parameter C_1, are characterized by high sensibility (reaction) towards external, particularly impact forces. Minimum value of C_1 ensures the maximum removal of natural oscillation frequencies, caused by the increased characteristics of transmission elastic link stiffness, but such mechanical systems, at the same time, are characterized by high clearances which represent the impact type.

The range of lower nonoptimal values of generalized parameters is particularly dangerous for machines with reverse working regimes and are characterized by periodical opening and closing of clearances.

It is clear from the foregoing that there occurs optimal alteration of the range of generalized parameter, practical realization of which ensure a minimal reaction of mechanical systems to the influence of external forces of any type.

For mechanical systems with the degree of freedom two this range

$$0.05 \le C_1 \le 0.18. \quad (15)$$

In the case of a differential equation of the sixth order (n=3)

$$0.08 \le C_1 \le 0.25. \quad (16)$$

For an equation of the eighth order

$$0.12 \le C_1 \le 0.3. \quad (17)$$

Thus, the simplest method of engineer solving of the problem of synthesis of mechanical systems for dynamic processes optimization, ensuring minimal dynamic loading of machine aggregates, reliability and decreas-

ing of metal-intensivity, consists in the algorithm based on the simplest arithmetic operations, demanding definition of the elastic-mass parameters of the designed objects and calculation of generalized dimensionless parameters.

It should be noted that this method of optimization synthesis of dynamic processes of mechanical system is true for linear systems, requiring some corrections for objects with nonlinear characteristics.

The author has defined the values of elastic and inertia parameters for two hundred rolling mills and other metallurgical equipment in the former USSR Republics and calculated the values of generalized parameters.

Analysis shows that 23 out of 200 objects examined are characterized by nonoptimal values of generalized parameters. 21 objects have been found to be subjected to fatigue failure of responsible links of transmissions, leading to great economic losses.

Technological loading (rolling moments) of some of them was so small that the failure of basis parts of transmissions was paradoxical.

The suggested approach to solving the problem enables us to throw light upon an analogous phenomenon and carry out some modernization of many metallurgical and mobile objects. Among them Rustavi plate rolling mill 2100, Novolipetsk metallurgical plants 120 ton converter, plate rolling mill – 2000, working stand No 5 at Karaganda metallurgical plant, transmissions of the Kutaisi lorries, etc.

This reconstruction eradicated the emergency working regimes at the above-mentioned objects and increased the stability and the speed of working regimes.

It is worth noting that to realize the aim discussed above it turned out to be enough to decrease the rigidity of engine shafts (not to increase?!). This fact seems paradoxical from the traditional point of view.

Had the mentioned objects been designed according to the method suggested by the author, their metalintensivity would have been decreased by 30%, and, at the same time, their productivity and stability would have increased considerably.

On the basis of scientific analysis of the obtained data that engineering method of the synthesis of structural schemes of mechanical systems was worked out.

The relationship between the ratio of a system's construction parameters and generalized parameters have been determined.

The range of rational alternation of the mentioned parameters has been determined. Their practical realization ensures maximal removal of lowest natural oscillation frequencies of mechanical systems, leading to complete elimination of internal resonance (oscillation beating) and accordingly, minimization of dynamic loading.

With numerical examples the influence will be shown of the correlation of construction parameters and structural schemes of mechanical systems on the nearness of natural oscillation frequencies and consequently on the values of the generalized parameter C_1.

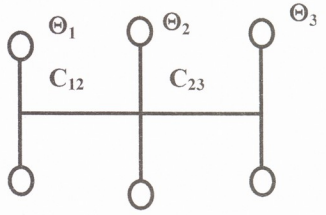


Fig. 2.

Let us consider several examples of different disposition and correlation of masses and rigidity of three mass equivalent torsion mechanical systems

1. $\theta_1 = \theta_2 = \theta_3 = 100$; $C_{12} = C_{23} = 10^4$;
2. $\theta_1 = \theta_3 = 100$; $\theta_2 = 1000$; $C_{12} = C_{23} = 10^4$;
3. $\theta_1 = \theta_3 = 100$; $\theta_2 = 10^4$; $C_{12} = C_{23} = 10^4$;
4. $\theta_1 = \theta_3 = 100$; $\theta_2 = 1000$; $C_{12} = 5 \times 10^4$; $C_{23} = 10^4$;
5. $\theta_1 = 10$; $\theta_2 = \theta_3 = 100$; $C_{12} = C_{23} = 10^4$.

It is known that the indicated three-mass system is described by a differential equation of the fourth order

$$X^{IV} + a_0 X^{II} + a_1 X = 0. \tag{18}$$

a_0 and a_1 are determined by expression (5)

In generalized dimensionless parameters, according to (2), expression (18) takes the form

$$X^{IV} + X^{II} + C_1 X = 0. \tag{19}$$

Then for case 1 we have:

$$a_0 = C_{12} \frac{\theta_1 + \theta_2}{\theta_1 \theta_2} + C_{23} \frac{\theta_2 + \theta_3}{\theta_2 \theta_3} = 400;$$

$$a_1 = C_{12} \cdot C_{23} \frac{\theta_1 + \theta_2 + \theta_3}{\theta_1 \theta_2 \theta_3} = 30000;$$

$$C_1 = \frac{a_1}{a_0^2} = \frac{30000}{160000} = 0.188.$$

The characteristic equation for case 1 in numerical form will be:

$$\lambda^4 - 400 \cdot \lambda^2 + 30000 = 0;$$

$$\lambda^2 = 200 \pm \sqrt{10000} = 200 \pm 100;$$

$$\lambda_1^2 = \beta_1 = 100; \quad \lambda_2^2 = \beta_2 = 300.$$

For case 2 we have

$$a_0 = 220; \quad a_1 = 12000; \quad C_1 = 0.248.$$

The characteristic equation

$$\lambda^4 - 220 \cdot \lambda^2 + 12000 = 0;$$
$$\lambda^2 = 110 \pm \sqrt{100};$$
$$\lambda_1^2 = \beta_1 = 100; \quad \lambda_2^2 = \beta_2 = 120.$$

Hence at increasing the generalized parameter C_1 natural frequencies of the oscillations of the system become close to each other, pointing to origination of beating of oscillations.

It will be recalled that for differential equations of the fourth order the maximum value of C_{1max}=0.25.

For the case 3 we have

$$a_0 = 202; \quad a_1 = 10200; \quad C_1 = 0.24998$$
$$\lambda^4 - 202 \cdot \lambda^2 + 10200 = 0;$$
$$\lambda^2 = 101 \pm \sqrt{1};$$
$$\lambda_1^2 = \beta_1 = 100; \quad \lambda_2^2 = \beta_2 = 102.$$

Thus, when the structural scheme is chosen unsuccessfully, i.e. when the middle mass is considerably (100 times) greater than extreme masses, the generalized parameter C_{1max}=0.24998≈0.25. Therefore natural frequencies become equal to β_1=100; β_2=102, which is the cause of origination of pure oscillations beating (internal resonance).

For case 4 we have

$$a_0 = 660; \quad a_1 = 60000; \quad C_1 = 0.137$$
$$\lambda^4 - 660 \cdot \lambda^2 + 60000 = 0;$$
$$\lambda^2 = 300 \pm \sqrt{48900};$$
$$\lambda_1^2 = \beta_1 = 109; \quad \lambda_2^2 = \beta_2 = 551.$$

For this example we see that with a 5-fold increase of rigidities C_{12} optimization of oscillation processes is ensured at the expense of removal of natural frequencies from each other to a considerable degree.

The criterion for the evaluation of the removal of natural frequencies from each other is a quantity of generalized parameter C_1, which for the case under consideration, is equal to C_1=0,137; i.e. it satisfies the condition of optimality determined by expression (15).

For case 5 we have

$$a_0 = 310; \quad a_1 = 12000; \quad C_1 = 0.124$$
$$\lambda^4 - 310 \cdot \lambda^2 + 12000 = 0;$$
$$\lambda^2 = 155 \pm \sqrt{11025};$$
$$\lambda_1^2 = \beta_1 = 50; \quad \lambda_2^2 = \beta_2 = 260.$$

It is clear that another effective means of obtaining the desirable value of generalized parameter, hence re-

Table 1.

Values of generalized parameter for different variants of fulfilment of three-mass design schemes

Variants	Correlation (ratio) of mass and rigid parameters (dimensionless values)					Generalized parameter C_1
	θ_1	θ_2	θ_3	c_{12}	c_{23}	
1	1	1	1	1	1	0.1875
2	1	2	1	1	1	0.22
3	1	5	1	1	1	0.243
4	1	10	1	1	1	0.248
5	5	1	1	1	1	0.135
6	2	1	1	1	1	0.164
7	3	1	1	1	1	0.150
8	10	1	1	1	1	0.125
9	2	1	2	1	1	0.140
10	3	1	3	1	1	0.110
11	5	1	5	1	1	0.076
12	10	1	10	1	1	0.043
13	1	1	1	2	1	0.090
14	1	1	1	3	1	0.062
15	1	1	1	2	2	0.187
16	1	1	1	5	5	0.187
17	2	1	1	2	1	0.040
18	2	1	1	2	2	0.163
19	1	10	2	2	1	0.220
20	1	10	1	5	1	0.135
21	1	10	1	10	1	0.082
22	1	5	1	3	1	0.082
23	1	5	1	5	1	0.134

moval of natural frequencies of oscillation, is to increase the correlation of the moments of mass inertia.

A more acceptable means of obtaining the necessary value of generalized parameter C_1, from the viewpoint of practical realization, is to increase the correlation of the rigidities of elastic links.

Numerous examples of calculations have shown that to prevent oscillation beatings it is enough to choose a correlation of rigidities equal to two and more. In such a system no internal resonance will arise.

Numerical examples of calculations of three-mass torsion mechanical systems have shown that symmetrical schemes are nonoptimal; extremely dangerous are structural schemes with middle greater mass. However, such schemes are optimal for the designer of vibromachines.

Values of generalized dimensionless parameters defining the level (values of amplitude) and the character of change of dynamic loadings (moments of elastic force) at different magnitudes of moments of inertia of mass and rigidity are given in Table 1.

Consideration of Table 1 shows variants of relationship of parameters that satisfy the conditions of optimality (15).

For four-mass torsion mechanical systems (Fig. 3) change of moments of elastic force is described by the differential equation of the sixth order

$$M_{i,i+1}^{IV} + a_0 M_{i,i+1}^{IV} + a_1 M_{i,i+1}^{II} + a_2 M_{i,i+1} = 0 \qquad (20)$$
$$(i = 1,2,3)$$

In generalized dimensionless parameters, according to (2), expression (20) takes the form

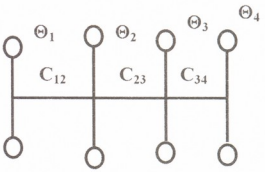


Fig. 3.

$$M_{i,i+1}^{IV} + M_{i,i+1}^{IV} + C_1 M_{i,i+1}^{II} + C_2 M_{i,i+1} = 0 \qquad (21)$$
$$(i = 1,2,3)$$

In these expressions the coefficients a_i (i=0,1,2) and C_i (i=0,1,2) are defined

$$a_0 = C_{12}\frac{\theta_1+\theta_2}{\theta_1\theta_2} + C_{23}\frac{\theta_2+\theta_3}{\theta_2\theta_3} + C_{34}\frac{\theta_3+\theta_4}{\theta_3\theta_4};$$

$$a_1 = C_{12}\cdot C_{23}\frac{\theta_1+\theta_2+\theta_3}{\theta_1\theta_2\theta_3} + C_{23}\cdot C_{34}\frac{\theta_2+\theta_3+\theta_4}{\theta_2\theta_3\theta_4} + C_{12}\cdot C_{34}\frac{\theta_1\theta_4+\theta_2\theta_3+\theta_1\theta_3+\theta_2\theta_4}{\theta_1\theta_2\theta_3\theta_4};$$

$$a_2 = C_{12}C_{23}C_{34}\frac{\theta_1+\theta_2+\theta_3+\theta_4}{\theta_1\theta_2\theta_3\theta_4}, \qquad (22)$$

$$C_1 = \frac{a_1}{a_0^{\frac{2}{3}}}; \qquad C_2 = \frac{a_2}{a_0^{\frac{1}{3}}}.$$

Table 2.

Values of generalized parameter for different variants of fulfilment of four-mass design schemes

Variants	Dimensionless parameter							Generalized parameter	
	θ_1	θ_2	θ_3	θ_4	c_{12}	c_{23}	c_{34}	c_1	c_2
1	1	1	1	1	1	1	1	0.277	0.018
2	10	1	1	1	1	1	1	0.250	0.010
3	1	10	1	1	1	1	1	0.260	0.015
4	10	10	1	1	1	1	1	0.240	0.006
5	1	10	10	1	1	1	1	0.280	0.016
6	10	1	1	10	1	1	1	0.150	0.003
7	5	1	1	5	1	1	1	0.180	0.0056
8	1	1	1	1	3	1	1	0.240	0.0120
9	1	1	1	1	5	1	1	0.190	0.0073
10	1	1	1	1	10	1	1	0.120	0.003
11	1	1	1	1	1	3	1	0.220	0.0120
12	1	1	1	1	1	5	1	0.170	0.0036
13	1	1	1	1	1	10	1	0.110	0.003
14	1	10	10	1	1	1	1	0.260	0.010
15	1	10	10	1	1	1	1	0.145	0.0035
16	1	10	10	1	1	1	2	0.275	0.0090
17	1	10	10	1	1	1	5	0.268	0.0071

Values of generalized dimensionless parameters C_1 and C_2, defining the level and the character of elastic force (M_{12}, M_{23}, M_{34}) in a four-mass torsion mechanical system are given in Tab. 2.

After using data given in the table the engineer easily ascertains what kind of relationship of moments of inertia and rigidity satisfy the conditions of optimality (16), at keeping of which minimization of dynamic loads and prevention of inner resonances will be provided.

In Tables 1 and 2 the values of digits 1…10 denote not absolute values of inertia masses and rigidity but their relationship. That is why the digits have no dimention.

One can conclude that the level and the character of change of dynamic loads in mechanical and electromechanical systems (in transmission of machines and mechanisms) depends not on real meanings of the parameters of the system but on their relationship, i.e. on the structure (by structure and order of position inertia and elastic parameters) of the objects under study.

მექანიკა

მექანიკური სისტემების რხევითი პროცესების და რაციონალური დაპროექტების მოდელირების ინჟინრული მეთოდები

რ. ადამია*

* *აკადემიის წევრი, რ. დვალის მანქანათა მექანიკის ინსტიტუტი, თბილისი*

ნაშრომში ანალიზური და ორიგინალური საინჟინრო მიდგომებისა და მეთოდების შემუშავებით გადაწყვეტილია მოძრაობის მრავალი თავისუფლების ხარისხის მქონე როტორული ტიპის მექანიკური დინამიკური სისტემების (სამანქანო აგრეგატების) გარდამავალი (რხევითი) პროცესების ოპტიმიზაციისა და რაციონალური დაპროექტების სამეცნიერო-ტექნიკური ამოცანები, რაც უზრუნველყოფს ანალოგიურ სისტემებში დინამიკური დატვირთვების მინიმიზაციას, ვიბრაციულ-რეზონანსული (რხევათა ცემის) მოვლენების აღმოფხვრას და ამით მათი საიმედოობისა და მედეგობის მნიშვნელოვან ამაღლებას.

REFERENCES

1. *R. Adamia* (1999), Dynamics of Machines. Tbilisi, 400 p. (Russian).
2. *R. Adamia, F. Mikhailenko* (1989), Limitations of Dynamic Loading in Machines. Tbilisi, 180 p. (Russian).

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Geophysics

Increase of Order of Local Seismic Energy Release Caused by Water Level Variation in the Enguri High Dam Reservoir

Vakhtang Abashidze*, Zurab Javakhishvili*, Thea Khutsishvili*, Ekaterine Mepharidze*, Natalia Zhukova*, Tamaz Chelidze, Teimuraz Matcharashvili***

* *M. Nodia Institute of Geophysics, Tbilisi*
** *Academy Member, M. Nodia Institute of Geophysics, Tbilisi*

ABSTRACT. **It was reported earlier that water level periodic variation in a high dam reservoir may lead to the synchronizing influence on local seismic activity. Taking into account that phase synchronization is not related to strong functional relationships, in the present study small dynamical changes in seismic activity caused by water level variation were investigated.**

For this purpose the recurrence quantitative analysis (RQA) approach was used. Analysis was carried out on data sets of water level daily variation and released daily seismic energy.

It was shown that when the external influence on the earth's crust caused by reservoir water becomes periodic the extent of the regularity of earthquake daily distribution essentially increases. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *dynamics, phase synchronization, seismicity, water reservoir.*

In the present work the character of dynamical changes in the regional seismic activity under phase synchronizing influence of water level variation in a large reservoir was investigated.

Investigation of dynamical aspects of the influence of large water reservoirs on the generation of local earthquakes is one of the well recognized scientific and practical problems [1-3]. Reservoir induced seismicity (RIS) has been observed at many reservoirs. At the same time, the character of influence in the time period following the initial RIS activity is still an open problem.

In our previous research based on field and laboratory data the evidence was presented showing that the decrease of seismic energy after RIS period may be caused by the periodic variation of the water level in the large reservoir [4]. As far as we deal with the weakest form of synchrony - phase synchronization [5], in the present work we aimed to investigate the character of dynamical changes in regional seismic activity caused by the influence of water level periodic variation in the large reservoir.

The data sets of water level variation and seismic activity around Enguri high dam, used in the present study, were obtained from data bases of the M. Nodia Institute of Geophysics (Georgia). Namely, daily water level variation data were gathered at the Enguri high dam reservoir located in Western Georgia, Caucasus (42.030 N, 42.775 E) in 1973-1995. Data sets of daily occurred earthquakes and released daily seismic energy were taken for magnitude threshold $M \geq 1.6$ within 90km around the reservoir. Time series of sequences of magnitudes and time intervals between consecutive earthquakes, unevenly sampled for 1973-1995, were analyzed.

To investigate dynamical changes in target processes the Recurrence Quantitative Analysis (RQA) method was used. RQA is convenient for nonstationary

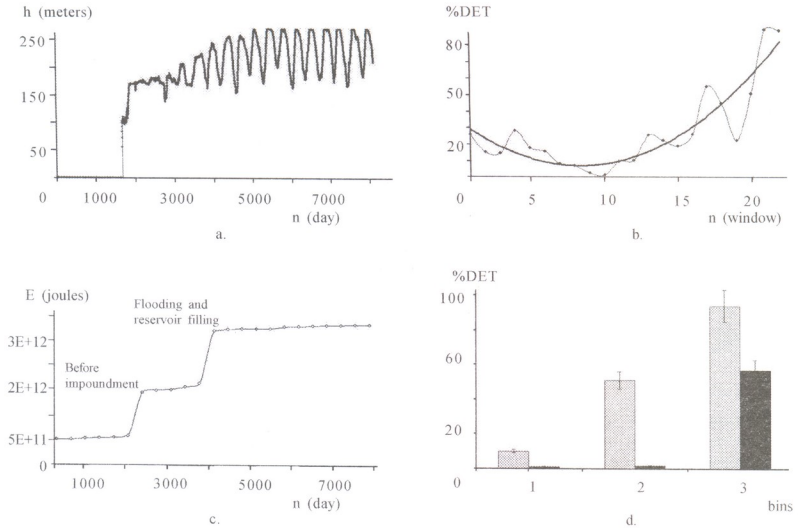


Fig. 1. a. Variation of water level in Enguri high dam reservoir above sea level in 1978-1995. b. RQA %DET of daily number of earthquakes calculated for consecutive one year sliding windows. c. Cumulative sum of released daily seismic energy. d. RQA %DET of magnitude (black columns) and interearthquake time interval (grey columns) sequences (1) before impoundment, (2) during flooding and reservoir filling and (3) periodic change of water level in reservoir.

and short real data sets [6]. In the present study recurrence determinism - *DET*(*t*) measure based on the analysis of diagonally oriented lines in the recurrence plot has been calculated [7].

According to recent publications, the dynamics of earthquake-related processes in the earth's crust are recognized as non-random, having both low and/or high dimensional nonlinear structures [8, 9]. One of the characteristic properties of processes in non-random systems, which are close to the critical state, is their high sensitivity to initial conditions as well as to relatively weak external influences. This is a general and very important property of practically unpredictable seismic processes. Indeed, insofar as we are not able to govern the initial conditions of lithospheric processes, even possibility in principle of controlling of the dynamics of seismic processes has immense scientific and practical significance. The way towards understanding such control mechanism lies through investigation of the dynamics of seismic processes, when a small external influence leads to phase synchronization.

The phase synchronization between water level periodic variation and seismic activity, observed in our previous and present studies, is recognized as the weakest form of synchrony when interacting nonlinear oscillators remain largely uncorrelated. Generally, interacting systems may have different dynamical features [5]. It is most important that, contrary to other forms of synchrony which lead to an increase of order in the behaviour of the synchronized system, phase synchronization does not require strong coupling between involved processes. This in turn means that the presence of order and character of the changes in dynamics of a phase synchronized system is not obvious.

We used RQA to investigate the character of dynamical changes in seismic process. As follows from our RQA results, when an external influence on the earth's crust caused by a water reservoir becomes periodic, the extent of the regularity of earthquake daily distribution (estimated as %DET) essentially increases (see Fig. 1b, bold line). This result was tested by comparing it with the surrogate data. It is important to mention that the influence of an increasing amount of water and its subsequent periodic variation essentially affects also the character of earthquakes magnitude and temporal distribution (see Fig. 1,c). The extent of order in the temporal (Fig. 1d, black columns) and magnitude (grey columns) distribution of the earthquakes, calculated as the value of %DET, substantially increases when the reservoir influence becomes periodic. Results of %DET calculation of corresponding surrogates are always less than 50%

of the original values (not shown here). It is interesting to note that the dynamics of earthquakes temporal and energetic distributions changes even under water level irregular variation, though not so much as under periodic variation.

Based on the results of our investigation, we conclude that the order in the dynamics of earthquakes daily occurrence, as well as in earthquake's temporal and energetic distributions increases when water level variation becomes periodic.

გეოფიზიკა

ენგურის წყალსაცავში წყლის დონის ვარიაციით გამოწვეული ლოკალური სეისმური ენერგიის გამოთავისუფლების მოწესრიგებულობის ხარისხის ზრდა

ვ. აბაშიძე*, ზ. ჯაფახიშვილი*, თ. ხუციშვილი*, ე. მეფარიძე*, ნ. ჟუკოვა*, თ. ჭელიძე**, თ. მაჭარაშვილი*

* *ი. ჯავახიშვილის გეოფიზიკის ინსტიტუტი, თბილისი*
** *აკადემიის წევრი, ი. ჯავახიშვილის გეოფიზიკის ინსტიტუტი, თბილისი*

ჩვენს წინანდელ ნაშრომებში ნაჩვენები იყო, რომ დიდ წყალსაცავში წყლის დონის პერიოდულმა ვარიაციამ შესაძლოა გამოიწვიოს ლოკალური სეისმური აქტიურობის ფაზური სინქრონიზაცია. იმის გათვალისწინებით, რომ ფაზური სინქრონიზაციისას პროცესებს შორის ფუნქციონალური კავშირი სუსტია, ჩვენ ამ ნაშრომში მიზნად დავისახეთ შეგვესწავლა თუ როგორ იცვლება სეისმური პროცესის დინამიკური მახასიათებლები წყალსაცავში წყლის დონის ვარიაციის სხვადასხვა რეჟიმში.

ნაჩვენები იქნა, რომ წყალსაცავში წყლის დონის პერიოდული ვარიაცია განაპირობებს მიწისძვრების რაოდენობისა და გამოყოფილი ენერგიის განაწილების პროცესის მოწესრიგებულობის ხარისხის ზრდას.

REFERENCES

1. *D. Simpson, W. Leith, C. Scholz* (1988), Bulletin of the Seismological Society of America, 78, 2025-2040.
2. *M. Assumpção, V. I. Marza, L.V. Barros, C. N. Chimpliganond, J. E. P. Soares, J. M. Carvalho, D. Caixeta, A. Amorim, E. Cabral* (2002), Pure Applied Geophysics, 159, 597–617.
3. *A. F. Nascimento, P. A. Cowie, R. J. Lunn, G. Pearce* (2004), Geophysical Journal International, 158, 1041-1052.
4. *J. Peinke, T. Matcharashvili, T. Chelidze, J. Gogiashvili, A. Nawroth, O. Lursmanashvili, Z. Javakhishvili* (2006), Physics of the Earth and Planetary Interiors, 156/1-2, 130-142.
5. *D. Pazo, M. A. Zaks, J. Kurths* (2003), Chaos, 13, 309-318.
6. *J. P. Zbilut, C. L. Jr. Webber* (1992), Physics Letters A, 171, 199-203.
7. *N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths* (2002), Physical Review E, 66, 026702.1-026702.8.
8. *T. Matcharashvili, T. Chelidze, Z. Javakhishvili* (2000), Nonlinear Processes in Geophysics, 7, 9-19.
9. *J. Rundle, D. Turcotte, W. Klein* (2000), (eds), Geocomplexity and the Physics of Earthquakes, American Geophysical Union, Washington.

Received February, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Hydrology

Measures for Predicting Accidents at Hydraulic Engineering Structures and their Prevention

Tsotne Mirtskhoulava*

* *Academy Member, Institute of Water Management, Tbilisi*

ABSTRACT. The paper deals with the urgency of research on accidents at hydraulic engineering structures. A brief description of some accidents is given and an approach to their prediction is proposed through the use of A.N. Kolmogorov's works devoted to defects. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *accident, reliability, risk, defect, prediction.*

The vast natural and technospheric accidents happening in recent decades have justly troubled the world community. The population is especially shocked by the fact that these incidents were attended by hundreds and thousands of human casualties. Dozens of descriptions of the disasters have been published. Significantly enough, the hitherto unknown sites of accidents (Soneso, Flicksoro, Bhopal) have become known through disasters occurring there and are frequently mentioned as symbols of hazard.

The population of many countries, including that of Georgia, cannot help being concerned about the fact that, notwithstanding the effective measures taken, no reduction of the losses resulting from accidents and disasters is noticeable; on the contrary, these losses are appreciably growing.

Special note should be taken of the increase of losses resulting from floods and high waters throughout the world, including Georgia. At the same time, the question of potential accidents at hydraulic engineering structures has become acute. This is accounted for by the fact that the term of service of these facilities has reached a critical threshold, entering the phase of obsolescence. Their majority were built 40 to 50 years ago, i.e. they have exhausted their viability resource. Practice shows that, of other existing structures, accidents at water development works are most hazardous for the population and the environment owing to the vast scale of damage caused by them; suffice it to recall the horrific tragedy caused by the Vaiont dam in Italy.

The foregoing leads one to the conclusion that averting accidents at hydraulic engineering structures and mitigation of the losses (exceeding tens of millions of dollars, to say nothing of human casualties) inflicted by them demand more attention on the part of scientists and must be given priority among the questions studied by technical sciences.

Thus, study of accidents at hydraulic-engineering structures aimed at working out measures for averting such disasters or mitigating the damage inflicted by them should unconditionally be considered the most urgent and vital problem to be investigated as the first priority.

Among the novelties awaited by the population from science, development of effective measures of combating accidents at structures is doubtless accorded one of the first places among those bringing useful satisfaction. As the Japanese, who are advanced in many spheres, teach us, we must always feel satisfaction from our work done. If this is not the case we must do so as to create such a situation. There can be no doubt that measures for averting accidents will evoke great satisfaction among broad strata of the population. Determination of the probable frequency of accidents and of the time of their occurrence should be considered one of the first measures of combating them. Knowledge of this important characteristic enables timely and quality preparation for combating accidents. This will minimize

the damage caused to the population and the environment.

There is no doubt that to date science stands heavily in the people's debt, for the population is under constant fear and threat of accidents, latter more often than not developing into catastrophes.

Notwithstanding the existing achievements in the study of this highly complex problem [2-10], many theoretical and practical tasks, which will shorten the long black list of disasters, stand in need of immediate in-depth study through wide and profound use of the latest achievements of different branches of science. Among the multiplicity of problems, whose solution is required from science by practice and safe economic development, study of the mechanism of the origin of accidents at structures, prediction of their frequency and working out measures – on the basis of these data – for averting them at hydraulic-engineering structures or for mitigating the damage inflicted should be considered a primary task.

In foreign specialist literature, to underline the magnitude of the damage caused by accidents at hydraulic-engineering structures, including great dams, the latter are referred to as hydrological bombs. Of course the simile is exaggerated, yet not a single year passes without the occurrence of an accident of the named structure in one or another "developed" country. That is why this question is given special attention at such forums as international forums on large dams and international congresses of the IAHR. The author is a direct witness of this who, until recently, was participant of these congresses and among their organizers. Storage reservoirs and their inseparable part - dams – are considered one of the most accident-prone facilities.

Numerous other structures occur in practice that may break down and inflict enormous losses to the environment and the economy. In recent times the probability of the occurrence of accidents has increased owing to the growth of the share of obsolescent structures. The substantial experience and knowledge, accumulated in the design and construction of dams, and progress in this direction should by no means blunt the attention of designers, builders, operatives and scientists to the safety of these major facilities. We should not be oblivious of the fact that many specialists believed atomic power stations to be absolutely safe. Yet, unfortunately, practice disproved this belief. However, it should be noted that today, too, atomic stations are highly reliable, so that starting their construction should not cause alarm.

The level of understanding of the reliability of safe functioning of hydraulic engineering structures has substantially grown, following the author's publications [2-10]. These publications encompass data on the prediction of the reliability and maintenance of hydraulic engineering structures of various purposes. However, his first study only mapped out the way of making a decision, and the necessity of expanding research for the development of methods of assessment of the risk of accident situations, without indication of the basic methods of the study of accidents, especially of measures towards predicting accidents and accident-prone situations and averting them. The results of research into the causes of accidents along these lines allow to predict a disaster with considerable approximation. For their part, these data facilitate a better understanding of the regularity of the functioning of designs and structures, directing the attention of operatives, scientists, designers and builders to improving the quality of design and construction; this will substantially ensure the prevention of accident-prone situations as well as accidents proper.

To achieve the goal set, it will be necessary to study this most complex problem, and primarily accidents at hydraulic engineering structures – analysis of measures for averting this phenomenon or mitigating the damage; this will call for the solution of a number of tasks, including: 1. Assessment of the pre-accident period; 2. Analysis of the process of accidents of structures; 3. Search for ways of predicting accidents; 4. Approximative prediction of accidents, using modern methods of the theory of random processes.

The principal steps to be taken towards averting accidents or mitigating the damage inflicted entail a number of theoretical, organizational and practical measures. This primarily requires the study (even approximative) of the frequency of the accidents or catastrophes at diverse-purpose facilities.

Judging from the literary sources available to the author, the available methods are not perfect either in this country of abroad, hence the urgent need for further studies, refinement of the existing results and development of a new approach.

Analysis of the accidents that have occurred will allow to develop designs that will be subject less to accident phenomena and their "vitality" will substantially increase.

Not only disintegration of the design should be considered an accident state of a structure but also the state in which normal exploitation of the facility is rendered difficult, when the design has cracks, major bends, the reinforcement is strongly corroded, or other similar defects are evident.

That the proposed method will allow consideration of the time factor in prediction should be considered a significant new approach. The existing traditional methods fail to take this parameter into account.

Although the chances of major accidents as well as catastrophes have diminished in recent years through the introduction of new perfect methods of calculation of construction and assembly work and exploitation, they,

regrettably, still occur – and not infrequently. Therefore it is impermissible to neglect this problem. In other words, research in this sphere is extremely urgent.

More often than not accidents occur at hydraulic engineering facilities, inflicting appreciable damage to the country s economy, as well as to ecological stability. According to the material gathered by the International Committee on the Breakdown of Dams, of the 1150 cases of damage occurring in 35 countries of the world, the following were found to be damaged: concrete dams and their bases (21%); rock-fill dams with their bases (2-4%); earth-fill dams with their bases (36%); damage in the reservoir zone (3-4%); damage in the downstream wall (0-5%). Analysis of these and other data [2-9] points to the occurrence of accidents at earth-fill dams as well, built in 1900-1919; suffosional processes proved most hazardous for the stability of earth-fill dams. Observation carried out for 200 years in Britain, involving 2000 earth-fill dams, showed that here, too, suffusion processes take place, caused by aging. Thus, the causes responsible for accidents may be divided in the following way: suffusion: 55%; spillover: 14%; other factors: 31%.

An accident occurred in Georgia too, involving a small dam at Tsqneti, which, regrettably, claimed 7 human lives. The commission that studied its causes was headed by the present writer and it is extensively discussed in his monograph published in Russia.

To demonstrate the acuteness of the problem, here are a few examples of dam accidents widely talked about world-wide. To be sure, the accidents at hydraulic engineering structures did not have such resources as power supply accidents had. Thus, in 1979 a grave accident occurred at two units of the atomic power station on Three-Mile-Island, USA; in 1984 a major explosion occurred at a chemical plant at Bhopal, India; the crash of the "Challenger" multiple use spacecraft with seven cosmonauts on board; special mention should be made of the explosion at the Chernobyl Atomic Power Plant block IV in April 1986. Obviously, these and similar accidents cannot be overlooked. The above cases, occurring in different countries of the world, point out that the process is intensified owing to the aging of structures. The principal cause of accidents at hydro-electric stations is aging as well. If we exclude accidents in extreme situations as well as those that occur in the first years of exploitation, then a special risk occurs at the stage of reservoir filling.

The following phenomena have been found to take place at the aging of earth-fill dams: the overall and local strength diminishes; at aging the strength of various materials (earth, concrete, polymer films, etc.) weakens. In this case, biodamage is frequent (destructive impact of bacteria, chemical and mechanical degradations; fungi, rodents, mollusks become active). In the course of time various metal items suffer corrosion, while wooden parts rot.

During exploitation filtration and suffusion processes are noticeable in dams. Alterations in a dam may be determined by analysis of the physical characteristics of water, e.g. the presence of sediment, opacity, colour, smell, electrical conductivity. These data allow judgment of the damage of concrete, reinforced and metal constructions by chemical and mechanical suffusion, and of the presence of microorganisms.

Chemical analysis of water and its composition – acidity, content of iron ions, chlorine, sulphates, etc. – yields information on the presence of an excessive amount of chemical substances in the filtration and suffusion process. With the increase of the age of a dam, as a rule, negative phenomena intensify, in the tail race: the depth of local scour and washout of the structure increases; the length of scour increases in the tail race; as a result of the impact of waves the cyclic processes of the freezing and melting of water become more frequent.

The experience of combating accidents and catastrophes shows that economic losses may be reduced to a certain extent through scientifically-grounded prediction. This allows to adopt an adequate measure to prevent the rise of causes that provoke accidents. In this connection it should be noted that in the 1970s in European countries accidents showed a tendency to grow, and they became more frequent, with human casualties. An international commission was set up and a special measure was worked out, under the name of the "Seveso Directive". This Directive is obligatory for all countries of Europe, envisaging the submission of a Report on Safety by the organization that possesses an accident-hazardous facility. It is highly interesting that following the introduction of this measure, from 1983, the number of accidents dropped appreciably, e.g. in 1986 it fell from 400 to 150, and in 1988 to 50.

Precise prediction of accident situations is not feasible owing to the multiplicity of their causes. No universal expression can exist for its prediction; the more so, we cannot expect to obtain exact analytic expressions. This task is unresolved at the present stage of scientific development. We may search for an approximative analytic dependence to identify accidents of individual type or groups. It should be emphasized specially that all accidents are determined by one principle: "Fortuitous accidents do not exist". Thus, each accident has its causative factors, with the sources and causes of these factors. Leading among these causes are the errors made by engineers, designers and superintendents of work. The probability of accidents caused by these errors amounts to 90% [1]. Although these errors are indirectly taken into account in designing by increasing the reliability level, yet causes still remain that can provoke accidents.

Until recently, along with complexity and multifactorialness, the lack of study of the regularity determining the state of reliability of hydraulic engineering facilities and the difficulty in obtaining initial information may be given as the reason for the absence of a prognostic method for the assessment of a safe state.

The Nobel Prize winner, the American Robert Aumann considers simplification of the possibility of obtaining information needed for research to be one of the principal criteria of assessing the modern achievements of science. This is not surprising, as we happen to live in an age of information that has reached such a level that it is considered the fourth factor of production, while the other three factors in conditions of the market economy are: land, labour and capital.

Ensuring the reliable exploitation of facilities calls for regular identification of the weak parts, units and most critical points of a facility, from the viewpoint of its failure. Timely detection of expected damage is considered the principal condition of ensuring the reliability of facilities.

Recently, we have come to face the necessity of solving a number of fresh problems posed by the increase of the number of obsolescent structures. In the new economic conditions this calls for an improvement of the maintenance of these facilities; also taking special measures towards giving advance warning about an expected breakdown of structures of this category. To this end, it is necessary to determine the critical moment that precedes the accident at the facility. This in turn will allow timely prolongation of the term of exploitation through taking adequate measures; the extended term will correspond to the existing physical state of the facility. In order to select the index of degradation of properties at each stage of exploitation of the facility it is important to identify such a prognostic variable that fully reflects the physico-mechanical parameters of the current state of the facility.

The breakdown of a facility may be forecast through timely assessment of its state of reliability – by identifying the critical moment of its weakening and decline of its tolerance to the environment and variable loads. This information enables one to select a lightened regime of exploitation – to put it figuratively, "*to allow facilities to leave the scene honourably, following their fulfillment of all the conceptions reflected in their design*".

Operative assessment of the state of a facility permits to limit the term of its exploitation, when such a need arises; an extension of the term allows to select a suitable regime of exploitation (reducing the level of load, depending on the intensity of damage, and so on). Operative assessment of the state of reliability of hydraulic engineering facilities is especially important in determining the moment of breakdown of obsolescent and weakened facilities; to work out measures for full utilization of their resource potential, and, if need be, to write them off (to avert an accident).

In such an extremely complex situation, search for a general expression for predicting the rise of an accident situation, accident or catastrophe is as a rule inconceivably difficult – short of impossible. Prediction is feasible only in individual simple cases.

As shown by analysis, accidents of engineering structures constitute a genuinely random process. Hence it is natural to use modern means of the theory of random processes to investigate this question. To this end it is advisable to find an approach that will allow to predict the phenomenon on the basis of data on past incidents, using the experience of the exploitation of same-type facilities. This is in the first place feasible by the use of powerful and at the same time elegant methods, proposed by the well-known scientist A.A. Markov, predecessor of the outstanding 20th century mathematician A.N. Kolmogorov. As we shall see further, the task is solved broadly and profoundly enough on the basis of Kolmogorov s formulae [1,2-9].

Similarly to the solution of many applied problems, here too, the task may be reduced to a reasonable use of the arsenal of corresponding mathematical means. As at present many analytical methods of solution have been worked out mostly for Markov processes, the striving is but natural to adapt many problems to the apparatus of the theory of Markov processes. I have chosen precisely this path in working on this topic.

As demonstrated by an analysis of the experience, accidents often occur in the work of some design suddenly, without any malfunction having been noticed. It is mostly in this way that metal and wooden constructions lose their stability.

Analogously to solving problems set in other branches of engineering, a model is proposed in which use is made of approximative estimations of numerical parameters, and account will be taken only of the main prognostic variables of the system. In this situation it is acceptable and profitable to use mathematical models that are successfully employed in solving other analogous problems. Fruitful use of fundamental biological principles, physical and other regularities in different branches of engineering is common knowledge. The view set forth allows to presume that the way proposed below is not only formalized, which incorrectly describes the process in "mathematical language". However, I should like to note that employment of stochastic models in practice – based on a single parameter – can express only the average indices of a system. It can give us only an approximative estimation of the functioning of the design, taking into account the factors of environmental impact and load during the exploitation of facilities.

Of the various available techniques of reliability theory, the methods of Markov processes [7] should be

considered the most effective, in particular, the use of the stochastic kinetic equation – widely used in physics in analyzing diffusional and other analogous processes.

Of course, in the study of a process it is advisable to take into consideration all the characteristics of the functioning of the system to be assessed. However, this is practically unfeasible. Hence, the identification of such a parameter should be considered one of the important stages in solving a problem, which is the principal prognostic variable among others and which will describe the functioning of the system under analysis with relative fullness. The results obtained will allow to assess the safety and permissible risk.

In working out measures for averting accidents it is logical and advisable to use systems analysis: - an approach widely used at present.

As noted above, gathering of observation data is an important stage of the study of an accident. But this is complex and labour-consuming, hence in the case of insufficiency of these data use may be made of the bootstrap method, techniques of Monte Carlo and expert assessments, as resorted to by the present writer in his other works [7].

It is expected that the recommendations worked out will enable a more objective determination of the length of exploitation of structures. The experience of exploitation of facilities shows that the terms set for the exploitation of many types are rather small, resulting in the premature stoppage of the facility, pointing to misused finances spent on their development, production and exploitation. An expression can be derived that will enable the assessment of the viability of a facility in conditions of exploitation. Account should be taken of the fact that, although some hydraulic-engineering structures still continue to function after the breakdown of some of their elements, their efficiency is low. It is said about such systems that they are reserved in excess, and the facility is of little viability. It will be recalled that the viability of a system is its property to perform its functions in the case of undesirable impact of outside forces on it, not envisaged under conditions of normal exploitation. Seismic and flood impacts are examples of such action.

Prediction of accidents by the proposed method will yield an answer related to the initial information (table of observations) used in the calculation. Thus, e.g., if sequences of accident situations are used, computations will yield the probable time of onset of an accident situation, as well as the level of risk of the occurrence of this phenomenon. At the use of observation data on accidents that have occurred, the prediction will yield the probability of accidents happening at a given time. The prognostic data obtained may be used in developing relatively perfect constructions and in raising the quality of structures. It should be noted that exact prediction of the hazard of accident situations and accidents is an extremely difficult – and often unfeasible – task. Therefore, not only an approximative expression should be considered acceptable that answers this import question but one that allows rough qualitative estimation of the frequency of these occurrences, and not qualitatively, in linguistic terms ('will occur frequently', 'very often', 'seldom', and so on).

Numerous observations of different facilities of various purposes give ground to assume that – under other equal conditions (obviously excluding cataclysms) – the onset of accidents depends to a considerable extent on defects. A defect of a facility implies a fault in the elements that may lead to a failure of the facility or lowers the level of its capacity for work. In many cases the presence even of small defects causes the breakdown of elements long before the onset of the fixed inter-repair time or the expiry of the resources of facilities, with grave consequences. In many elements, defects accumulate in the course of time, among which some are major defects of three-dimensional form: loose states, pores, various inclusions, cracks.

In view of the extreme importance of the description of defective states, they have been the subject of long and fairly in-depth investigation. Attention was given to the study of this problem by many outstanding mathematicians, including the great mathematician of the end of the 20th century A.I. Kolmogorov [1]. The enormous material losses caused, including human casualties, call for urgent and radical measures towards averting their occurrence. To this end, in the first place one must have an objective foresight of the impending danger. All grounds exist for the assumption that the world community – and not only individual advanced countries – will in the near future develop measures towards lowering the presence of defects in structures, thereby substantially reducing the hazard of accidents and preventing grave accidents in the technosphere. Observations of accidents demonstrate that they largely depend on defects in the facilities.

The thorough study of the characteristics of defective states, set forth in the foregoing, suggests its application in predicting the accident-proneness of various facilities. The first step in this direction is to begin with the use of the findings of the study of the presence of defects in facilities and of the factors causing them and directly related to accidents. Obviously, under other equal conditions, the more the number of defects in any randomly taken facility, the oftener it is subject to accidents. Observations and analysis show that the correlation between the number of defective items in a sample X and the number of defective items in the remaining – uncontrolled – batch X-x is positive – equals zero or is negative, if the dispersion number of the controlled item in batches is, respectively, higher, equals, or is lesser than the binomial dispersion.

$$\sigma_x^2 = NPq\,. \qquad (1)$$

In this connection, in predicting the process of the onset of accidents – by itself an exceptionally complex task – the use of rigorously derived mathematical expressions for the description of the presence of defects seems appropriate.

The formula given in [1], describing the covariation between the number of defective items in a sample and the number of defective items in the remaining batch serves as such an expression. Thus, we shall have: a mathematical expression for covariation between the number of accident-prone facilities in a sample and the number of accident-prone facilities in the remaining part of the group, obtained in [1].

$$E\{x(X-x)\} = \frac{n(N-n)}{N(N-1)}\left[\sigma_x^2 - Npq\right]. \qquad (2)$$

Clearly enough, under positive covariation, the level of accident-proneness of the uncontrolled group $P_{N-n} = (X-x)/(N-n)$ will be close to the level of the accident- proneness of the sample $P_\pi = x/n$; under covariation values close to $\sigma_x\,\sigma_{X-x}$ - product of standard diviations of the numbers of accident-prone facilities in the sample and the uncontrolled part of a batch, with a probability close to unity, it may be asserted that, $P_n = P_{N-n} = P_N$. This fact constitutes the main principle of the classical theory of statistical acceptance inspection. As zero covariation P_n and P_{N-n} are independent random values and, therefore, random inspection is useless, for according to its results one cannot judge about the level of the presence of defects of the remaining part of the batch. Finally, under negative covariation two cases are possible, $P_n < P_N < P_{N-n}$ and $P_n < P_N < P_{N-n}$. These have the following probability meaning: the items in a batch are arranged with a definite regularity in such a way that the level of the number of defects in the remaining part of a batch is correspondingly less or more than that of the batch. In the limiting case, when $E\{x(X-x)\} = -\sigma_x\sigma_{X-x}$, any given level of the number of defects $P_\pi \geq P_N$ has its absolutely definite the number of defects $P_n \geq P_N$. The following designations are adopted in the formula: N, n - volumes of the batch and sample: X, x - numbers of defective items in a batch and sample; p - level of the number of defects, $q = 1 - p$; $g_N(X)$ - a priori distribution of the number of defective items in batches submitted for control; $N_n(x)$ - unconditional distribution of the number of defective items in the samples; $f_n(x/X)$ - conditional distribution of defective items in a batch (hypergeometric distribution); $\Phi_{N-n}(X-x/x)$ - a priori distribution of the number of defective items in the un controlled part of the batch at a given number of defective items in the sample; $j(x, X)$ - joint distribution of the number of defective items in the batches submitted to testing; $E\{x(X-x)\}$ - covariation between the numbers of defective items in the sample and control part.

Realization of the inexhaustible potentialities of branches of the theory of random processes and of reliability theory to forecast accidents and accident-prone situations on the basis of processing the observation data, in time, is one of the principal ideas in the foregoing works.

Detection of accidents through the presence of defects in the proposed method is effected by replacement of the presence of defects in formula (2) with accident rate. However, in the majority of practical cases the characteristics of the presence of defects are unknown, which inevitably causes difficulties in prediction. However, it is much simpler to determine these characteristics than to determine the accident rate directly.

Along with the above-said, in analyzing accidents the use of the so-called methods of leading indications of failures appear to be promising, which will undoubtedly facilitate timely implementation of the following measures connected with combating accidents: identification of weak points of constructions at exploitation, under group leading; assessment of the parameters of distribution of mean-time-to failure, allocating a specified life for concrete elements, typical of a group of facilities. In the case of group leading the group of facilities is identified that is characterized by a lead in operating time. At individual leading, several leader-indicators – models of the element under consideration – fall to each facility, having – under equivalent conditions of load – the same function of distribution of mean-time to failure and operating-time lead.

Conclusion

One of the principal tasks of the present paper is an attempt made once again to draw attention to the thorough and objective study of the causes of accidents, allowing an insight into the regularities and conditions of operation of constructional systems and their elements; to draw the attention of scientists to the solution of inadequately researched problems in the field of design and construction, as well as to flaws in design and construction, causing destruction, the removal of which should prevent accidents and raise the reliability of structures.

ჰიდროლოგია

ჰიდროსაინჟინრო ნაგებობების ავარიების პროგნოზირება და თავიდან აცილება

ც. მირცხულავა*

* *აკადემიის წევრი, წყალთა მეურნეობის ინსტიტუტი, თბილისი*

ნაშრომში განხილულია ჰიდრონაგებობების ავარიების კვლევის აქტუალურობა. მოცემულია ზოგიერთი ავარიის მოკლე აღწერა და პროგნოზირებისადმი მიდგომა ა.ნ. კოლმოგოროვის ნაშრომების გამოყენებით, რომელიც დეფექტიანობას ეძღვნება.

REFERENCES

1. *K. Yu, A.N. Belyaev* (1960), Kolmogorov, Economic plans of acceptance inspection. IV All-Union Mathematical Congress. Leningrad. (Russian).
2. *Ts.E. Mirtskhoulava* (1975), Erozionnye i selevye processy i bor'ba s nimi, vip. 4, M., VNIIGiM, 1975, 48-62 (Russian).
3. *Ts.E. Mirtskhoulava* Avarii: uroki, prognoz, mery po bezopasnomu funktsionirovaniju objektov gidrotekhniki. Moscow. (Russian).
4. *Ts.E. Mirtskhoulava* (1993), Ecological disturbances (prediction of the risk of disturbances, measures for reducing the hazard). Tbilisi, 438 p. (Russian).
5. *Ts.E. Mirtskhoulava* (1998), Gidrotekhnicheskoe stroitel'stvo, 5:16-19 (Russian).
6. *Ts.E. Mirtskhoulava* (2003), Opasnosti i riski nekotorykh vodnykh i drugikh sistem. Vidy, analiz, otsenka. V 2-kh knigakh 503 p. (Russian).
7. *Ts.E. Mirtskhoulava* (1987), Reliability of Hydro-reclamation Installations. – Rotterdam: A.A. Balkema, 297 p.
8. *Ts.E. Mirtskhoulava* (2000), Mud floods and their prediction in the Caucasus and Central Asia. /*Floods*. Vol. II. Edited by D.J. Parker. – London, New York, p. 216-222.
9. Guidelines for Floods Loss Prevention and Management in Developing Countries. Aarhus Group Experts of Flood Prevention (S.K.Banerjee, I.Bogardi, A.Fahmy, H.Kikkawa, P.Maistr, Ts.E.Mirtskhoulava, S.Nesadurai, G.R.Phippen, L.S.Rahm, T.Tahara, J.W.Vam der Made), Natural Resources. /Water Series N5, United Nations, New York, 1976.
10. Construction Industry Research and Information Association (CIRIA) (1977) Rationalization of Safety and Serviceability Factors in Structural Codes, Report 63. CIRIA, London.

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Hydrology

Catastrophic Phenomena in the Caucasian Nival Belt

Nodar Begalishvili*, Vasil Tsomaia*

* *Institute of Hydrometeorology, Tbilisi*

(Presented by Academy Member Ts. Mirtskhoulava)

ABSTRACT. The dynamics of glacial catastrophes in the Caucasian nival zone for 1776-2002, their stimulating factors and results are studied. The hazard assessment of glacial catastrophes and elaboration of protective measures are highlighted. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *glacier, ablation, mudflow, glacial pulsation.*

The existence of glacial and nival (eternal snow) zone is one of the distinguishing features of the Caucasus natural landscape. It mainly covers the adjacent territory of the Caucasus range watershed, which is located at the elevation of 3000-5000 m a.s.l. This zone is characterized by severe natural conditions with cold long winter time and short cool spring. Therefore one of the major peculiarities of its landscape is the existence of seasonal or everlasting snow cover and glaciers. Nival zone is the main nutrition source of freshwater resources and rivers, which is a great natural wealth. At the same time this zone is characterized by frequency of natural disasters, for instance glacier catastrophes. According to hazard rating glacier catastrophes belong to the strongest category: their characteristic feature is big velocities - 50-120 m/sec, 20-100 mln m^3 volume of glacial drift, big destructive force of about 200-300 t/m^2. Glacier avalanche destroys and buries everything it meets during movement, blocking gorges with 100-120 m thickness snow and glacial drift. In spite of its fierceness, it is possible to avoid glacier catastrophe. To this end it is necessary to study its nature and make an assessment of its possible manifestation. Glacier catastrophes occur during glaciers pulsation, i.e. during growth (movement forward) and decline (retreat) periods. Therefore catastrophe-causing factors differ from each other. In the growth period catastrophe-causing factor is a huge mass of glacier which is mainly estimated in the form of ice. It differs from other glaciers. For example, the catastrophe starts on the Kolka glacier when its thickness reaches 150 m, and on the Devdoraki glacier exceeds 75 m, inclination of glacier surface (a) in the cited case totals 10 and 16°, respectively. Thus the larger the glacier surface inclination, the less its thickness is and it represents glacier's critical parameter (h_{cr}) which is calculated using the old [1] and renewed formula:

$$h_{cr}=172\,\alpha^{2}[(0.9+\rho)^{6}+(0.99+\rho^{2})^{6}],$$
$$h_{cr}=172\,\alpha^{-2}\cdot 10^{0.1+1.93\rho}, \qquad (1)$$
$$h_{cr}=172\,\alpha^{-2}\cdot e^{0.23+4.44\rho}.$$

During the period of glacier retreat a decrease of the values of its indices takes place. Nevertheless, catastrophe might happen. It is caused by the action of other stimulating factors [2]. To them belong earthquakes (Devdoraki glacier, 1832), rock-avalanche fall on the glacier surface (Abano glacier, 1909), accumulation of a great amount of water from melted snow, rainfall and glacier in glacier fractures, ice pockets, ice shafts, interglacial hollows, subglacial cavities (Kolka glacier, 1969 and 2002). As a result of filling the mentioned forms with water the stability of glacier mass breaks, resulting in catastrophic events, origination of floods and glacial debris flows. During such hazards maximal water discharge (Q_m) is calculated by the formulas:

At the breakthrough area $Q_m = 1.9\,Bh^{3/2}$. (2)

On transition site $Q_{mi} = \dfrac{L}{L+L_i}\cdot Q_m$, (3)

where B denotes the width (m) of the newly cut glacier bed at the breakthrough area; h is the depth of break-

through (m); L is the length of the expected dam (km); L_i is the distance from the breakthrough area to any section of transition site.

The boundary of the destructive force spreading (l km) of glacial air wave and the distance passed by glacial avalanche (L_{tr}) is calculated by formula [1],

$$L_{tr} = \frac{W}{b \cdot h_{cr}}, \quad (4)$$

$$l = 1.6(0.16\sqrt{W} + \sqrt[3]{W}), \quad (5)$$

where W *is* the volume of ice avalanche (m^3); b is the average width of the bottom (m).

The formulas (1)-(5) were checked and satisfactory results have been obtained [1]. This is also proved by completely new data about the catastrophe on Kolka glacier which took place on 20 September, 2002 [2]. It was caused by colossal accumulation of rain water and water formed as a result of melted snow and glacier. Accumulated water surged up the glacier and turned it into catastrophic mudflow at a distance of 12 km. To check this we consider the results of estimated characteristics of glacial catastrophe that occurred on the Kolka glacier based on the initial data and those obtained from literature sources.

The length and width of glacier in the glacier circus is L=3.5 km, B=750 m and b=200 m, respectively, glacier area is f=2.6 km^2, inclination of glacier surface is α=8° (7-9° [2]), volume of the Kolka glacier W_{gl}=1.1·10^8 m^3, water maximal discharge at the breakthrough site, formula (2) Q_{sm} = 500000 m^3/sec, density of glacial water ρ = 0.6 g/cm^3, glacier's critical thickness, formula (1) h_{cr} = 50 m, the traveled distance of glacial water mixture, formula (4) L_{tr}=11 km (12 km [1]) and the boundary of destructive force of glacial wave spreading (l) formula (5) l=3.2 km (3.0-3.5 km [2]).

According to the calculated results the distance traversed by avalanche water-diluted mass flow and that of spreading air wave destructive force coincide with actual (given in brackets) values.

Thus the obtained results represent a scientific basis for technical and economic grounding of prevention of avalanche hazards. The following recommendations are given to deal with avalanches: an increase of glacial ablation or melting, conduit of channels in the obstruction mass to let impounded water pass, to dig a tunnel at the foot of the opposite slope, etc.

ჰიდროლოგია

კატასტროფული მოვლენები კავკასიონის ნივალურ ზონაში

ნ. ბეგალიშვილი*, ვ. ცომაია*

* *ჰიდრომეტეოროლოგიის ინსტიტუტი, თბილისი*

(წარმოდგენილია აკადემიკოს ც. მირცხულავას მიერ)

ნაშრომში დახასიათებულია 1776-2002 წლებში კავკასიონის ნივალურ ზონაში მყინვარული კატასტროფის დინამიკა, მათი მასტიმულირებელი ფაქტორები და შედეგები; შემოთავაზებულია ემპირიული ფორმულები, რომელთა დახმარებით განისაზღვრება მყინვარის კრიტიკული სისქე, მყინვარის ჩამონაქცევით წარმოქმნილი ჰაერის ტალღის დარტყმის ძალა და გავრცელების მანძილი, ხეობის ჩახერგვით წარმოშობილი დაგუბებული ტბის გარღვევასთან დაკავშირებული კატასტროფული წყალმოვარდნების წყლის მაქსიმალური ხარჯის გამოთვლა გარღვევის ადგილას და მდინარის დინების ტრანზიტულ უბნებზე. მოყვანილია შემოწმების შედეგები და რეკომენდაციები საშიშროების თავიდან ასაცილებლად.

REFERENCES

1. *V. Tsomaia* (1983), Opasnye gidrometeorologicheskie javlenija na Kavkaze. L., 158-184 (Russian).
2. Express information (2002), Catastrophe on Kolka Glacier. MGI, vyp.93, 221-228 (Russian).

Received January, 2007

Bull. Georg. Natl. Acad. Sci. vol. 175, no. 3, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Petrology

On Genetic Identity of Rkvia and Beretisa Granitoid Intrusives (Dzirula massif)

Avtandil Okrostsvaridze*, Rusudan Chagelishvili*

* *A. Janelidze Geological Institute, Tbilisi*

(Presented by Academy Member D. Shengelia)

ABSTRACT. In the canyon of the river Dumala, near the village Beretisa granitoids are outcropped by a narrow zone. These rocks were first outlined and studied by the authors. On the basis of geological, mineralogical, petrochemical and geochemical studies we can assume that Beretisa's granitoids and Upper Hercynian generation Rkvia granitoid intrusive located in its western part within 12 km, are absolutely identical and distinguished in Rkvia-Beretisa intrusive complex. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *granitoid, intrusive, outcrops.*

Dzirula massif is one of the best studied pre-Alpine units in the Caucasus from the petrological point of view, which has been intensively explored during the last decade [1-4]. However, some controversial issues about its geological structure still exist, including the problem of the genetic identity of Rkvia granitoids and granitoids outcropped in the river Dumala and in its tributaries.

As there is much information on the Rkvia intrusive [4-6], we propose only a short review. The intrusive crosses Upper Proterozoic gneissic quartz diorites and at the same time is overlaid by Upper Paleozoic quartz porphyries, Liassic sediments and Upper Cretaceous and Lower Neogene limestones. The crystallization age of Rkvia intrusive is 352±mln. years (Rb-Sr method), as for the time of its inversion, it is 303 ±2 mln. years (^{40}Ar-^{39}Ar method) [4]. The intrusive is obviously of a crust origin, generated as a result of anataxis of Sialic rocks. The geometry of the intrusive shows that injection of magmatic melt was caused by meridional compression subsidence.

After the 12th km of the eastern extension of the Rkvia intrusive, near the village of Beretisa Upper Cretaceous limestones are underlain by granitoids. They are outcropped within canyons of the river Dumala and its left tributary Akhashmula, being represented by narrow zones (strips), with a total area of up to 4 km^2. This

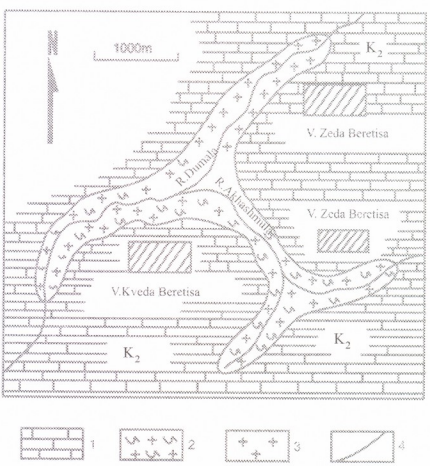


Fig. 1. Sketch of the Beretisa intrusive outcrop. 1. Upper Cretaceous limestone, 2. Late Hercynian microcline granites and pegmatites, 3. Early Hercynian microcline porphyry granites, 4. Geological borders.

Table 1.

Chemical analyses of granitoids of the Beretisa intrusive (mas.%).

Samp.№	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	FeO	MnO	CaO	MgO	K_2O	Na_2O	P_2O_5
1-05	72.36	0.19	14.42	2.02	0.41	0.05	0.98	0.32	4.36	3.32	0.02
2-05	73.52	0.15	14.45	2.08	0.50	0.07	0.93	0.37	3.58	3.39	0.01
3-05	75.13	0.18	12.18	1.85	0.30	0.03	1.23	0.46	3.63	3.05	0.02
5-05	71.78	0.19	14.69	2.15	0.57	0.04	0.39	0.37	4.72	3.05	0.06
6-05	73.16	0.15	13.05	1.89	0.40	0.03	0.98	0.28	4.38	3.94	0.02
7-05	74.51	0.01	13.92	0.45	0.08	0.01	0.03	0.10	6.05	3.01	0.06
8-05	76.33	0.03	12.90	0.64	0.20	0.01	0.05	0.14	5.02	3.29	0.04
9-05	77.03	0.18	12.18	1.46	0.40	0.03	0.09	0.38	2.23	3.94	0.04
10-05	72.46	0.24	14.07	2.50	0.22	0.05	1.15	0.20	3.65	3.94	0.03

Samples: 1-05, 2-05, 3-05, 5-05, 6-05, 10-05 –porphyritic granites; samples 7-05, 8-05, 9-05 –pegmatites.

outcrop was for the first time outlined (Fig. 1) and studied by the authors [7].

The investigations carried out showed that by its textural-structural characteristics, petrologic and mineralogical composition, petrochemically, geochemically and according to distribution of the rock types, the outcrop of Beretisa granitoids is absolutely identical with the Rkvia intrusive. It is obviously a fragment of the granitoid intrusive, hence we call it Beretisa intrusive.

The unaltered granitoids of the Rkvia and Beretisa intrusives are of milky-whitish color. As for the altered varieties, K-Na feldspars acquire a pinkish hue, owing to which the whole rock has the same coloring.

The central part of the west edge of the intrusive (cross-section of the river Buja) is built up with porphyritic granites. On both peripheries they are replaced by granitic pegmatites and aplites. In the cross-section of the river Dumala, where only a fragment of the intrusive is outcropped, porphyritic granitoids are substituted by granitic pegmatites and aplites, as in the Buja cross-section.

Both outcrops are alike according to their structural-textural peculiarities. In particular, most of the granitoids are of porphyritic structural generation; where the main mass is represented by medium-grained rocks of massive texture. In both cross-sections porphyric minerals of the main mass are idiomorphic crystals of K-Na feldspars (3-5 cm). The rock building minerals of the main mass are silicic (acid) plagioclase, K-Na feldspar, quartz and biotite. Secondary minerals are muscovite, sericite, chlorite, epidote

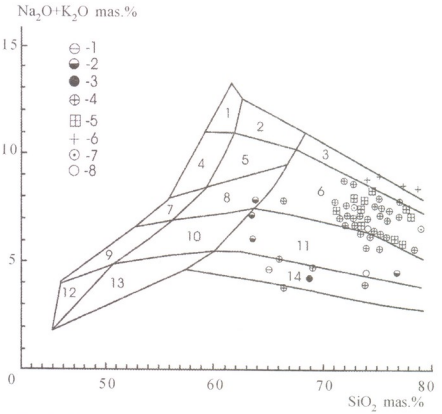


Fig. 2. Rocks of the Rkvia-Beretisa intrusive complex on (Na_2O+K_2O)-SiO_2 discrimination diagram of granitoids (Middlemost, 1985).

Fields: 1. alkali systems; 2. alkali quartz-syenites; 3. alkali granites; 4. syenites; 5. quartz-syenites; 6. granites; 7. monzonites; 8. quartz-monzonites; 9. monzodiorites; 10. quartz-monzodiorites; 11. grano-diorites; 12. gabbros; 13. quartz-diorites.

Symbols: 1. granite-gneisses; 2. migmatites; 3. tonalites; 4. porphyritic granites of Rkvia intrusive; 5. porphyritic granites of Beretisa intrusive; 6. even-graind granites of Rkvia intrusive; 7. aplites; 8. pegmatites.

and caolinite. Accessory minerals are zircon, apatite, sphene, monacite and ilmenite. According to chemical composition, petrochemical parameters and peculiarities of rare elements distribution, both intrusives are almost the same. In the (Na_2O+K_2O)-SiO_2 discrimination diagram the figured points of the rocks of Rkvia and Beretisa intrusive are distributed in the same field of normal granites.

Rocks investigated according to the A/CNK parameter belong to peraluminium S type. In the D_1-D_2 discrimination diagram figure points of the Rkvia and Beretisa granitoids also are focused in the same S type field. The same happens in the R1-R2 geodynamic diagram, where almost all points are located in the field of syncollision formations.

According to the peculiarities of distribution of rare elements, both intrusives are identical. For example, they have higher concentration of Ni, Co, Cr, Hf, Ta, Sn, Th and U in comparison with the clarke concentration. As to Cu, Mo, Pb, Rb, Li, Sr, Y, Nd, Ga and V contents, they are within normal [7].

Thus the Rkvia and Beretisa intrusives are geologically, petrologically, mineralogically, petrochemically and geochemically absolutely identical. From the viewpoint of their distribution in space, they may be considered as different manifestations of the whole protolithic and tectono-magmatic activity. This gives ground to identify the Rkvia-Beretisa intrusive complex as a separate unit.

პეტროლოგია

რკვიისა და ბერეთისას გრანიტოიდული ინტრუზივების გენეტური იდენტურობის შესახებ (ძირულის კრისტალური მასივი)

ა. ოქროსცვარიძე*, რ. ჩაგელიშვილი*

* *ა. ჯანელიძის გეოლოგიის ინსტიტუტი, თბილისი*

(წარმოდგენილია აკადემიკოს დ. შენგელიას მიერ)

მდ. დუმალას ხეობაში, სოფელ ბერეთისასთან ავტორების მიერ კონტურით შემოსაზღვრულია და შესწავლილია პორფირისებრი გრანიტოიდები. მათს გეოლოგიურ, პეტროგრაფიულ, პეტროქიმიურ და გეოქიმიურ კრიტერიუმზე დაყრდნობით დაშვებულია, რომ ბერეთისას გრანიტოიდები და მის დასავლეთით 12 კმ-ში მდებარე, ზედა ჰერცინული გენერაციის რკვიის გრანიტოიდული ინტრუზივი გენეტიკურად და ასაკობრივად იდენტურ წარმონაქმნებს წარმოადგენენ და სწორედ ზედა ჰერცინული გენერაციის მიერაა გამოყოფილი რკვია-ბერეთისას ინტრუზიული კომპლექსი.

REFERENCES

1. *I. Gamkrelidze, D. Shengelia* (2005), Precambrian-Paleozoic regional metamorphism, granitoid magmatism and geodynamics of the Caucasus. Moscow: Scientific World, 460 p.
2. *D. Shengelia, A. Okrostsvaridze* (1998), Report RAS, 359, 6, 801-803.
3. *K. Chikhelidze* (1998), Cand. thesis., Tbilisi, 25 p.
4. *A. Okrostsvaridze, D. Clarke, P. Reinolds* (2002), Coll. Papers Geol. Inst. Georg. Acad. Sci., new ser., v.117, p. 173-186.
5. *P. Topuria* (1938), Bull. GIN GSSR, 4, p. 361-475.
6. *R. Manvelidze* (1983), Coll. Papers Geol. Inst. Georg. Acad. Sci., new ser., v.81, 124 p.
7. *R. Chagelishvili* (2006), Cand. thesis, Tbilisi, 25 p.

Received January, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Physical Geography

Natural Anthropogenic Mining Complexes and the Problems of their Optimization

Zurab Seperteladze*, Eter Davitaia*, Tariel Kikvadze*

* *I. Javakhishvili Tbilisi State University*

(Presented by Academy Member Z. Tatashidze)

ABSTRACT. The present work concerns optimization issues of natural anthropogenic industrial mining complexes. Particularly, statistic interrelation between physical and geographic factors and reclamation time period is obtained by means of linear regression analysis. In combination with appropriate empiric data base, it allows to make scientifically well-reasoned statement about optimization of environmental conditions. *© 2007 Bull. Georg. Natl. Acad. Sci.*

***Key words**: landscape, landscape and ecologic situation, natural territorial complex (NTC).*

Realization of a complex approach to nature protection and rational use of natural resources is one of the basics of stable development of a country. Complex studies are of exceptional significance in mining regions, where proceeding with industrial processes without consideration of the sensitive landscape and ecologic situation is impossible.

It needs to be noted that natural landscapes formed as a result of technogenesis substantially differ from the earlier complexes by their morphologic parameters, structure and biocenosis as well as by substance circulation, biologic production and specifics of industrial utilization. In an ore extraction zone (within radiuses starting from several tens of meters to several kilometers) disturbance of the natural balance formed over many centuries takes place and a significant part of the agricultural lands gets out of use. There are several hundreds of hectares damaged in this way in Georgia, forming a considerable part of the entire territory of the country (1.5% approx) [1].

Slowing down of autoregulation of technogenic geologic complexes and activation of rehabilitation processes are possible only by means of different measures of optimization. Hence, the need of treatment of general theoretical issues for functioning of natural territorial complexes (NTC) and physical and geographical regional structures of ore, an extraction zone as well as shedding light on the nature of their space-time dynamics. This requires: a) development of a new approach to the theoretical basics of interaction between nature and society; b) prediction of the ability of autorehabilitation in harsh anthropogenic (technogenic) conditions, c) balanced regulation of utilization of natural resources and reproduction processes in mining zones, d) typification of technogenic complexes, elaboration of zoning principles and cartographic modeling [2].

Of the above mentioned problems, in this paper we shall speak about one of the main problems of optimization of natural anthropogenic mining complexes, such as reclamation.

Reclamation by itself takes place during technical impact on the environment which increases the asymmetry between the landscape components. In other words, it causes relatively fast progress of succession rehabilitation series. Therefore, landscape complex formed by reclamation is impossible to be the same as its pre-exploitation one as tectonic succession is characterized by increased asymmetry between the landscape components relatively to the natural one. Consequently, realization of this kind of landscape optimization has to be carried out sensibly and purposefully [3].

Among many other factors, physical and geographical ones are very important for successful realization of

landscape optimization processes. The time period, being one of the main characteristics of reclamation related to restoration of distorted landscape, also depends on the abovementioned (explanatory) factors. In order to determine this value, studies were carried out on 8 plateaus in Tchiatura district (Rgani, Merevi, Perevisa, Itkhvisi, Bunikauri, Tabagrebi, Mghvimevi and Darkveti) [4].

In order to reveal the relation between different values, methods of mathematical statistics are often used in practice [5]. One such method - regression analysis - is distinguished for the presence of independent (explanatory) and dependent (resultant) variables. Such separation of variables is made by a researcher issuing from the matter of the problem:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots a_nx_n + E,$$

where Y is a value of the dependent (resultant) factor and $x_1, x_2, \dots x_n$ are the independent (explanatory) ones, E is the fluctuation caused by unpredictable (incidental) events. The constant values – $a_0, a_1, a_2 \dots a_n$ are determined by the regression equation.

We have used this method to estimate the reclamation time period of damaged landscapes of Tchiatura ore based on the data obtained from the empiric mining and scientific literature. The following relationship was obtained with rather high statistical validity:

$$Y = 0.0044x_1 + 0.039x_2 + 0.006x_3 + 0.06x_4 - 0.2702x_5 + \\ + 0.5587x_6 - 0.1406x_7 + 0.003x_8 + 0.0028x_9 + 0.0035,$$

where Y is the reclamation time period (resultant factor). The explanatory factors are: x_1 – the power of the roof layer (m), x_2 – the horizontal section of the relief (km/km^2), x_3 – the vertical section of the relief (m/km^2), x_4 – the relief inclination (deg.), x_5 – the soil humus layer thickness (sm), x_6 – indicator of the soil acidity(pH), x_7 – the maximal difference between project and surface indices of the dirt pile (m), x_8 – the dirt pile area (m^2) and x_9 – the dirt pile volume ($\cdot 10^3$m^3).

Elasticity coefficient of i-*th* explanatory factor ($Э_i$) is one of the important statistical characteristics that are used for the analysis of regression relationship:

$$Э = a_i \frac{\overline{x_i}}{\overline{y}},$$

where $\overline{x_i}$ and $\overline{y}$ are average values of i-*th* and output factors. These factors show the amount of percent of change of the dependent factors at 1% change of i-*th* factor when the others stay constant. In particular, for the physical and geographical factors that were selected by us, the following values of elasticity coefficients were obtained:

$$Э_1 = 0.1302, \quad Э_2 = 0.4895, Э_3 = 0.0499, Э_4 = 0.4042,$$
$$Э_5 = 0.0833, Э_6 = 0.5956, Э_7 = 0.0958,$$
$$Э_8 = 0.1624, Э_9 = 0.0617.$$

As one can see, the influence of x_2, x_4 and x_8 factors are important in the region under study.

The accuracy of the obtained statistical dependence is determined by the volume and objectivity of the empiric information. Due to shortage of necessary materials and some other reasons, the maximum needed information was not obtained. So, we have no claim to high accuracy of our results, although prognosis on the mining region optimization with satisfactory correctness is feasible by means of generalization of the obtained dependence and reasonable conclusions about the ecological quality of separate reclamation zones can be made.

ფიზიკური გეოგრაფია

სამთამადნო ბუნებრივ-ანთროპოგენური კომპლექსები და მათი ოპტიმიზაციის პრობლემები

ზ. სეფერთელაძე*, ე. დავითაია*, ტ. კიკვაძე*

* *ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

(წარმოდგენილია აკადემიკოს ზ. ტატაშიძის მიერ)

წარმოდგენილ სტატიაში განხილულია სამთო სამრეწველო ბუნებრივ-ანთროპოგენური კომპლექსების ოპტიმიზაციის საკითხები. კერძოდ, წრფივი რეგრესიული ანალიზის მეთოდის გამოყენებით მიღებულია

რეკულტივაციის დროის ხანგრძლივობისა და ფიზიკურ-გეოგრაფიულ ფაქტორთა სტატისტიკური ურთიერთდამოკიდებულება, რაც სათანადო ემპირიულ მონაცემთა ბანკის შემთხვევაში, შესაძლებლობას იძლევა გამოითქვას მეცნიერულად არგუმენტირებული მოსაზრება რეგიონის ბუნებრივი პირობების ოპტიმიზაციის შესახებ.

REFERENCES

1. *E. Davitaia* (1988), Sakartvelos metsnierebata akademiis Vakhushti Bagrationis geograpiis institutis shromebi. **VII**, Tbilisi, 100-105 (Georgian).
2. *E. Davitaia, Z. Seperteladze* (2002), Materials of International Conference "Mtiani regionebis mdgradi ganvitarebis problema". Tbilisi, 27-28 (Georgian).
3. *D. B. Ukleba* (1983), Antropogennye landshafty Gruzii. Tbilisi, 157p. (Russian).
4. *V.I. Fedotov* (1985), Tekhnogennye landshafty. Teorija, regional'nye struktury, praktika. Voronezh, 273p. (Russian).
5. *N. Draiper, G. Smith* (1986), Prikladnoi regressionnyi analiz. M., 273 (Russian).

Received January, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Physical Geography

Geographo- Hydrological Arguments on Possible Expansion of the Tskaltubo Cave System

Givi Gigineishvili*†, Zurab Tatashidze, Kukuri Tsikarishvili***

* *Vakhushti Bagrationi Institute of Geography, Tbilisi*
** *Academy Member, Vakhushti Bagrationi Institute of Geography, Tbilisi*

ABSTRACT. The Tskaltubo Cave System is one of the most important speleological objects which is designed for effective tourist exploitation. Because of its complex geological setting the entire system and its unknown branches in particular have not been studied yet. On the basis of geographo-hydrological methods their provisional directions have been defined and new vectors of speleological research stated. *© 2007 Bull. Georg. Natl. Acad. Sci.*

***Key words**: karst, theoretical and actual runoff, underground water circulation, topographic and underground basins.*

One of the most significant trends of hydrology is mountain hydrology. It is based on the basic regularity expressed by the vertical zonality of the elements of river runoff in mountainous areas. [2]. The mean height of the river basin index is taken for the most efficient and widely approved hypsometric characteristic. As a rule, there is a close and synonymous link between the mean height of river basins and elements of runoff. The larger is the territory under investigation the more probable is this link, which is of differential nature. Not only Georgia but the whole Caucasus is covered by such dependence curves. In case when azonal or anthropogenic impacts are registered in the area the mentioned regularity is disturbed and abstraction noted. The stronger is the influence of the outer factors upon the runoff the greater is abstraction. Thus we can theoretically define the volume and regime of the runoff of the rivers which are under the impact of azonal or anthropogenic factors, as well as determine the parameters when the rivers are not subject to any outer influence.

To determine one of the strongest azonal factors, i.e. to determine theoretical runoff under the conditions of karst expansion is very essential since we can determine the range of its effect by comparison of the indices of theoretical and actual runoff. On the basis of this procedure it is possible to estimate the effect of karst processes upon the local runoff in the process of the study of the Tskaltubo cave system and its vicinity. Besides, areas of inflow and outflow of underground karst water were identified, i.e. the mechanism and direction of water change /circulation/ were defined. Comprehensive research into underground water change is one of the major goals of karst hydrology. It will enable us to create the picture of not only the area of redistribution of runoff but to establish provisional offshoots of the speleological system, to outline the location of underground passages and their hitherto unknown branches.

Since the territory under study is maximum 800 m. above sea level and can be the main recharge source of karst water, we should take this hypsometric mark for determination of theoretical runoff. Heavily karsted territory to the north-south of Tskhunkuri settlement, on the right bank of the river Semi, attests to this. Its maximum height is 636 m. Let us call this 2.7 km^2 territory a potential inflow hearth of Tskhunkuri. Besides, a well defined possible inflow hearth is located to the right of

the Semi river, north-east of Melouri, with 1.5 km^2 area. Maximum hypsometric mark of the Okunela spot is 758m. We may as well call it a potential inflow hearth. It took its name from the village of Okunela, lately turned into ruins.

Both these territories are situated beyond the topographic, i.e. formal drainage basin limits of the Kumi river. But if we take karst hydrological principles for granted, topographic watershed loses its importance, if underground streams flow from one formal basin into another through karst passages or conduits. This is called underground water change, which can be positive or negative.

The basic principle of karst hydrology is especially important in the study of the Tskaltubo-Kumistavi speleological system, aimed at tracing unknown underground passages and their branches. This means that if the actual runoff of the Kumi river does not correspond to the amount of theoretical runoff of that height zone we should search for some other areas beyond the theoretical drainage system basin supplying the Kumi river with water through underground conduits. If this is confirmed, we should look for some new underground branches of the Tskaltubo cave system in the direction of inflow currents without fail.

It has been ascertained that the topographic or formal drainage system up to the village of Qvilishori on the Kumi river is 14 km^2. The underground river Ghliana receives its alimentation just from the same basin and it is not feasible to separate the surface basins of these two underground rivers on 1:50 000 scale topographic maps.

Now we have to determine the volume of water which is supplied by the 14 km^2 topographic basin and is considered to be the surface drainage system of the Kumi and Ghliana rivers.

If there exists a reliable zonal background of distribution of runoff, then it is possible to estimate those variations which are caused by azonal, anthropogenic and other local factors, including karstic. However, these estimations do not claim to be accurate, as only the zonal runoff of the territory by the river basin was considered and sometimes the data of hydrological stations located on quite vast territories.

Whenever we are interested in the zonal runoff of this or that uninvestigated river or the river beyond the impact of outer factors, whose basin is within the area of the curve, then it is desirable to posses at least some data on the river runoff which will help to make calculations of the runoff more precisely.

At the earliest stage of investigation of the Tskaltubo-Kumistavi cave system we requested the Hydrometeorological Service of Georgia to arrange hydrometeorological observation posts on the Kumi and Ghliana rivers in 1987. The annual cycle of observation was made up for everyday observation data. That year, the mean annual discharge on the Kumi river was 0.64m^3, on the river Ghliana it amounted to 0.16 m^3/sec. The total discharge of these two cave streams was formed within the limits of one common basin with the surface of 14 km^2 topographic area.

To determine the actual area of the drainage system of these underground rivers, the relation between the mean water discharge of 1987 and the mean perennial discharge should be defined.

It is more conclusive to estimate the rate of runoff per liquid atmospheric precipitation since the alimentation of the mentioned subsurface rivers is absolutely dependent on it.

For this purpose we have chosen the perennial data of the Kutaisi Meteorological Station, which is the most reliable and closest to the area under study. The annual rate of total sums of atmospheric precipitation was 1477 mm per observations of 1936-90, whereas the sum of precipitation of the same series in 1987 totalled 1718 mm. It means that the precipitation fall in Kutaisi in 1987 exceeded the perennial rate by 16 %. Consequently we may assume that the mean annual discharge of the Kumistavi and Ghliana rivers in 1987 must have also exceeded their joint perennial rate. Per simple mathematical calculations with Kutaisi atmospheric data taken as a basis the rate of the runoff of the Kumistavi and Ghliana rivers should total 0.688 m^3/sec.

To define the area of the drainage system basin of the given runoff the zonal rate of the runoff was used again, which within the range of 500-1000 m altitude totaled 50 l/sec.km^2 at the Tskhenistskali-Rioni interfluve. Due to appropriate estimation, it was found that the rate of the Kumistavi -Ghliana runoff was accumulated from 13.76 km^2 area. This shows that the area of the territory of 14 km^2 drainage system basin, depicted by us on the map, absolutely coincides with the area estimated according to calculations based on the runoff data. Such coincidence is good on the one hand, and points to the preciseness of estimation of theoretical runoff but on the other hand, doubts still arise since there exists an objective reason that the Kumistavi-Ghliana actual underground drainage system does not coincide with the surface of the formal basin on the map. The Above-mentioned 2.7 km^2 potential inflow hearth of Tskhunkuri and 1.5 km^2 Okunela hearth belong to such territories. Besides, in the southern part

of the territory under study there exist some undefined sites from where the recharge of underground karst water is possible. First of all to such sites belongs the Qvilishori potential inflow zone situated in the vicinity of Qvilishori settlement at the north edge of the town of Tskaltubo.

One of the spots of underground water discharge may prove to be the outfall of the Tskaltubo river, which deserves more attention than it was paid until now.

Let us go back to the issue of atmospheric precipitation and say that besides the observations and estimations of the runoff carried out in 1987, special work on precipitation gauging was accomplished in the area above the underground Tskaltubo-Kumistavi system and its vicinity. As a result of observations very high indices were registered at the three observation posts organized by the Hydrometeorological Service of Georgia at our request. For instance the precipitation gauge at Melouri at 450 m above sea level registered an annual precipitation sum of 2331 mm, the second precipitation gauge at 250 m situated by the Kumistavi village, near our research station and at Opitcho funnel of discharge registered 2184 mm. At the exit of Kumistavi cave at 127 m altitude 2155 mm precipitation was registered, whereas the Tskaltubo meteorological station recorded 2060 mm at 150 m. In the city of Kutaisi the amount of precipitation reached 1718 mm. All this shows that in the area of Kumistavi cave and its vicinity in the low and piedmont zones of the Rioni-Tskhenistskali interfluve significant gradient changes were observed in the sums of atmospheric precipitation within the range of 350-400 m altitude.

In order to carry out a through investigation of the Tskaltubo-Kumistavi cave system and its profound karstic and hydrological study it is necessary to restore the special permanent hydrometeorological observation network, without which it will not be possible to widen the subsurface research of the cave and ensure its safe and reliable operation for tourist recreational purposes in future.

The Tskaltubo river, whose outfall is in the very territory of the town of Tskaltubo is doubless recharged by karst water. Permanent observations of its runoff were carried out from 1936 to 1940.

Three full annual cycles of observations and gauging were accomplished (1936,1938, 1940). The mean annual discharge totalled 1.72m^3/sec. The mean monthly discharge was also included in these calculations., which were obtained during the years of unsystematic observations (1937, 1940).

The area of the drainage system basin (i.e. topographic basin) up to the water gauging station is 25 km, as registered in the official reference book.

As observations of the Tskaltubo river runoff were very rare, it became necessary to evaluate it on the basis of atmospheric precipitation. Again the most reliable data turned out to be of the Kutaisi Meteorological station which were used in calculating the mean perennial discharge of the Tskaltubo river 1.49 m^3/sec. near the town of Tskaltubo.

The runoff module calculated according to this discharge was 59.6 l km^2/sec. which considerably exceeded the zonal index characteristic of the part of the Rioni river basin where the Tskaltubo flows. This zonal index is 35 l/sec. km^2. The area of the topographical basin of this river cannot provide the amount of the runoff which is available under the existing hydrometeorological conditions. It became necessary to define topographical boundaries of the actual drainage system where the Tskaltubo receives alimentation. The area from where this river could get its alimentation is 49.1 km^2 . All kinds of negative forms of relief which might be considered as surface water consuming hearths have been estimated.

On the basis of methodological geographic-hydrological study it has been found that the height of the layer of the zonal runoff is 1100mm in the basin of the river Tskaltubo and nearby, whereas per the data of actual observations it is 960 mm. We took zonal indices of the runoff to calculate the area of the basin which totaled 42.6 km^2. It can be viewed as a great coincidence (difference is 6.5 km^2) and means that we may proceed with the search for underground channels and streams flowing through them within the limits of the topographic drainage system basin defined by us, where the Tskaltubo gets its alimentation.

It should be noted that at this stage of research the results of study should not be viewed as absolutely authentic, but we have outlined the direction of our further investigations and identified well-founded boundaries of water tracing required for carrying out experiments.

წყალტუბოს მღვიმური სისტემის შესაძლო გავრცელების გეოგრაფიულ-ჰიდროლოგიური არგუმენტები

გ. გიგინეიშვილი*†, ზ. ტატაშიძე**, კ. წიქარიშვილი*

* კახუშტი ბაგრატიონის გეოგრაფიის ინსტიტუტი, თბილისი
** აკადემიის წევრი, კახუშტი ბაგრატიონის გეოგრაფიის ინსტიტუტი, თბილისი

წყალტუბოს მღვიმური სისტემა უმნიშვნელოვანესი სპელეოლოგიური ობიექტია და იგეგმება მისი გამოყენება მასობრივი ტურიზმისათვის. დიდი სირთულის გამო, იგი ბოლომდე არ არის გამოკვლეული. დასადგენია მისი დღემდე უცნობი განშტოებები. კვლევის გეოგრაფიულ-ჰიდროლოგიური მეთოდიკის საფუძველზე განისაზღვრა მღვიმური სისტემის შესაძლო გავრცელების მიმართულებები და სპელეოლოგიური ძიების ახალი ვექტორები.

REFERENCES

1. *L. Vladimirov, D. Shakarishvili, T. Gabrichidze* (1974). Vodnyi balans Gruzii. Tbilisi (in Russian).
2. *L. Vladimirov, G. Gigineishvili, A. Javakhishvili, N. Zakarashvili* (1991), Vodnyi balans Kavkaza i ego geograficheskie zakonomernosti. Tbilisi, 144 (in Russian).

Received June, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Physical Geography

The Anthropogenic Transformation of Natural Landscapes of Sub-Mediterranean Semi-Humid Foothills

Tamar Mamukashvili*

* *Iv. Javakhishvili Tbilisi State University*

(Presented by Academy Member Z. Tatashidze)

ABSTRACT. In sub-Mediterranean semi-humid foothills the stability of landscapes has been established – the degree of anthropogenic influence on them, the sections of ecological tensions and the territory distinguished for high landscape diversity have been determined. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *sustainable, geoecology, re-anthropogenization.*

Determination of sustainable landscapes is one of the main tasks of geoecological investigation. Sustainability of landscapes is the capacity of resistance to natural and anthropogenic factors – **"Immunity"** [3].

The landscapes of sub-Mediterranean semi-humid foothills of Georgia, enclosed within the natural habitat of intense anthropogenic influence (area 2587 km^2), vividly presenting a many-sided picture: 632 km^2 (25%) area – on the average occupies a modified (20-50%) landscape; A territory with the area of 448 km^2 (18%) is represented by rather transformed (50-80%) landscapes; a territory of 910 km^2 (35%) is covered with rather modified (80-95%) landscapes; and the territory of the area of 548km^2 (28%) – are practically completely transformed (95-100%) landscapes.

On the basis of comparison of available literary and cartographic sources (a landscape map of the Caucasus, 1:1000 000, 1979) and our investigations in the sub-Mediterranean semi-humid foothills four landscape categories are represented [1]. The 15^{th} erosion-denudation relief of hilly foothills with the landslide and mudflow processes, shibliak, shrubs-hornbeam-oak-wood derivatives, here and there with arid sparse wood, with lammergeier field and partial phrygana (500 – 800m).

The 16^{th} hilly erosion-denudation relief of hilly foothills with landslide and mudflow processes, shrub-hornbeam-oak-wood derivatives, shibliak, at some places with arid sparse wood, with phrygana and beard grass steppes, here and there with badlands (700 – 1000m).

The 17^{th} denudation-accumulative relief of hilly foothills with landslide and mudflow processes, with complexes of hammer-like field, shibliak, rarely with phrygana and wood derivatives (700 –1000m).

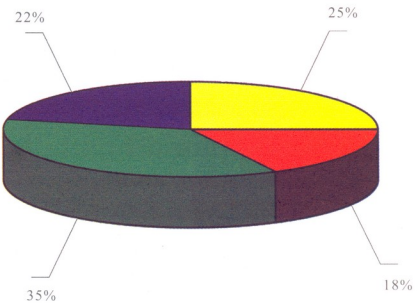
The 18^{th} hilly erosion-accumulative relief of hilly foothills with landslide and mudflow processes, shrub-hornbeam-oak-wood derivatives and shibliak (300 – 800m).

On the basis of researches carried out we can conclude that rather transformed (50-80%) landscapes are least represented, while strongly modified (80-95%) landscapes are represented widely.

According to the genuses the anthropogenesis acquires the following image:

In the 15^{th} category the landscapes modified by (20-50%) occupy 41% (180 km^2) of category, and the landscapes modified by (80-95%) occupy 59% (258 km^2).

In the 16^{th} category we encounter a whole spectrum of transformation. Here, the change by (20-50%)



occupies 20% (250 km^2); The least represented is (50-80%) change 18%-(224km^2). The (80-95%) transformation is spread at 21% (268km^2) and the (95-100%) change is represented in 41% of the category (536km^2), being a dominant in the category. And the least is represented the 50-80% change, 224 km^2 (18% of the category).

In the 17th category the priority development has a 20-50% transformation. It covers 53% (205km^2) of the category. By 50-80% is changed 58km^2 of the category (15%), and 124 km^2 (which constitutes 32% of the category) experiences a 80-95% transformation.

In the 18th category two degrees of anthropogenesis are recorded. 61% (260km^2) of the category comprises 50-80% modification, and the remaining 39% (166km^2) occupies a 20-50% modified landscape.

According to the above indicated data 25% (635km^2) of sub-Mediterranean semi-humid foothills landscapes are modified at an average rate (20-50%); 18 % (448 km^2) is transformed seriously considarably (50-80%); 35% (910 km^2) are strongly modified (90-100%).

Environment protection and nature use are nearly identical to maintaining and protection of self-restoration of landscapes. [2]

Among the sub-Mediterranean semi-humid foothills we shall distinguish ecologically tensile sections to which the Kaspi and Kazreti districts have been assigned regions belong. Kaspi is included in the 16th category "The hilly erosion-denudation narrow mountainside, with active gravitation processes of steep slope, here and there with badlands, constructed of molas rocks, with arid rare wood on dark brown and washed soils" and in the 15th category "the hilly erosion-denudation foothills, constructed with volcanic deposit rocks, shrub-hornbeam-oak-wood derivatives, meadow steppes and shibliak (300–800m) on leached brown and meadow brown soils".

The territories with high landscape diversity are distinguished in the region. This is Martkopi stationary and adjacent territory which is comprised in the 17th category of research region.

In Martkopi stationary and its neighborhood we have distinguished 2 landscape sites. In each landscape sites there are distinguished two valleys. And in valleys there are 20 tracks and 42 facies, giving ground to consider it as a territory distinguished for high landscape diversity where there takes place a process of **reanthropogenesis.** It is distinguished for its high landscape diversity and is a protected territory of some kind. In the category of our research region – "The hilly erosion-accumulative foothills, built of conglomerates, clays and sand-stones, with complexes of beard grass steppes, shibliak, phrygane and wood derivatives, on carbonate, typical brown and sometimes black earth soils".

The data of Martkopi physical-geography stationary may be extended in the entire 17th category of the research region. By expedition and semi-expedition studies and with the use of large-scale maps we can distinguish and give a detailed description of the respective morphological units. And for the remaining 15th, 16th and 18th categories it is possible to establish a definite correlation and detailed characterization.

By using "semaphore maps", I have assigned a red color to the sections of tension (Kaspi, Kazreti), a green color to the territory distinguished for high landscape diversity (Martkopi) and for remaining territory of the research region yellow color would probably be relevant.

ფიზიკური გეოგრაფია

საქართველოს სუბხმელთაშუაზღვიური სემიჰუმიდური მთისწინების ბუნებრივ ლანდშაფტთა ანთროპოგენური ტრანსფორმაცია

თ. მამუკაშვილი*

* *აკადემიის წევრი, ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

(წარმოდგენილია აკადემიკოს ზ. ტატაშიძის მიერ)

ნაშრომში დადგენილია სუბხმელთაშუაზღვიური სემიჰუმიდური მთისწინების ბუნებრივ ლანდშაფტთა მდგრადობა ანთროპოგენური ზემოქმედების პირობებში და განსაზღვრულია საკვლევ რეგიონში ეკოლოგიური დაძაბულობის უბნები და მაღალი ლანდშაფტური მრავალფეროვნებით გამორჩეული ტერიტორია.

REFERENCES

1. *N. Beruchashvili, T. Dekanoidze, N. Jamaspishvili* (2003), The Caucasus Geographical Journal, No. 3.
2. *N. Elizbarashvili* (2005), The Geoecological Fundamentals of Landscape Planning. Tbilisi.
3. *Z. Seperteladze, T. Mamukashvili* (2004), Geography of Georgia, No. 3.

Received December, 2006

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Biological Sciences

Botany

Botany – Progress and Prospects

Peter H. Raven*

* *Missouri Botanical Garden, St. Louis, Mo., USA*

ABSTRACT. Progress in botany has been very rapid over the past half century. Earlier, it was scarcely possible to imagine what would be learned about hormone action, growth, or many other key fields of botany. A few of important new discoveries are mentioned here. The complete sequencing of the genomes of different kinds of plants began in 2000. The information gained from genomic studies soon will make possible a rigorous study in plant evolutionary history. At the same time, the accelerating effects of habitat reduction; the widespread and growing presence of alien invasive species; the gathering of plants in nature for personal use and commercial purposes; and global warming combined threaten to eliminate two-thirds or more of all plant species during the course of this century. As scientists and informed citizens, we must give the global situation our most serious attention. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *genome, hormone action, plant evolutionary history.*

Progress in botany has been very rapid over the past half century, with problems that had been thought to be intractable yielding to modern methods that would earlier have been impossible to conceive, much less to apply. Although Watson and Crick proposed the double helical model for the structure of DNA in 1953, it was not until the unraveling of the genetic code in the early 1960s and the demonstration of how the transcription of proteins actually worked that we began to understand the basic tenets of molecular biology and their central role for the whole field of biology. Earlier, it was scarcely possible to imagine what would be learned about hormone action, growth, or many other key fields of botany, and few people had any premonition of the massive changes that would alter science and the face of the globe over the past 50 years [1,2].

The world population in 1950 was approximately 2.5 billion people; in the ensuing years it has grown to an estimated 6.5 billion! We were probably using about 50% of global productivity then on an ongoing basis, whereas now our rate of use is estimated at 120%. With half of the world's people living on less than $2 per day and one out of eight literally starving, the combination of population numbers, consumption levels, and technology is literally reducing the productive capacity of the earth to lower and lower levels with every passing day. Why do we seem to care so little? I first became aware of these problems along with Paul Ehrlich and other colleagues at Stanford about 40 years ago. Gradually it became apparent to us that we were driving species to extinction at an ever-increasing rate, higher by 2-3 orders of magnitude than anything experienced for the past 65 million years, and that the cost of our existence on Earth was rapidly becoming an insupportable burden [3-5].

But what have we learned about botany over these years? There have been so many surprises and important discoveries that I can offer only a few for consideration here.

The complete sequencing of the genomes of different kinds of plants began in 2000, with *Arabidopsis,* which was shown minimally to be of ancient hexaploid derivation, even though it has the smallest genome among the flowering plants. Complete genomic se-

quences were soon added for rice, poplar, the lycophyte *Selaginella*, with tomatoes, maize, wine grape, and the legume *Medicago* soon to follow. These sequences have made it possible to identify the gene corresponding to a mutation in *Arabidopsis* within a few weeks of the formation of a segregating population between a plant containing a mutant allele and the wild type. The DOE Community Sequencing Program has also supported the ongoing sequencing of genomes or major sections of genomes in manioc, potato, tobacco, a moss, *Mimulus guttatus*, *Sorghum bicolor*, *Capsella bursa-pastoris,* and *Arabidopsis lyrata,* a perennial congener of *A. thaliana* [1,2,6].

The large amount of related activity, evident in the pages of any journal in the field, reminds us that the "$1,000 genome" is going to become a reality sooner rather than later. As costs decrease rapidly, we will have greatly enhanced ability to assess variation within and among populations, thus making it much easier to learn about adaptation and the process of evolution in plants and to conserve them.

The information gained from genomic studies soon will make possible a rigorous study of the evolutionary history of key innovations in plant evolutionary history such as vascular tissue, leaves, seeds, and flowers. It will also make possible the determination of the genetic basis of significant innovations in the features of mutant individuals and different species and genera of plants. Elegant QTL work in *Mimulus* has documented the genetic basis for species differences. And, genomic studies have also led to a better understanding of the genetics of domestication, particularly in maize, where key genes have been identified and results of selection for a chromosome region documented (selective sweeps) [7,8].

Also discovered during the last few years is the fact that plant disease-resistant proteins (R proteins) usually detect pathogens indirectly by the damage they do to host cell components, rather than by identifying the pathogen's molecules directly. Recent studies have shown that at least one case of "non-host resistance" (i.e., in which a plant species does not allow growth of a particular pathogen species) is due to active resistance mechanisms encoded by multiple genes. This raises the possibility that it may be possible to engineer stable non-host resistance into crop species.

Of fundamental importance has been the independent discovery by two groups of investigators of the receptor for auxin, which was discovered some 80 years ago — the first plant hormone to be described. This receptor has turned out to be an *Arabidopsis* F-box protein (one of about 700 such proteins in Arabidopsis). Such proteins act in eukaryotic organisms to target regulatory proteins for degradation in a signal-dependent manner. This finding – a beautiful piece of work on a long-standing problem – hints at how plant cells "sense" and respond to this protein, and thus provides a key for investigating the action of plant proteins in general. Recently, the signaling mechanism by which plants sense and respond to gibberellin has also be found to involve an F-box protein, suggesting that the phytohormones may act via similar mechanisms [6,9,10].

Of special significance has been the identification of a molecule, called FT, that has all the hallmarks of the hitherto elusive florigen, and published in three articles in Science in 2005. The FT gene is induced in leaves within hours after plants receive a stimulus that promotes flowering, and its product, the FT protein, acts at the growing tips of the plant to activate the flowering process. The gap between the two sites is bridged through movement of FT RNA from the leaf to the growing tip.

The role of micro RNA, which was poorly understood in 1999, has now been shown to be important in many aspects of plant growth and development.

Transcription factors and other proteins have been shown to move in a regulated way through plasmodesmata, the "plant information superhighway," by Bill Lucas at the University of California – Davis.

Jeff Palmer and his colleagues have demonstrated the massive horizontal transfer of mitochondrial genes from diverse land plant and fungal donors to the basal New Caledonian angiosperm *Amborella.* This has been a startling discovery whose significance for plant evolution in general and mode of origin are receiving further studies in the Palmer laboratory. Certainly the horizontal transfer of genes to plant mitochondria is frequent, but not to the massive degree in which it has occurred in *Amborella* [11].

Another great surprise about angiosperm phylogenetics came in 2007, with the recognition that the tiny grasslike plants of the family Hydatellaceae, endemics of Australia and New Zealand, were the sister group of the Nymphaeaceae, the water lilies. Comparisons with the early angiosperm fossils *Archaefructus* followed, with the promise of exciting advances in understanding for many years to come.

Plant phylogenetic studies have expanded rapidly in precision and in coverage of different groups. Careful developmental studies linked with comparisons of the genetic basis for the patterns observed have much to offer in understanding the relationships of plant groups at all levels, and the basic patterns of relationship that have been emerging over the past 15 years or so – often radically different from what had been suspected earlier – appear durable. Informative studies of fossil plants have begun to teach us much about the nature of the earliest angiosperms. Along the way, numerous discoveries in the fossil record have proved patently false earlier conventional wisdom that held that there simply were not be enough fossil flowers to make any difference in our understanding of angiosperm evolution. Some of

the earliest fossil flowers apparently represent entirely extinct major taxa. By the mid and Late Cretaceous, ancient taxa with clear relationships with groups such as Chloranthaceae and other extant angiosperm clades appear. The resolution of such fossils will provide interesting results for years to come.

The Angiosperm Phylogeny Group (APG) has contributed much to the establishment of monophyletic groups in angiosperms, a trend obviously beneficial to the establishment of sound classifications with predictive value. The phylocode, with its cognitive formlessness, has provided a way of organizing information that some students of phylogeny have found useful, but, in principle, since it does not indicate the relationships of taxa nor help us locate information about them it has not been widely accepted.

Notable in recent years has been the sturdy growth of the Tree of Life Project, which will provide a sound basis for understanding the relationships of major taxa within the next few years. Numerous surprises, such as the rooting of the Equisetales within the ferns, and, controversially, the Gnetales within the conifers, will clearly be encountered along the way. With respect to the latter hypothesis, in which Gnetales are seen as sister to Pinaceae, there has been much doubt, but further critical evaluation is clearly necessary in view of the material presented.

In terms of the material available for systematic botany and its availability, the total number of plant specimens in the world's roughly 3,000 herbaria is growing at the rate of about 10 million specimens per year, with approximately 345 million specimens in the world's herbaria today. The total number of distinct vascular plant species validly described has not been reliably estimated, but there are clearly at least 325,000 of them, with what I would estimate as 100,000 more still to be named and defined. Over 100,000 species are cultivated in botanical gardens already, and the gardens themselves have grown by about a third over the past decade, with about 2,700 operating today.

Major increases in the availability of information about plants on the World Wide Web foretell even greater increases in such useful information in the future. For example, the African Plants Index now includes high-resolution images of about 80% of the types of African plant species, and will go on line later this year; and a similar project for the types of Latin American plants, again backed by the A. W. Mellon Foundation of New York, was started in 2005.

The literature of systematic botany is likewise becoming available on line: the Missouri Botanical Garden's *Botanicus* project, funded by the Keck Foundation and the Institute of Museum and Library Services, has already recorded over 400,000 pages of pre-1923 systematic literature in a searchable format, with 2,500 additional pages being added each week. In this way, the complete literature of systematic botany will become universally available relatively soon. All of the *Botanicus* information is linked to the Missouri Botanical Garden's *Tropicos* 2 database, the most comprehensive and widely consulted database on plants.

At the same time, the accelerating effects of habitat reduction; the widespread and growing presence of alien invasive species; the gathering of plants in nature for personal use and commercial purposes; and global warming combined threaten to eliminate two-thirds or more of all plant species during the course of this century. Following a call for increased efforts to conserve the world's plants at the 1999 International Botanical Congress in St. Louis, a Global Strategy for Plant Conservation was approved within the Convention for Biological Diversity in 2002. The Global Strategy then established specific, ambitious goals for the preservation of plant diversity that are intended to be met by 2010. The efforts made to realize these goals are clearly having an important impact on plant conservation throughout the world, starting with our knowledge about the amount of diversity that exists.

Transgenic crops have now been grown on more than 1 billion acres (in aggregate) throughout the world, amounting to approximately an eighth of the total cultivated land globally. More than a decade of experience has demonstrated no damage related to the cultivation of these crops, which have offered proven economic and environmental benefits. Even more impressive gains are in sight for the decades to come. Plants that exhibit improved levels of cold, freezing, salt, and drought tolerance have been developed and are expected to improve crop productivity in regions where it is limited by these factors. Although investigations continue, there seems little doubt that the intensive production of adequate supplies of food on the least amount of land possible will contribute a great deal to the preservation of biodiversity at a time of maximum threat to its continued existence.

The appearance of the Millennium Ecosystem Assessment in 2005 demonstrated that human beings over the past 50 years have degraded ecosystems more rapidly than any earlier time. These changes have allowed major increases in human well-being but at the same time rapidly diminished the benefits that future generations will be able to obtain from ecosystems. Achieving the Millennium Development Goals will require significant changes in policies, institutions, and practices, but few nations seem ready to embrace these changes fully and certainly they have not been accepted as a basis for action here in the United States.

As scientists and informed citizens, we must give the global situation our most serious attention. We live in a more diverse world today than will ever exist again, but our efforts will play a major role in shaping the contours of that future world and the opportunities that its citizens will enjoy. As botanists, we have a great deal to

contribute, and exciting future discoveries await us in all of the subfields of our discipline.

Acknowledgments: This review is based on a talk presented at the XVII International Congress in Vienna, Austria, in July 2005, revised for the annual meeting of the Botanical Society of America in August 2006, and published in the Bulletin of the Botanical Society of America subsequently.

ბოტანიკა

ბოტანიკა – პროგრესი და პერსპექტივა

პ. რეივენი*

* *მისურის ბოტანიკური ბაღი, სენტ ლუისი, მონტანა, აშშ*

უკანასკნელი ნახევარი საუკუნის განმავლობაში ბოტანიკა სწრაფად განვითარდა. უწინ ძნელად წარმოსადგენი იყო ის, რაც ახლა ვიცით ჰორმონების მოქმედების, ზრდის და ბოტანიკის ბევრი სხვა მნიშვნელოვანი საკითხის შესახებ. აქ საუბარია რამდენიმე ღირსშესანიშნავ ახალ აღმოჩენაზე. სხვადასხვა მცენარის გენომების სრული გაშიფრვა 2000 წ. დაიწყო. გენომების გამოკვლევის შედეგად მიღებული ცოდნა მალე შესაძლებელს გახდის მცენარეთა ევოლუციის ინტენსიურ შესწავლას. ამავე დროს, ჰაბიტატთა მზარდი შემცირება, უცხო ინვაზიური სახეობების ფართო გავრცელება, მცენარეთა შეგროვება ველურ ბუნებაში პირადი და კომერციული საჭიროებისათვის და გლობალური დათბობა საფრთხეს უქმნის მცენარეთა ყველა სახეობის ორი მესამედის ან მეტის არსებობას ამ საუკუნეში. როგორც მეცნიერები და გათვითცნობიერებული მოქალაქეები, ვალდებულები ვართ გლობალურ ვითარებას უდიდესი ყურადღებით მოვეკიდოთ.

REFERENCES

1. *P. H. Raven* (Editor) (1997), Nature and Human Society. National Academy Press, Washington DC.
2. *P. H. Raven & L. R. Berg* (2001), Environment. Harcourt College Publishers, Philadelphia, Pennsylvania.
3. *P. Ehrlich & P. H. Raven* (1965), Butterflies and plants: a study in Coevolution. Evolution 18: 586-603.
4. *A. O. Richardson & J. D. Palmer* (2007), Horizontal gene transfer in plants. J. Exp. Bot. 58 (1): 1-9.
5. *G. B. Johnson & P. H. Raven* (2004), Biology. Holt, Rinehart & Winston Publishers, Austin, Texas.
6. *J.-Y. Lee, K. Taoka, B.-C. Yoo, G. Ben-Nissan, D.-J. Kim, & W.J. Lucas* (2005), Plasmodesmal-associated protein kinase in tobacco and *Arabidopsis* recognizes a subset of non-cell-autonomous proteins. Plant Cell 17: 2817-2831.
7. *A. L. Angert* (2006), Demography of central and marginal populations of monkeyflowers (*Mimulus cardinalis* and *M. lewisii*). Ecology 87(8): 2014-2025.
8. *R. L. Gilbertson, M. R. Rojas, & W. J. Lucas* (2005), Plasmodesmata and phloem: conduits for local and long-distance signaling. Annual Plant Reviews 18: 162-187.
9. *P. Ehrlich & P. H. Raven* (1969), Differentiation of populations. Science 165: 1228-1232.
10. *A. L. Angert* (2006), Growth and leaf physiology of monkeyflowers with different altitude ranges. Oecologia 148:183-194.
11. *U. Bergthorsson, A. O. Richardson, G. J. Young, L. R. Goertzen, & J. D. Palmer* (2004), Massive horizontal transfer of mitochondrial genes from diverse land plant donors to basal angiosperm *Amborella*. PNAS, 101, no 51:17747-17752.

Received June, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Molecular Biology

Function of the Yeast Rad57 Protein Depends on a Conserved Lipid-Binding Motif

Gela G. Tevzadze*

* *University of Chicago, Chicago, Illinois, USA*

ABSTRACT: The Rad57 protein of the yeast *Saccharomyces cerevisiae* has been extensively studied and is known to have multiple functions critical for vegetative growth and meiosis. Here we present evidence suggesting that these functions are regulated by lipid binding. This binding likely occurs via a GW…W motif conserved in the phosphatidylinositol kinase (PIK) protein family and in a meiosis-specific yeast protein Spo73. Earlier studies indicate these proteins respond to and are regulated by phosphatidylinositol phosphate (PIP) or phosphoinositide signaling molecules. Specific mutations in the GWLVGW motif of Rad57 abolish the ability of lipid binding, and significantly impair Rad57 functions essential for vegetative growth, meiotic recombination and spore germination. These results imply the multiple roles of Rad57 are governed via lipid signaling, and can be modulated by direct binding of specific lipid second messengers to a specific motif in Rad57p. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *lipid-binding, phospatidylinositol phosphates, Rad57, yeast, spore viability*

INTRODUCTION. Lipid signaling is considered one of the most potent and diverse mechanisms regulating various important cellular processes, such as protein traffic and re-localization, exo- and endocytic vesicle formation, actin nucleation and cytoskeletal assembly. Phosphatidylinositol (PI) and its phosphorylated derivatives (PIP), which act as signaling molecules, have several characteristics which make them ideal for this function: *i*) their low cellular abundance compared to other lipids (e.g., phosphatidylserine, phosphatidylcholine); *ii)* multiple sites of phosphorylation within the inositol ring, providing the ability to generate wide variety of differentially phosphorylated derivatives from a single progenitor (PI); *iii)* rapid interconversion between various PIPs due to activities of multiple lipid kinases and phosphatases [1]. Regulation by lipid signaling often includes modular protein domains that bind specific PIP-molecules [2]. In addition to well-characterized domains (FYVE, PH, PX, C2), there are indications that several (less obvious) motifs can bind lipids as cofactors required for enhancement of their biochemical activity and/or for maintaining protein stability. Lipid-binding may also occur as a secondary function within a domain with a previously known distinct role [3, 4].

In our previous studies, we identified a motif (GWC..W), shared by the PIK-related kinases and a mid-late meiosis-specific protein Spo73 (Tevzadze *et al.*, in preparation). This motif is the only common fragment for Spo73p and TOR-kinases, which are quite disparate in size (14 kDa for Spo73, >200 kDa for TORs) as well as in function: Spo73 is involved in a specific step of spore wall formation, whereas TOR kinases (components of the TORC1 and TORC2 complexes) control wide variety of cellular functions, e.g., translation, transcription, autophagy, and nutrient transport [5-8]. TORs are lipid-kinase homologs, but have only protein kinase activity [9]. Spo73 activity is dependent on a putative phospholipase Spo1, which is thought to produce Lyso-PI second messengers [10-13]. These observations led us to suggest that the motif acts as a sensor of lipid signals, i.e., is involved in direct binding to second messenger molecules produced by upstream activators. Indeed, the GWC..W motif was shown to be critical for protein function, as well as for binding specific lipids both in Spo73 and TOR kinases (Tevzadze *et al.*, in preparation). The GW pair is conserved in the PIK-kinase protein family

and is critical for binding. For Spo73, the presence of either residue is important for meiotic function of the protein and its PIP-binding abilities.

The importance of the GW pair prompted us to examine the distribution of this motif or its variants throughout the yeast genome. In this paper, identification of GWlvGW in the Rad57 protein is reported and its importance demonstrated in PIP-binding and Rad57p function during vegetative growth and meiosis.

MATERIALS AND METHODS

Strains: *E. coli* DH5α strain was used for the maintenance and propagation of plasmids, and BL21(DE3) was employed for expression of the Rad57 protein fragments from an IPTG-inducible promoter in the pGEV2 plasmid ([14]; see also below). The *S. cerevisiae* strains used in this study are listed in Table 1.

Plasmids: *Plasmids for bacterial expression of Rad57p internal fragments.* An internal ~100 bp *RAD57*-fragment containing the sequence encoding the ^{298}GWLVGW^{303} domain was PCR-amplified from the yeast genomic DNA. The *Bam*HI and *Xho*I sites, engineered at the 5' and 3' termini, respectively, were used for cloning into the pGEV2 vector. The pPW24C plasmid contains this fragment from *RAD57* (bases 811-909) fused by its N-terminus to the IgG-binding domain of Protein G; it encodes residues 271-303 (the motif is at the C-terminus). Two derivatives of pPW24C contain mutations which alter the ^{298}GWLVGW^{303} motif to ^{298}AWLVAW^{303} (pPW29) or to ^{298}GALVGA^{303} (pPW30) in the otherwise identical fragments encoding residues 271-303.

Plasmids for integration of the rad57 *mutant alleles:* For introduction of the mutant ^{298}AWLVAW^{303} or ^{298}GALVGA^{303} alleles, a 2 kb *Sac*I fragment (containing

Table 1

Yeast Strains Used in this Study

Strain	Relevant Genotype	Source
W303-1A	*MAT***a** *ade2 can1-100r his3-11,15 leu2-3,112 trp1-1 ura3-1*	R. Rothstein*
W303-1B	*MATα ade2 can1-100r his3-11,15 leu2-3,112 trp1-1 ura3-1*	R. Rothstein*
GTY484	W303-1A *RAD57:URA3:rad57(5'Δ)*	This study
GTY485	W303-1B *RAD57:URA3:rad57(5'Δ)*	This study
GTY486	W301-1A *rad57-g,g:URA3:rad57(5'Δ)*	This study
GTY487	W303-1B *rad57-g,g:URA3:rad57(5'Δ)*	This study
GTY490	W301-1A *rad57-w,w:URA3:rad57(5'Δ)*	This study
GTY491	W303-1B *rad57-w,w:URA3:rad57(5'Δ)*	This study
GTY494	W303-1A *rad57Δ:TRP1*	This study
GTY495	W303-1B *rad57Δ:TRP1*	This study
GTY502	W303-1A *rad57Δ:KanMX6*	This study
GTY503	W303-1B *rad57Δ:KanMX6*	This study
SK1-A	*MAT***a** *arg4 his4X::LEU2 ho::LYS2 leu2::hisG lys2 ura3*	D. Bishop**
SK1-B	*MATα arg4 his4B::LEU2 ho::LYS2 leu2::hisG lys2 ura3*	D. Bishop**
GTY508	SK1-A *RAD57:URA3:rad57(5'Δ)*	This study
GTY509	SK1-B *RAD57:URA3:rad57(5'Δ)*	This study
GTY512	SK1-A *rad57-g,g:URA3:rad57(5'Δ)*	This study
GTY513	SK1-B *rad57-g,g:URA3:rad57(5'Δ)*	This study
GTY516	SK1-A *rad57-w,w:URA3:rad57(5'Δ)*	This study
GTY517	SK1-B *rad57-w,w:URA3:rad57(5'Δ)*	This study
GTY520	SK1-A *rad57Δ:KanMX6*	This study
GTY521	SK1-B *rad57Δ:KanMX6*	This study

* Columbia University, New York
** The University of Chicago

1 kb of the 3' region of the *RAD57* ORF and 1 kb of the 3' UTR) was PCR-amplified from yeast genomic DNA and cloned into pRS306, a *URA3*-marked integrative vector [15]. The G298A G302A ("g, g") and W299A W303A ("w, w") mutations were introduced by the recombinant PCR method [16], and the amplified fragments were sequenced throughout their length to confirm that no additional mutations were generated due to the PCR errors. To introduce the resulting pPW35 (pRS306-*RAD57*), pPW36 (pRS306-*rad57-g,g*) and pPW39 (pRS306-*rad57-w,w*) plasmids into the yeast genome, they were linearized by the *Bgl*II restriction endonuclease, which has a unique recognition site within the *RAD57* ORF, ~200 bp upstream of the fragments encoding the GWLVGW wild-type or mutant motifs. Integration of these plasmids into the *RAD57* locus created an insertion of the *URA3* marker, flanked by a full-size *Rad57* (wild-type or mutant) allele on one side and a 5' deletion allele of *RAD57* (lacking the 5'UTR and ~250 bp of the 5' ORF) on the other. Proper integration was confirmed by PCR, sequencing and by mating with *rad57Δ* strains (the deletions marked with *TRP1* or *KanMX6* for W303, and *KanMX6* for SK1). Sporulation and tetrad analysis of the heterozygous *rad57Δ* diploids demonstrated segregation of *URA3* opposite to the *rad57Δ*-markers.

Media, Growth, and Sporulation: *E. coli* growth media (LB and SOC) were described previously [17]. The yeast media YPDA, YPA, Synthetic Complete (SC) for growth, and SPII and SPIII for sporulation were described previously [18, 19]. Isolation of genomic DNA for Southern blot analysis was performed as described elsewhere [20]. Transformation of yeast cells was performed by high-efficiency lithium acetate protocol [21, 22]. RbCl-mediated transformation of *E. coli* strain DH5a was performed as described earlier [12, 23]. Sporulation of the SK1 background strains was done as described [24], heteroallelic recombination frequency was assayed based on the number of His+ colonies forming as a result of recombination between the *his4-X* and *his4-B* heteroalleles [24, 25].

Assay of γ-Radiation Sensitivity: The W303 background haploid strains were grown overnight in 5 ml YPDA at 30°C with aeration and diluted to OD600=1.0 in the same volume and media. Cultures were treated with 800 Gy (at 0.5 Gy/sec) in a GammaCell 220 industrial irradiator equipped with the cobalt-60 source (Nordion). After irradiation, cultures were serially diluted in YPDA and 2.5 μl of 10^0 to 10^{-5} dilutions were spotted on YPDA plates together with untreated controls. Plates were incubated at 30°C and 23°C, since the *rad57* null mutants are ts regarding their sensitivity to radiation. While radiation sensitivity for the control *rad57Δ::TRP1* deletion strains was evident at both temperatures, incubation at 23°C did give the most dramatic differences between the isogenic *RAD57* and *rad57Δ* strains.

Lipid-Binding Assays using PIP-strips (Echelon Biosciences Inc., Salt Lake City, Utah, USA) were done as suggested by manufacturer (www.echelon-inc.com), with several modifications. The fragments of Rad57 alleles expressed in the BL21(DE3) bacterial strain (see above) were diluted to 0.5 μg/ml in TBST (100 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.1% Tween 20) containing 0.1% ovalbumin (Sigma) and incubated with PIP-strips or arrays for 1 h at room temperature. The strips were then washed 3 times for 15 min in TBST + 0.1% ovalbumin and incubated again for 1 h at room temperature with horseradish peroxidase-conjugated anti-Protein G polyclonal antibody raised in rabbit (1:10,000 dilution, AbCam). Strips were washed as before and activity of HRP bound to the filter was assayed by ECL (Amersham). Signal intensities for individual lipid spots were calculated using the NIH Image 1.62 software and normalized to the control spots containing no lipids.

RESULTS AND DISCUSSION

Search for the *S. cerevisiae* Proteins Containing the GW...W Motif: The search was done using the Pattern Matching (PatMatch) algorithm provided by the Stanford Genome Database (http://db.yeastgenome.org/cgi-bin/ PATMATCH/nph-patmatch). PatMatch allows to search for the presence of short (<20 residues) peptide sequences in the yeast ORFs. The GW...W motif was found in 33 proteins (Table 2) among the 5869 sequences (i.e., the entire *S. cerevisiae* genome) searched. The more restricted searches specified smaller sets, e.g., 3 hits for GWC..W (Tor1, Tor2, Spo73); 4 hits for GW.Y.W (Ygl262w, Mep1, Mep3, Spo73) and GW..SW (Sup35, Ypl162c, Eaf3, Spo73) each. These variations were chosen since they are present in the Spo73 sequence (GWCSYW), where this motif was initially identified. Finally, a duplication of the GW pair, separated by two other residues (GW..GW), was identified in only two yeast proteins: GWlvGW for Rad57p (residues 298-303 in the 460 aa protein) and GWriGW for Bna3p (278-282, 444 aa).

The *BNA3* locus encodes arylformamidase, involved in biosynthesis of nicotinic acid from tryptophan via kynurenine pathway [26]. Studies of this protein were considered beyond the scope of this paper (and our expertise). Since the presence of the GW..GW motif in Rad57 raised an intriguing possibility of meiotic recombination being regulated by lipid signaling, here we concentrated on the importance of the GWlvGW motif in Rad57 function.

The *RAD57* gene encodes a 52 kDa-protein that stimulates strand exchange by stabilizing Rad51p binding to single-stranded DNA via forming a heterodimer with Rad55 [27]; it is involved in the recombinational repair of double-strand breaks in DNA during vegetative growth [28] and meiosis [29], and has a minor role

Table 2

Proteins with the GW...W Motif in the *S. cerevisiae* Genome

Locus/Gene*	Motif	Starts	Ends**	Locus Information (*Saccharomyces* Genome Database Annotation)
YAL048C/GEM1	GWLAQW	390	395	Mitochondrial membrane GTPase, regulating mitochondrial morphology
YBR021W/FUR4	GWFQRW	592	597	Uracil permease, localized to the plasma membrane
YBR136W/MEC1	**GWLPFW**	**2363**	**2368**	Checkpoint protein and PI kinase superfamily member; monitors and participates in meiotic recombination
YBR155W/CNS1	GWISKW	369	374	TPR-containing co-chaperone
YCR026C/NPP1	GWRSAW	36	41	Hypothetical protein
YCR101C	GWGVVW	53	58	Hypothetical protein
YDR004W/RAD57	**GWLVGW**	**298**	**303**	Protein that stabilizes the binding of Rad51p to single-stranded DNA; forms heterodimer with Rad55p
YDR093W/DNF2	GWTGIW	1394	1399	Non-essential P-type ATPase, localizes to the plasma membrane, late exocytic or early endocytic membranes
YDR172W/SUP35	GWYLSW	306	311	Translation termination factor eRF3; altered protein conformation creates the [PSI(+)] prion
YDR234W/LYS4	GWFLKW	640	645	Homoaconitase, catalyzes a step in the lysine biosynthesis pathway
YDR335W/MSN5	GWFSVW	219	224	Karyopherin involved in nuclear import and export; cargo dissociation involves binding to RanGTP
YER046W/SPO73	**GWCYSW**	**108**	**113**	Meiosis-specific protein, required for spore wall formation but dispensible for both nuclear divisions
YGL262W	GWKYYW	82	87	Hypothetical protein
YGR121C/MEP1	GWAYQW	164	169	Ammonium permease
YIR028W/DAL4	GWFQRW	594	599	Allantoin permease
YJL016W	GWVKVW	205	210	Cytoplasmic protein of unknown function
YJL060W/BNA3	GWRIGW	277	282	Arylformamidase, involved in biosynthesis of nicotinic acid from tryptophan via kynurenine pathway
YJL133W/MRS3	GWKGFW	278	283	Mitochondrial iron transporter of the mitochondrial carrier family (MCF), functionally redundant with Mrs4p
YJR066W/TOR1	**GWCPFW**	**2465**	**2470**	PIK-related protein kinase and rapamycin target; subunit of TORC1
YKL203C/TOR2	**GWCPFW**	**2469**	**2474**	PIK-related protein kinase and rapamycin target; subunit of TORC1 and TORC2
YKR052C/MRS4	GWKGFW	268	273	Mitochondrial iron transporter of the mitochondrial carrier family (MCF), functionally redundant with Mrs3p
YLR088W/GAA1	GWLPCW	587	592	Subunit of the GPI-protein transamidase complex, removes the GPI-signal, attaches GPI to proteins in the ER
YLR195C/NMT1	GWKKDW	128	133	N-myristoyl transferase, acts on several proteins involved in cellular growth and signal transduction
YMR078C/CTF18	GWLRQW	140	145	Subunit of a complex with Ctf8; is required for sister chromatid cohesion and the DNA damage checkpoint
YMR266W/RSN1	GWIFFW	157	162	Membrane protein of unknown function

Table 2 (continued)

Locus/Gene*	Motif	Starts	Ends**	Locus Information (*Saccharomyces* Genome Database Annotation)
YNL172W/APC1	GWPDLW	663	668	Largest subunit of the Anaphase-Promoting Complex/Cyclosome (APC/C)
YOR376W-A	GWTILW	6	11	Identified by fungal homology and RT-PCR
YPL162C	GWSDSW	191	196	Hypothetical protein
YPL257W	GWFLIW	75	80	Hypothetical protein
YPR023C/EAF3	GWKSSW	83	88	Esa1p-associated factor, nonessential component of the NuA4 acetyltransferase complex
YPR127W	GWVKHW	281	286	Hypothetical protein
YPR138C/MEP3	GWAYQW	163	168	Ammonium permease of high capacity and low affinity
YPR193C/HPA2	GWQRLW	21	26	Tetrameric histone acetyltransferase; acetylates histones H3 and H4 *in vitro*, exhibits autoacetylation activity

* Proteins relevant for the present and related studies (Rad57, Spo73, PIK-kinase family proteins) are shown **in bold**
** Amino acid residue coordinates where the motif **Starts** and **Ends** are shown for each protein.

Note: several other PIK-kinase protein family members contain the GW.PF. variant of the motif: Tel1 (GWSPFY, 2782-2787), Mec1 (GWLPFW, 2363-2368), Sch9 (GWSPFF, 608-613), Tor1 (GWCPFW, 2465-2470), Tor2 (GWCPFW, 2469-2474). The Cbk1 serine-threonine protein kinase contains GWPPFC (607-612).

in mating efficiency [30]. Rad57 participation in the assembly of recombinational complexes was also demonstrated by immunocytological methods [30]. Null alleles of *rad57* exhibit cold-sensitive phenotype, consistent with the role of Rad57 in stabilization of the DSB repair/recombination complexes [31]. This notion is further supported by the fact that *rad57* mutant phenotype can be suppressed by high-copy *RAD51* [32] or by specific *rad51* mutant alleles with the enhanced abilities of DNA-binding [33]. Most likely, suppression occurs due to the increased stabilization of Rad51-binding to DNA, thus rendering Rad57 function obsolete. Cold-sensitive *rad57* alleles have decreased fidelity of DSB repair, as demonstrated by double-strand gap (DSG) repair studies: significant portion of plasmid-borne DSGs are not restored accurately, indicating the mechanism of re-circularization of a "DSG-linearized" plasmid is ligation of cohesive ends, rather than recombinational repair [34]. Cold-sensitivity is also evident in the enhanced sensitivity of *rad57* mutants towards ionizing radiation (IR), with more dramatic phenotype at 23°C compared to 30°C. Furthermore, *rad57* haploids exhibit higher IR-sensitivity compared to isogenic diploids [35].

Based on our recent findings that the GW…W motif is important for PIP-binding of Spo73 and the PIK-related kinase protein family (Tevzadze *et al.*, in preparation), we proposed that Rad57 also binds specific phosphatidylinositols and binding ability is defined by the presence of the GW...GW domain. This paper reports that: 1) Rad57 binds specific phoshatidylinositol phosphates (PIPs), but not unphosphorylated PI; 2) The strongest binding is detected for $PI(3,4)P_2$ and $PI(3,4,5)P_3$; 3) mutations in the GW..GW domain reduce (Gly/Ala) or abolish (Trp/Ala) binding; 4) The Trp/Ala (i.e., the motif GALVGA), but not Gly/Ala (AWLVAW) mutations, generate loss-of-function alleles. Loss of function was demonstrated both in vegetative growth and sporulation by assaying γ-radiation sensitivity, spore viability and meiotic recombination efficiency.

A Fragment of Rad57 with the C-Terminal GWlvGW Motif Binds PIPs: In previous studies [12], our approach has been to test lipid-binding abilities of a protein by expressing its relatively small fragment (encoded by ~100-110 bp) as a fusion with Protein G in a bacterial system [14, 36], purify the expressed protein and examine its binding to various groups of lipids spotted on a nitrocellulose filter. This approach allows to relatively quickly test lipid-binding and, due to a small size of the fragment, does not generally create solubility problems. To further employ this strategy, lipid-binding of Rad57 was examined for a 33 aa fragment (residues 271-303) containing GWlvGW at its extreme C-terminus. This fragment exhibits strong binding to PIPs with a significant preference towards $PI(3,4)P_2$ and $PI(3,4,5)P_3$ (Fig. 1, left panel).

Next, we inquired whether introducing mutations in the GW..GW residues will alter lipid-binding abilities of Rad57. Replacing both glycines to alanines caused a significant reduction in PIP-binding (Fig. 1, central panel). The effect of replacing the tryptophans to alanines was even more dramatic, completely abolishing lipid-binding for *rad57-w,w* (Fig. 1, right panel).

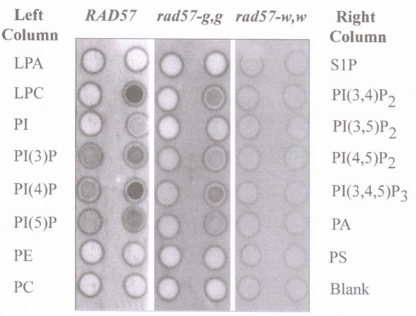


Fig. 1. Typical results of lipid-binding overlay assays, using PIP-strips (Echelon Biosciences Inc.), for the Rad57 fragments containing the wild-type motif GWLVGW (*RAD57*, left panel) or mutated AWLVAW (*rad57-g,g*; central panel) and GALVGA (*rad57-w,w*; right panel) motifs. The strips were successively incubated first with the fusion proteins purified on IgG sepharose columns (Amersham) and then with anti-Protein G antibody conjugated with horseradish peroxidase (AbCam) to detect the protein bound to lipid spots. The PIP-strips contain 100 pmol spots of 15 different lipids and a control blank spot. The positions of test lipids on the left and right sides of the strips are shown on both sides of the test samples: LPA, Lysophosphatidic acid; LPC, Lysophosphocholine; PI, phosphatidylinositol (unphosphorylated inositol ring); PI(3)P, phosphatidylinositol-3-monophosphate; PI(4)P, phosphatidylinositol-4-monophosphate; PI(5)P, phosphatidylinositol-5-monophosphate; PE, phosphatidylethanolamine; PC, phosphatidyl-choline; S1P, sphingosine-1-phosphate; $PI(3,4)P_2$, phosphatidylinositol-3,4-biphosphate; $PI(3,5)P_2$, phosphatidylinositol-3,5-biphosphate; $PI(4,5)P_2$, phosphatidylinositol-4,5-biphosphate; $PI(3,4,5)P_3$, phosphatidylinositol-3,4,5-triphosphate; PA, phosphatidic acid; PS, phosphatidylserine; the lowest spot on the right does not contain any lipids.

One possible explanation of the mutant alleles' inability to bind PIPs is a low stability of the mutated proteins. However, the abundance and stability of bacterially expressed wild-type and mutant proteins were comparable (monitored on PAGE gels by Coomassie staining, data not shown), indicating that decreased (for "g,g") or abolished (for "w,w") ability to bind PIPs genuinely reflects a change in Rad57 function(s), rather than low levels of the mutant proteins.

Thus, the GWlvGW motif in Rad57 is important for binding the PIP-molecules, which is indicative of regulation via lipid signaling. To further explore this possibility, we inquired how the Gly/Ala and Trp/Ala mutations within GWlvGW affect the diverse functions of Rad57p.

The Tryptophans in the GWlvGW Motif Are Also Essential for Rad57 Function. The effect of both mutants on Rad57 function was examined using two approaches: first, their sensitivity to IR was assayed in vegetatively growing haploid strains and compared to wild-type and deletion mutants. Haploid strains were chosen since they are more sensitive to IR than isogenic diploids [35]. Second, the effect on sporulation was monitored by determining spore viability (in the W303 and SK1 backgrounds) and heteroallelic recombination at the *his4-X/his4-B* hot-spot in SK1 [25].

Tests of radiation sensitivity clearly show that the *rad57-w,w* allele renders the cells IR-sensitive at levels similar to the deletion allele (Fig. 2, compare row 5 to rows 7 and 8), whereas the *rad57-g,g* allele is almost identical to wild-type (Fig. 2, compare rows 3 and 4).

Meiotic recombination in strains carrying the *rad57-w,w* alleles is impaired as well: while the frequency of His+ colonies (forming as a result of recombination between the *his4-X* and *his4-B* heteroalleles) is 200-500 per 10^4 viable cells for wild-type and *rad57-g,g* strains, the *rad57-w,w* diploids accumulate only 40-60 His+ colonies per 10^4 viable cells. This defect, however, is not as dramatic as the one for *rad57Δ*, where only 4-8 His+ colonies per 10^4 viable cells are detected. Thus, the efficiency of meiotic recombination, monitored at the *his4* loci, is significantly decreased in *rad57-w,w* mutants compared to wild-type and *rad57-g,g*.

Consistently, spore germination efficiency is lower in *MAT***a**/*MATα rad57-w,w* diploids than in the isogenic *RAD57* and *rad57-g,g* strains, derived both from W303 and SK1 backgrounds, although the survival for *rad57-w,w* is higher than for *rad57Δ* strains (Table 3). Interestingly, while the viability of spores derived from *RAD57/rad57Δ* and *rad57-g,g/rad57Δ* heterozygotes is almost indistinguishable from *RAD57* homozygous diploids, spore viability of *rad57-w,w/rad57Δ* is dramatically reduced even compared to *rad57-w,w* homozygotes, although still higher than for *rad57Δ* diploids. This result indicates that while the "w,w" mutant allele function is severely impaired, it's not completely abolished. Possibly, some residual function of Rad57p, acting indepen-

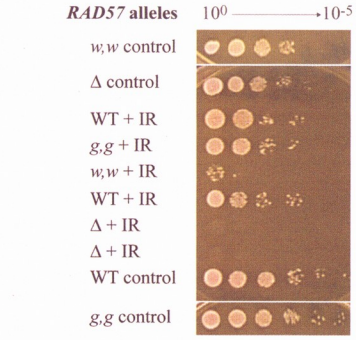


Fig. 2. Sensitivity to γ-irradiation of the wild-type Rad57 and various mutant alleles. Growth of each strain after treating with 800 Gy (denoted as "+ IR") is compared to the untreated control. Serial dilutions from 10^0 to 10^{-5}, spotted on YPDA media, are shown for each strain.

Table 3

Spore Germination in *RAD57* and Mutant Strains

Strain, Relevant Genotype	No. Tetrads Dissected	No. Complete Tetrads Survived	% Spore Viability
W303:			
a/α *RAD57*	11	11	100 (44/44)
a/α *rad57-g,g*	11	10	98 (43/44)
a/α *rad57-w,w*	33	12	73 (96/132)
a/α *rad57Δ*	11	0	5 (2/44)
a/α *RAD57/rad57Δ*	5	5	100 (20/20)
a/α *rad57-g,g/rad57Δ*	6	5	92 (22/24)
a/α *rad57-w,w/rad57Δ*	27	4	64 (69/108)
SK1:			
a/α *RAD57/rad57Δ*	5	5	100 (20/20)
a/α *rad57-g,g/rad57Δ*	6	5	92 (22/24)
a/α *rad57-g,g*	11	9	93 (41/44)
a/α *rad57-w,w*	11	0	47 (21/44)
a/α *rad57-w,w*,liquid	22	0	27 (24/88)
a/α *rad57Δ*,liquid	22	0	0 (0/88)
a/α *rad57-g,g/rad57Δ*	11	11	100 (44/44)
a/α *rad57-w,w/rad57Δ*	22	0	14 (12/88)

Sporulation was done on solid media (SPIII) at 30°C for 5 days unless noted otherwise.

Liquid sporulation was carried out in SPII at 30°C for 48 hours for a/α *rad57-w,w* and a/α *rad57Δ* SK1 homozygous strains, since the deletion strains formed no asci on solid media.

dently from the GWlvGW motif, still exists in these strains.

From these data, we infer that replacement of conserved tryptophans to alanines in the GWLVGW motif, while causing a dramatic decrease of lipid-binding, also affects Rad57 function as demonstrated by increased radiosensitivity in vegetatively growing cells, lower frequency of heteroallelic recombination, and decreased ascospore viability in meiosis. The mutant carrying the replacement of glycines to alanines, causing only a moderate effect on lipid-binding, is not defective for either of the assayed functions of Rad57, i.e., radiation resistance, induction of meiotic recombination and ascospore viability.

In conclusion, this paper presents the initial studies of a previously unexplored facet of Rad57 function, specifically, examining the possibility that Rad57 is governed by lipid signal transduction and can be controlled via direct binding of PIPs, potent second messenger molecules. Precise role(s) of PIP-binding for Rad57 is yet to be elucidated and might include, for example, 1) increasing protein activity via PIP-binding, or 2) facilitating the interaction of Rad57 with the Rad55 protein. Strong binding by $PI(3,4)P_2$ and $PI(3,4,5)P_3$ (and no binding by $PI(3,5)P_2$) is also indicative of specific binding of PIPs phosphorylated both in the third and fourth positions of the inositol ring. Quantitative analysis of Rad57-binding to various PIP-derivatives, and examining its dependence upon key enzymes regulating intracellular PIP concentration will further elucidate physiological relevance of Rad57 regulation by lipid second messengers and its involvement in lipid signaling networks.

Acknowledgements: Thanks are due to the following colleagues at the University of Chicago: Douglas K. Bishop for valuable advice and suggestions, Rochelle E. Esposito for support of this work, Ali Javaheri for his help in designing and carrying out the experiment on γ-irradiation of yeast strains, and Petra Wade for providing technical assistance.

მოლეკულური ბიოლოგია

საფუარის Rad57 ცილის ფუნქცია დამოკიდებულია ცხიმების მბმელ კონსერვატულ მოტივზე

გ. თუმზაძე*

* *ჩიკაგოს უნივერსიტეტი, ჩიკაგო, ილინოისი, აშშ*

Saccharomyces cerevisiae საფუარის ცილა **Rad57** საფუძვლიანად არის შესწავლილი და გააჩნია როგორც ვეგეტატიური ზრდისათვის, ასევე მეიოზისათვის აუცილებელი ფუნქციები. სტატიაში წარმოდგენილი შედეგების საფუძველზე შესაძლოა ვიფარაუდოთ, რომ ამ ფუნქციების რეგულირება ხდება სასიგნალო მოლეკულებთან (ფოსფატიდილინოზიტიდებთან ან ფოსფატიდილინოზიტოლთან) ბმის გზით. ეს ბმა ხორციელდება **GW...W** მოტივის საშუალებით, რომელიც კონსერვირებულია როგორც ფოსფატიდილ-ინოზიტოლ კინაზების (**PIK**) ცილების ოჯახში, ისევე მეიოზისათვის სპეციფიურ **Spo73** ცილაში. ადრინდელი შედეგების საფუძველზე მიჩნეულია, რომ ამ ცილების ფუნქცია ცხიმოვან სასიგნალო მოლეკულებთან ურთიერთმოქმედებაზეა დამოკიდებული. **Rad57**-ის **GWLVGW** მოტივის მუტირება სპობს ცხიმებთან ბმას და მნიშვნელოვნად აზიანებს ამ ცილის სხვადასხვა ფუნქციას, რომლებიც ვეგეტატიური ზრდის, მეიოზური რეკომბინაციისა და სპორების გაღივებისთვისაა აუცილებელი. ამ შედეგების ერთობლიობა მიუთითებს, რომ **Rad57**-ს როლი საფუარის უჯრედის ზრდასა და გაყოფაში იმართება ამ ცილის ცხიმოვან სასიგნალო მოლეკულებთან უშუალო ბმის მექანიზმით.

REFERENCES

1. *P. J. Cullen, G. E. Cozier, G. Banting and H. Mellor* (2001), Curr. Biol., **11,** 21: R882-R893.
2. *M. A. Lemmon* (2003), Traffic, **4,** 4: 201-213.
3. *Y. Fang, M. Vilella-Bach, R. Bachmann, A. Flanigan and J. Chen* (2001), Science, **294,** 5548: 1942-1945.
4. *T. A. Hornberger, W. K. Chu, Y. W. Mak, J. W. Hsiung, S. A. Huang and S. Chien* (2006), Proc. Natl. Acad. Sci. USA, **103,** 12: 4741-4746.
5. *K. P. Wedaman, A. Reinke, S. Anderson, J. Yates, J. M. McCaffery and T. Powers* (2003), Mol. Biol. Cell, **14,** 3: 1204-1220.
6. *S. Wullschleger, R. Loewith and M. N. Hall* (2006), Cell, **124,** 3: 471-484.
7. *K. Inoki and K. L. Guan* (2006), Trends Cell Biol., **16,** 4: 206-212.
8. *C. K. Tsang and X. F. Zheng* (2007), Cell Cycle, **6,** 1: 25-29.
9. *T. Hunter* (1995), Cell, **83,** 1: 1-4.
10. *G. G. Tevzadze, A. R. Mushegian and R. E. Esposito* (1996), Gene, **177,** 1-2: 253-255.
11. *G. G. Tevzadze, J. V. Pierce and R. E. Esposito* (2003), Yeast, **20,** Suppl. 1, S62.
12. *G. G. Tevzadze, J. V. Pierce and R. E. Esposito* (2007), Genetics, **175,** 3: 1213–1227.
13. *G. G. Tevzadze, H. Swift and R. E. Esposito* (2000), Chromosoma, **109,** 1-2: 72-85.
14. *J. R. Huth, C. A. Bewley, B. M. Jackson, A. G. Hinnebusch, G. M. Clore and A. M. Groneborn* (1997), Prot. Sci., **6,** 11: 2359-2364.
15. *R. S. Sikorski and P. Hieter* (1989), Genetics, **122,** 1: 19-27.
16. *R. M. Horton, H. D. Hunt, S. N. Ho, J. K. Pullen and L. R. Pease* (1989), Gene, **77,** 1: 61-68.
17. *T. Maniatis, E. F. Fritsch and J. Sambrook* (1982), Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor).

18. *S. Klapholz and R. E. Esposito* (1982), Genetics, **100,** 3: 387-412.
19. *S. Klapholz, C. S. Waddell and R. E. Esposito* (1985), Genetics, **110,** 2: 187-216.
20. *C. S. Hoffman and F. Winston* (1987), Gene, **57,** 2-3: 267-272.
21. *R. D. Gietz and R. A. Woods* (1994), in *Molecular Genetics of Yeast: A Practical Approach*, J. R. Johnston (Oxford University Press Inc., New York, NY), 121-134.
22. *R. D. Gietz and R. H. Schiestl* (1995), Meth. Mol. Cell. Biol., **5,** 1: 255-269.
23. *T. Ano and M. Shoda* (1992), Biosci. Biotechnol. Biochem., **56,** 9: 1505.
24. *D. K. Bishop, D. Park, L. Xu and N. Kleckner* (1992), Cell, **69,** 3: 439-456.
25. *E. Alani, R. Padmore and N. Kleckner* (1990), Cell, **61,** 3: 419-436.
26. *C. Panozzo, M. Nawara, C. Suski, R. Kucharczyka, M. Skoneczny, A.-M. Bécam, J. Rytka and C. J. Herbert* (2002), FEBS Lett., **517,** 1-3: 97-102.
27. *P. Sung* (1997), Genes Dev., **11,** 9: 1111-1121.
28. *V. M. Glazer and A. V. Glazunov* (1997), Genetika, **33,** 9: 1221-1228.
29. *A. Schwacha and N. Kleckner* (1997), Cell, **90,** 6: 1123-1135.
30. *S. L. Gasior, A. K. Wong, Y. Kora, A. Shinohara and D. K. Bishop* (1998), Genes Dev., **12,** 14: 2208-2221.
31. *L. S. Symington* (2002), Microbiol. Mol. Biol. Rev., **66,** 4: 630-670.
32. *S. L. Hays, A. A. Firmenich and P. Berg* (1995), Proc. Natl. Acad. Sci. USA, **92,** 15: 6925-6929.
33. *G. S. Fortin and L. S. Symington* (2004), EMBO J., **21,** 12: 3160-3170.
34. *A. V. Glazunov and V. M. Glazer* (1999), Curr. Genet., **34,** 6: 430-437.
35. *T. Saeki, I. Machida and S. Nakai* (1980), Mutat. Res., **73,** 2: 251-265.
36. *S. A. Dames, J. M. Mulet, K. Rathgeb-Szabo, M. N. Hall and S. Grzesiek* (2005) J. Biol. Chem., **280,** 21: 20558-20564.

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Plant Physiology

Influence of Temperature on Morphological Indices of Soybean and Pea Seedlings

Eva Chkhubianishvili*, Nancy Kacharava*, Luiza Kobakhidze*

* *N. Ketskhoveli Institute of Botany, Tbilisi*

(Presented by Academy Member N. Nutsubidze)

ABSTRACT. **Changes in morphology of soybean (*Glycine hispida* (Moench) Max) and pea (*Pisum sativum* L.) seedlings under the influence of high and damaging temperatures of air (40°C for 1h, 2h, 3h, and 45°C for 1h and 2h respectively) have been evaluated. Heating at 40°C for 1-2h was found to stimulate the growth of seedlings, while exposure to the same temperature for 3h, or at 45°C for 1h caused an abating of the process. Prolonged (2h) exposure to 45°C resulted in the death of soybean seedlings. Pea seedlings appeared to be more resistant to increased temperature (45°C for 2h), but changes in root morphology – necrosis of the main root apex and retardation of growth and lateral root renovation – were noted. Among the soybean organs cotyledons proved the most resistant to overheating. Exposure to temperature led to a regular decrease of water content in seedlings, associated with the temperature level and duration of treatment. Increased temperature (45°C) for 1-2h caused the water content to diminish to about 50%. This served as the cause of growth inhibition.** © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *temperature, linear growth, water content, seedlings, soybean, pea.*

Plant's reaction to increased temperature depends on the extent and duration of the exposure to stress-factor. A number of investigations are dedicated to the study of the effect of overheating on the metabolism and productivity of plants [1-3].

The temperature range between 10°C to 25°C is regarded to be optimal for most plants, and 35°C is the upper level of temperature for plant metabolism. An increase of the air temperature to 40°-45°C leads to damage of the photosynthetic apparatus and finally to the plant's death [4]. Therefore, it is known that high temperature stimulates the synthesis of low molecular proteins called "heat shock proteins" [5, 6].

Data on the stimulative effect of increased temperature on growth and application of heat shock as a stimulator of plant tolerance to other stress factors are also presented in the literature [7].

Mainly full-grown individuals are used in studying the effect of stressors on plant, though in natural conditions seedlings are more sensitive to temperature.

Seeds of legumes – soybean (*Glycine hispida* (Moench) Max.) and pea (*Pisum sativum* L.) were previously soaked in water and placed for germination on a wet filter paper at 25°-27°C in darkness for two days. After germination the seedlings were exposed to white illumination for 24h at 50cm distance from the light source. 3 day old seedlings were exposed to increased temperatures of different duration (40°C for 1, 2, and 3h, and 45°C for 1 and 2h). 4 days after treatment morphological observation of seedlings was carried out. Also dry matter and water content of separate organs of seedlings were determined.

Experiments were conducted twice. Each variant consisted of 3-4 biological replications with 10 seedlings. The experiments were analogous to those conducted by T. A. Borisova on water-melon seedlings [7]. Statistical analysis of the data was carried out following the computer program MS-Excel. The diagrams demonstrate the water content of separate organs of soybean and pea in percentage. In the case of morphometric indi-

Table

Influence of different temperatures on growth of 7-day seedlings of soybean and pea

Plant	Seedling's organ	Index	Control	40°C for 1h	40°C for 2h	40°C for 3h	45°C for 1h	45°C for 2h
Soybean	Main root	Length, mm	43.75±4.23	124±10.5	144±14.1	81.75±4.3	37.5±3.67	Dies
		Biomass, mg	106±9.2	419±15.3	219±13.6	169±10.4	139±9.6	
	Lateral root	Maximal length, mm	48.25±29.4	48.75±4.31	50.77±3.06	49.77±5.0	35.62±2.17	Dies
	Hypocotyls	Length, mm	91.75±6.73	106.33±7.18	99.1±8.32	85.66±7.85	51.12±4.34	Dies
		Biomass, mg	273±8.1	282±9.6	297±6.8	228±4.9	159±7.6	
	Cotyledons	Length, mm	17.33±1.14	19.05±1.04	17.95±1.24	17.55±0.69	15.22±0.35	Dies
		Width, mm	9.35±0.94	10.33±0.74	10±0.86	9.6±0.59	8.55±0.72	
		Biomass, mg	759±15.6	433±11.3	416±8.8	350±10.6	308±9.9	
Pea	Main root	Length, mm	42.37±4.6	90.3±8.37	75.8±7.05	61.0±6.0	41.25±3.89	30.18±3.07
		Biomass, mg	98±8.6	230±10.9	198±8.6	196±6.3	81±6.9	51±4.3
	Lateral root	Maximal length, mm	21.88±2.88	37.8±3.56	26.8±2.74	20.8±2.36	16.16±1.5	9.0±1.0
	Hypocotyls	Length, mm	10.1±0.38	12.1±1.34	13.25±1.8	8.81±1.54	6.44±1.0	5.0±0.75
		Biomass, mg	30±3.2	59±4.4	40±3.6	23±1.5	20±1.5	20±2.0
	Epicotyls	Length, mm	20.62±2.03	47.3±4.56	37.7±3.63	27.0±2.26	25.8±2.94	14.8±1.3
		Biomass, mg	60±6.4	150±8.8	110±7.3	74±4.0	70±5.1	43±39

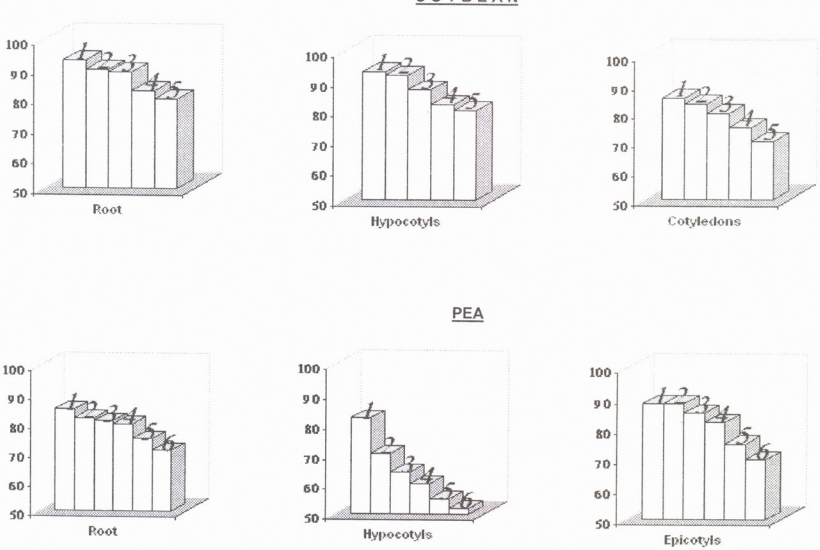


Fig. Influence of temperature on water content in organs of soybean and pea seedlings (%) 1) Control, 2) 40°C for 1h, 3) 40°C for 2h, 4) 40°C for 3h, 5) 45°C for 1h, 6) 45°C for 2h

ces the Tables contain mean arithmetical values and their standard deviations.

Experiments with short-term effect of high temperature showed that the growth reaction of seedlings differed. Exposure of 3-day seedlings of soybean and pea to 40°C for 1-2h stimulated the growth of all organs of the plants. The reaction to temperature stress depended on the duration of exposure. 3h exposure to 40°C and 1h to 45°C caused retardation of seedlings' growth (Table).

Soybean seedlings appeared to be more sensitive to increased temperatures: 2h exposure of 3-day seedlings to 45°C caused the death of most individuals on the 7th day. Necrosis of the main root apex and inhibition of the lateral roots growth, as well as damage of the cotyledons were noticed.

Visual observation first of all showed clearly the damage of the main root and its diminished growth. Cotyledons of the soybean seemed to be more resistant to overheating. This may be accounted for by the physiological peculiarities of cotyledons, determined by their functions: protection of the first leaves and provision of nutrients and energy till their autotrophic transformation [7].

In pea seedlings exposure to 40°C revealed its positive effect on the growth of all organs. Increasing the duration and level of temperature impact to 45°C for 2h caused regular diminishing of the growth intensity of the organs, but seedlings' resistance to stress turned out to be higher, compared with soybean. According to our data, the lethal temperature for soybean was 45°C at 2h exposure, while pea seedlings were more stable, although changes in root morphology, necrosis of the main root apex and inhibition of the lateral root growth were observed (Fig.).

Maintaining the stable level of water content is important for plant adaptation to unfavorable environmental conditions. Comparison of the water content of soybean and pea showed that pea seedlings contained less water than soybean did. Increasing the temperature and duration of the exposure led to the diminishing of water content. 45°C temperature for1-2h caused a reduction of the water content of seedlings by 50% and more in some organs (Fig.). This may be explained by the intensification of transpiration, connected with increasing temperature, causing a disbalance between water uptake by roots and its evaporation by hypocotyls in soybean, or epicotyls in pea, resulting in the growth inhibition of the experimental plants.

მცენარეთა ფიზიოლოგია

ტემპერატურის გავლენა სოიისა და ბარდის აღმონაცენების მორფოლოგიურ მახასიათებლებზე

ე. ჩხუბიანიშვილი*, ნ. კაჭარავა*, ლ. კობახიძე*

* *ნ. კეცხოველის ბოტანიკის ინსტიტუტი, თბილისი*

(წარმოდგენილია აკადემიკოს ნ. ნუცუბიძის მიერ)

შესწავლილია ჰაერის მომატებული ტემპერატურის (40°C 1, 2 და 3 სთ-ის განმავლობაში, 45°C 1 და 2სთ-ის განმავლობაში) გავლენა სოიისა (*Glycine hispida* (Moench) Max) და ბარდის (*Pisum sativum* L.) აღმონაცენების მორფოფიზიოლოგიურ მახასიათებლებზე. მომატებული (40°C) ტემპერატურის 1-2 საათით ზემოქმედება იწვევდა აღმონაცენების ორგანოთა ზრდის სტიმულაციას, ხოლო მაღალი ტემპერატურების ზემოქმედებას (3 საათით 40°C, ან 1 საათით 45°C) თან ახლდა სოიის აღმონაცენების დაღუპვა. ბარდის აღმონაცენები უფრო გამძლე აღმოჩნდა მაღალი (45°C 2სთ-ით) ტემპერატურისადმი, თუმცა აღინიშნებოდა ცვლილებები ფესვის მორფოლოგიაში: მთავარი ფესვის წვერის ნეკროზი, გვერდითი ფესვების ზრდისა და ადგენის შეფერხება. სოიის აღმონაცენის ორგანოებს შორის ლებნები ყველაზე გამძლე აღმოჩნდა ტემპერატურული შოკისადმი. ტემპერატურის ზემოქმედებამ შესწავლილ მცენარეთა აღმონაცენებში გამოიწვია

ტენის შემცველობის კანონზომიერი შემცირება, რაც ტემპერატურის სიდიდისა და ზემოქმედების ხანგრძლივობის მატებასთან ერთ დაკავშირებული. ტემპერატურის მომატება 45°C-მდე 1-2 საათით თითქმის 50%-ით, და ზოგ ორგანოში მეტადაც, ამცირებდა წყლის შემცველობას ალმონაცენებში, რამაც თავის მხრივ ზრდის ინჰიბირება გამოიწვია.

REFERENCES

1. *J. Levitt* (1972), Responses of Plants to Environmental Stress. NY. L., 697p.
2. *P. A. Genkel* (1982), Fiziologiya zhara i zasukhoustoychivosti rastenii. M., 280p (Russian).
3. *O. K. Atkin, M. G. Tjoelker* (2003), Trends Plant Sci., **8**, 343-351.
4. *I. S. Chauhan, T. Senbonu* (1996), J. Plant Physiol., **149**, 727-734.
5. *R. S. Boston, P. V. Viitanen, E. Veirling* (1996), Plant Mol. Biol., **32**, 191-222.
6. *N. Nedunchezhian, K. Annamalainathon, G. Kulandaivelu* (1992), Physiol. Plant., **85**, 503-506.
7. *T. A. Borisova, S. M. Bugarje, N. V. Meshkova, P. V. Vlasov* (2001), Fiziologiya Rastenii, **48**, 1, 589-595 (Russian).

Received December, 2006

Plant Physiology

Influence of Simulated Acid Rain on Nitrate Reductase Activity of Some Legumes and Vegetables

Medea Kikvidze*, Shota Chanishvili*

* *N. Ketskhoveli Institute of Botany, Tbilisi*

(Presented by Academy Member N. Nutsubidze)

ABSTRACT. Seeds (kidney bean and basil) and leaves (kidney bean, basil, beet and cabbage) of experimental plants were treated with pH2.5 water solution of sulphuric acid. The activity of nitrate reductase of treated plants, which usually reveal high activity of this enzyme, has been investigated. Spraying plants with acid solution increased the enzyme's activity. Since nitrate reductase is regarded as detoxifier of polluted air, it may be supposed that tested plants are resistant to acid pollution of the environment. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *acid rains, nitrate reductase, kidney bean, basil, beet, cabbage.*

Pollutant emissions of industry and transport, such as CO, NO_2, SO_2 etc., cause significant pollution of air on the territory of Georgia. These gases dissolve in rain water and return to the earth surface in the form of acidic precipitations [1, 2].

Recently the effect of acid rains on cultivated plants and woody plants has been investigated. The negative effect of metallurgical, chemical and other industries on vineyards, orchards and vegetables has been shown [3]. Acidic precipitations are responsible for disorders in plant vegetation, decreasing the assimilation area and growth, accumulation of toxic substances in roots, leaves and fruits. Increasing of titrable acidity and diminishing of dry matter accumulation has also been mentioned. Industrial gases cause inhibition of photosynthesis and photosynthetic productivity of plants, in particular, noncyclic phosphorylation and ATP synthesis, and related synthesis of CO sink are retarded, causing disbalance of the metabolic processes in plant [3].

The purpose of the given study was to investigate the activity of nitrate reductase, the main enzyme of plant nitrogen metabolism, in acid-treated cultivated plants. Two cultivars of kidney bean (*Phaseolus vulgaris* L.) with red and white seeds and vegetable cultures of beet (*Beta vulgaris* L.), red and white forms of cabbage (*Brassica capitata* L.) and red and green forms of basil (*Ocimum basilicum* L.) were taken for experiments.

In one series of experiments nitrate reductase activity was studied both in leaves of sprayed plants and those emerged from acid-treated seeds (kidney bean and basil).

Seeds and leaves of experimental plants were treated with pH2.5 water solution of H_2SO_4: seeds were soaked in acid solution for 24h, while leaves of tested plants were sprayed with acid three times with five days interval. Material for analysis was taken 10 days after the last spraying. The activity of nitrate reductase was studied after Mulder [4]. It was expressed in γ of NO_2 released during 30 min, per g of fresh material.

In the case of kidney bean and basil experimental material was picked in flowering and fruit-bearing phases, in the cases of beet and cabbage leaves for analysis were collected in the phase of intensive vegetation.

According to the obtained data it is clear that generally high activity of nitrate reductase was revealed in flowering phase (Table 1). Leaves of white-seed form of

Table 1

Nitrate reductase activity in leaves of kidney bean and basil (amount of NO_2 in γ per g of fresh weight, released in 30min)

Plant	Variant	Nitrate reductase activity, γ	
		Flowering phase	Fruit-bearing phase
Red seed kidney bean	Control	49.8	32.8
	Soaked seeds	87.8	38.5
	Sprayed leaves	98.8	62.2
White seed kidney bean	Control	98.4	68.8
	Soaked seeds	99.4	72.4
	Sprayed leaves	109.4	90.6
Blue leaf basil	Control	22.8	18.4
	Soaked seeds	14.4	10.2
	Sprayed leaves	16.2	12.4
Green leaf basil	Control	16.0	12.4
	Soaked seeds	12.6	12.0
	Sprayed leaves	14.0	8.4

Table 2

Nitrate reductase activity in leaves of beet and cabbage (amount of NO_2 in γ per g of fresh weight, released in 30min)

Plant	Variant	Nitrate reductase activity, γ
Beet	Control	73.4
	Sprayed leaves	98.0
Red cabbage	Control	31.7
	Sprayed leaves	64.2
White cabbage	Control	18.4
	Sprayed leaves	19.0

kidney bean were distinguished for especially high activity of the enzyme.

Soaking seeds and spraying leaves with acid solution increased the activity of nitrate reductase in both cultivars of kidney bean in the studied phases of vegetation, while in basil both acid treatments caused the diminishing of enzyme activity. It must also be mentioned that soaking of seeds caused more negative effect, compared to leaf spraying.

Spraying of beet and cabbage plants with acid rain caused essential stimulation of nitrate reductase activity of leaves. From the obtained results it is clear that nitrate reductase activity was the highest in beet (Table 2). Spraying of leaves significantly increased the enzyme's activity in beet and red form of cabbage, while in white form of cabbage small differences were revealed between the control and experimental variants.

According to the experimental results, high index of nitrate reductase activity was revealed in white and red sorts of kidney bean and beet. Treating seeds and leaves with acid solution increased the index. Since nitrate reductase is regarded as a detoxifier of polluted air [5], we may suppose that the studied plants are resistant to polluted environment.

მჟავე ნალექების გავლენა ნიტრატრედუქტაზას აქტივობაზე ზოგიერთი პარკოსანი და ბოსტნეული კულტურის ფოთლებში

მ. კიკვიძე*, შ. ჭანიშვილი*

* *ნ. კეცხოველის ბოტანიკის ინსტიტუტი, თბილისი*

(წარმოდგენილია აკადემიკოს ნ. ნუცუბიძის მიერ)

საცდელ მცენარეთა თესლი (ლობიო და რეჰანი) და ფოთლები (ლობიო, რეჰანი, ჭარხალი და კომბოსტო) დამუშავდა pH2.5 მჟავიანობის მქონე გოგირდმჟავას წყალხსნარით. დამუშავებულ მცენარეთა ფოთლებში შესწავლილ იქნა ფერმენტ ნიტრატრედუქტაზას აქტივობა, რომელიც ნორმალურ პირობებში მაღალია ამ მცენარეებში. მჟავათი დამუშავებამ გამოიწვია ფერმენტის აქტივობის გაზრდა. ვინაიდან ნიტრატრედუქტაზა დაბინძურებული ჰაერის ერთგვარ დეტოქსიკატორად ითვლება, შეგვიძლია ვივარაუდოთ, რომ შესწავლილი მცენარეები გარკვეულ გამძლეობას აქლენენ მჟავე ნალექებით დაბინძურებული გარემოს მიმართ.

REFERENCES

1. *M. Treshou* (1988), Zagryaznenie vozdukha i zhiz'n rasteniy. L. (Russian).
2. *G. Kontsik, M. Makarov, V. Kiseliova* (1998), Printsipy i metody otsenki ustoichivosti pochv k kislotnym vydeleniyam. M. (Russian).
3. *D. Dzugarian* (1988), Toksicheskoe deistvie tekhnogennogo zagryazneniya na agrotsenozy. "Ekotoksikologiya i okhrana prirody", Riga (Russian).
4. *S. Sergeichik, A. Sergeichik* (1988), Vliyanie gazoobraznykh toksikantov na aktivnost' peroksidazy i nitratreduktazy list'ev drevesnykh rastenii. Thesis of training, Riga (Russian).
5. *E. L. Mulder, R. Van Veen Boxma* (1959), Plant and Soil, **10**, 4.

Received November, 2006

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Plant Growing

Effect of Different Growth Media on *in vitro* Propagation of Grapevine Cultivar "Chkhaveri"

Nargiza Alasania*, Sophia Manjgaladze*, Nino Lomtatidze*, Sophia Tskvitinidze*

* *Batumi Botanical Garden*

(Presented by Academy Member V. Papunidze)

ABSTRACT. Artificial Gamborg (B5) and Murashige-Skoog (MS) nutrient media have been exploited for *in vitro* propagation of endemic grapevine cultivar "Chkhaveri". Cultivation media were added by phytohormones benzylaminopurine (BAP, group of cytokinins), and indole acetic acid (IAA, group of auxins). The Murashige-Skoog (MS) cultivation medium added by 8 μM BAP solution turned out to be the optimum for propagation of grapevine explants. Buds grew rapidly on the medium, the developed plants were of light green color and at the end of subcultivation shoot height reached 20-25 mm, average number of developed shoots was equal to 5-6 units. Adding auxin to the cultivation medium was ineffective. Auxin negatively influenced the cultivation, process probably due to the content of endogenous auxin in the explant. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *explant, micro shoots, phytohormones, benzylaminopurine (BAP), naphthyl acetic acid (NAA), indole acetic acid (IAA).*

Viticulture has a long and interesting history in Ajara region. Best of all wine-growing is developed in Keda District of mountainous Ajara [1].

The aim of our experiments was to establish optimum conditions for rapid mass propagation of the oldest endemic grapevine cultivar "Chkhaveri" and to choose concentrations of phytohormones which promote fast growth and propagation of buds.

Micro shoots developed in the first stage of the experiment have been used as explants. Explants were cultivated on the modified Gamborg (B5) and Murashige-Skoog (MS) artificial nutrient media (pH 5.8-5.9) [2, 3]. The following growth regulators: benzylaminopurine (BAP), naphthyl acetic acid (NAA) or indole acetic acid (IAA) were added to the media at different concentrations. Subcultivation of explants was carried out every 25 days. Explants were incubated in phytotron at 26°C temperature, at 16/8 h photoperiod, at 2-3 Klux illumination. Evaluations were carried out at the end of each subcultivation [3-5].

Mineral composition of cultivation media, concentration and proportion of phytohormones differently affected propagation coefficient. The following ratios of BAP/IAA have been used in the experiment:

8μM/ 1μM; 8μM/0.5μM; 8μM/-
14μM/1μM; 14μM/0.5μM; 14μM/-
20μM/1μM; 20μM/0.5μM; 20μM/-

None of cytokinin (BAP) concentrations added to Gamborg medium gave desirable effect. Only emergence of negligible amount of adventitious buds has been attested. In particular, development of buds *de novo* and growth in height of micro shoots proceeded slowly on the Gamborg medium (B5) added by BAP at 8mM concentration. During the further cultivation plant leaves became yellow and their abscission took place.

Table

Effect of phytohormones on *in vitro* propagation of grapevine cultivar "Chkhaveri"

Phytohormone concentration, μM		Number of adventitious buds	Average height of shoots, mm
BAP	IAA		
8	-	5.2	20-25
14	-	5.0	20-25
8	0.5	2.3	12
14	0.5	2.0	10
8	1	1.7	10
14	1	1.5	10

Adding BAP solution of 10mM concentration to Gamborg medium (B5) caused significant delay of propagation and growth processes. Condition of the cultures worsened, shoots became brittle and perished. Further elevation of BAP concentration yielded a similar result. It can be concluded that the components of Gamborg medium ineffectively influenced the morphogenetic potential of grapevine plant. This was manifested as repressive effect of phytohormones as well. Contrary to this, addition of phytohormones to the modified nutrient medium of Murashige-Skoog (MS) positively affected microclonal propagation of grapevine explants.

The results of experiments have shown that substitution of one nitrogen source KNO_3 with another one NH_4NO_3 in grapevine *in vitro* culture acted as inductor of initiation and realization of morphogenetic potential of the plant.

Murashige-Skoog (MS) medium added by 8M BAP solution turned out to be the optimum for grapevine cultivation. On the 3rd-4th days after subcultivation development of new buds at the basal part of the explant started. The buds grew rapidly, the developed plants were of light green color and at the end of subcultivation shoots reached 20-25 mm in height, average number of shoots was 5-6 units.

To reach the higher propagation coefficient BAP concentration was increased up to 14mM, but we failed to obtain a significantly different result. Elevation of BAP concentration up to 20mM caused shortening of apical shoot, its thickening and increase of leaf size, probably due to surplus concentration of cytokinin (Table).

Thus, summarizing points of growth and development of grapevine explants 8mM concentration of BAP can be regarded as an optimum one. As regards auxins, addition of IAA to the cultivation media negatively affected culture growth probably due to the content of endogenous auxins in the explant tissues [2, 6, 7]. So, synthesis of auxins took place in tissues of the explant. Halving of auxin concentration (0.5mM), caused strong retardation of processes of culture growth and development. Thus the use of auxin, namely IAA, for microclonal propagation of explants of grapevine cultivar "Chkhaveri" has been found to be ineffective.

მემცენარეობა

საკვები არის გავლენა ვაზის ჯიშის "ჩხავერის" გამრავლებაზე *in vitro* სისტემაში

ნ. ალასანია*, ს. მანჯგალაძე*, ნ. ლომთათიძე*, ს. ცქვიტინიძე*

* *ბათუმის ბოტანიკური ბაღი*

(წარმოდგენილია აკადემიკოს ვ. პაპუნიძის მიერ)

ვაზის ენდემური ჯიშის "ჩხავერის" კულტურაში გამრავლებისათვის გამოყენებულ იქნა გამბორგის (B_5) და მურასიგე-სკუგის (MS) ხელოვნური საკვები არეები. საკვებ არეებს ემატებოდა ფიტოჰორმონები ბენზილამინოპურინი (ბაპ) და ინდოლმძარმჟავა (იმმ). *in vitro* ვაზის ექსპლანტთა გამრავლებისთვის

ოპტიმალური აღმოჩნდა MS საკვები არე ბაპ-ის 8 მკმ კონცენტრაციით. ამ არეზე კვირტები სწრაფად იზრდებოდა, ვარჯი იყო მთლიანად ღია მწვანე შეფერილობის და სუბკულტივირების ბოლოს ყლორტების სიმაღლე 20-25 მმ-ს აღწევდა, ხოლო საშუალო რაოდენობა 5-6 ერთეული იყო. საკვებ არეში აუქსინების დამატება არაეფექტური აღმოჩნდა. ამ ფიტოჰორმონის დამატება უარყოფითად მოქმედებდა კულტურაზე, რაც საგარაუდოდ გამოწვეული იყო ექსპლანტში ენდოგენური აუქსინის შემცველობით.

REFERENCES

1. *N. Gelashvili.* (1961), Wine-growing, part I, Tbilisi, 415 (Georgian).
2. *B. A. Dospekhov.* (1985), Metodika polevogo opyta, M., 35 (Russian).
3. *J. Murashige, F. Skoog.* (1962), Physiolog. Plant., 87.
4. *O. Z. Gamborg et al.* Exp. Cell. Res., 1968, 88.
5. *Y. Sun, S. Jin, S.Liang , Y. Nie.* (2004), Plant Cell, Tissue and Organ Culture, **76**, 2, 190-197.
6. *K. A. Schestibarov, R. V. Mikhailov, S. V. Dolgov.* (2003), Plant Cell, Tissue and Organ Culture, **72,** 2, 140-145.
7. *Tang Wei, L. Harris, V. Quthavon.* (2004), Plant Cell, Tissue and Organ Culture, **75,** 2, 250-260.

Received December, 2006

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Agricultural Sciences

Soil Sciences

Heavy Metal Pollution of Soils and Food Crops due to Mining Wastes in the Mashavera River Valley

Peter Felix-Henningsen[*], Tengiz F. Urushadze[**], Eliso I. Narimanidze[£], Lars-Christoffer Wichmann[£], Dietrich Steffens[§], Besik B. Kalandadze[¥]

* *Institute of Soil Science and Soil Conservation, Justus-Liebig-University, Giessen, Germany*
** *Academy Member, Georgian State Agricultural University, Tbilisi*
£ *Centre for International Development and Environmental Research, Giessen, Germany*
§ *Institute of Plant Nutrition, Justus-Liebig-University, Giessen, Germany*
¥ *Iv. Javakhishvili Tbilisi State University*

ABSTRACT. The fertile irrigated soils of the Mashavera valley, Georgia, have high agricultural yield potential. The river water used for irrigation, however, is polluted with mining waste from a copper and gold mine situated in the mountainous region of the middle reaches of the Mashavera river. Furthermore waste water from a floatation plant, erosion material from floatation waste deposits and acid mine drainage leads to high concentrations of dissolved and suspended sulphidic heavy metals. The Cu, Zn and Cd concentrations of mud from irrigation channels and the Mashavera river are extremely high. It is estimated that the annual transfer of heavy metals (HM) by irrigation water is in the range of several g*ha^{-1} for Cd and several kg*ha^{-1} for Cu and Zn. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *heavy metals, pollution of soils, irrigated and non-irrigated soils.*

Georgia is rich in different mineral resources, which were already known and mined over the past centuries. One of the most important and biggest non-ferrous metal deposits is located in SE Georgia in the province Kvemo Kartli at the mountain fringe of the Small Caucasus. In the region of the small town Bolnisi, located in the transition area between the middle and lower reaches of the Mashavera river, poly-metallic ore deposits, mainly copper, zinc and gold, have been mined since 1974 in a large opencast mine at the village of Kazreti.

Another branch of economic activity is intensive agriculture on the fertile soils of the alluvial and terrace plains of the Mashavera river. Due to the semi-arid climate of East Georgia, with arid phases during the vegetation period, sufficient yields of fruits and food crops from the fertile soils, mainly kastanozems and chernozems, can only be obtained by irrigation with water from the Mashavera. Then 2 to 3 yields per year are possible.

Decades of mining of copper and precious metals have caused severe environmental problems. Mining waste, which is deposited on the mountain slopes around the opencast mine as well as deposits of waste from a flotation plant consisting of fine ground rock debris, cover an area of about 240 ha. Mining and flotation wastes are rich in fines, which contain remains of sulphides. Due to an extremely acid environment from sulfuric acid, caused by oxidation of sulphides near recent surface deposits and the release of heavy metals (HM), the stockpiles are unvegetated. Adequate measures to stabilize the slopes by planting vegetation were never taken. Therefore, runoff from the slopes leads to rill and gully erosion and severe contamination of the Kazretula river and other

small creeks near Kazreti, which meet the Mashavera river, by fines with adsorbed heavy metals and remnants of sulphides. The Mashavera presently is one of the most polluted rivers of Georgia.

The use of the Mashavera waters for irrigation led to pollution of the fertile soils by suspended fines, rich in heavy metals. The polluted soils, on the other hand, can also be a potential source for the contamination of the food chain by heavy metals.

Accordingly a research project, generously funded by the German Volkswagen Foundation, focused on amounts and spatial distribution of the heavy metals as a consequence of deposition of mining wastes on irrigated soils of the Mashavera valley as well as on the eco-toxicological importance for the food chain.

The study area is situated in SE Georgia, about 60 km SW of the capital Tbilisi, in the administrative district Bolnisi, and contains the middle and lower reaches of the Mashavera valley.

The ore deposits of the Bolnisi region show a close relationship to volcano-sedimentary rocks of the Upper Cretaceous (Upper Turonian up to Lower Santonian), locally penetrated by intrusive rocks.

The most important ore deposits of Georgia occur in the Bolnisi district, especially deposits of non-ferrous metals. They show a close relationship to hydrothermal systems that occur in the surroundings of Santonian volcanic dykes, tuff deposits and tectonic faults [1,2].

The hydrothermal fluids caused a diffuse, stratiform dissemination of metals in the superficial rocks as well as a crystallisation of sulphidic ore veins. The metal sulphides are associated with gold, which mainly occurs in metasomatic alterated banks of quartzite-sericite rocks of about 100 m thickness [2,3]. Mining of the copper sulphide deposits started in 1975. The "Madneuli" mining company produces high-grade copper concentrates by floatation, which are exported to different countries [4]. The extraction of gold from ground quartzite rocks by cyanide (heap-leach procedure) started in 1994. The open mine pit is situated 10 km SW of Bolnisi village and covers an area of about 350 ha in the summit area of a mountain range between the Mashavera valley in the west and the Poladauri valley in the east. Therefore surface and subsurface waters from the mine and the mine spoils, which cover the slopes adjacent to the pit, form acid mine drainage (AMD) with pH values < 3 and high dissolved HM concentrations. AMD is collected in sedimentation basins by a channel and pump systems, but a lack of financial and technical possibilities in the past prevented further processing of the AMD by neutralization with lime and extraction of the dissolved copper by binding agents. In addition, highly contaminated waste water from the Madneuli floatation plant contributes to the pollution of the river Mashavera [2,3]. Violations against existing security regulations caused a temporary closure of mining in the past.

Fine grained floatation wastes are deposited on two heaps situated about 2-3.5 km NW of the open mine pit, covering a total area of 240 ha. The sulphidic rock material is subject to intensive biochemical oxidation, which causes release of large amounts of AMD and dissolved HM. The steep slopes of the large heaps are bare and not fixed by vegetation, thus causing sheet and rill erosion and transport of fines to the Mashavera and Kazretula rivers, while AMD runoff directly contaminates the rivers by HM. In addition, the physical state of the deposits was not taken into account and resulted in slope failures [5]. During the dry summer months the heaps of floatation waste are the source of dust, which amounts to about 30 Mg a^{-1} [4] and pollutes large areas in the surroundings with HM. Thus the non-irrigated arable soils and uncultivated soils of the mountain regions also display relatively high "natural" background concentrations of HM.

The natural steppe vegetation formations [6] of the Mashavera valley were changed by agriculture.

This main channel branches off from the Mashavera river directly below the floatation plant and the large deposits of floatation waste at Kazreti. Due to the strong load of the Mashavera river water with fines, which derive from the waste deposits and the floatation plant, the irrigation channels also conduct water heavily contaminated by HM.

The fields of the northern side of the Mashavera valley are irrigated with water from a channel, which branches off from the Mashavera river upstream of Kazreti. Therefore this water is free from contamination.

Due to the continental type of climate the typical main soil orders belong to kastanozems and chernozems [7,8], which frequently degraded to phaeozems as a consequence of long periods of intensive irrigation. The main soils of the study area are calcic kastanozems, which show an accumulation of calcium carbonate in the form of concretionary mottles in the subsoil but lower humus content and a more brownish colour than chernozem. They are associated with calci-vertic chernozems with clay contents up to 65 mass-%. The medium to strong humic topsoils show a blocky to polyhedral structure due to the high clay content. They are weakly alkaline and have a rather high CEC.

The humus content decreases with increasing intensity of soil cultivation in the sequence grape fields and vineyards, orchards, house gardens, and arable fields. As a further reason for the high humus contents in topsoils of vineyards, a lower mineralization rate can be assumed resulting from an accumulation of Cu due to strong irrigation (see below) as well as from the use of copper bearing fungicides [9]. Arable fields, on the other hand, are subject to soil erosion during periods of rotation fallow, which diminishes the humus content in slope positions and leads to the formation of colluvium in depressions and on the flat valley floor. On slopes with severe erosion, the calcic horizon is exposed at the surface and causes the formation of calcisols.

The clay minerals of chernozems and kastanozems mainly consist of chlorite-smectite mixed-layer minerals

(corrensite) and the composition shows no difference between the saprolite from pyroclastics (rhyolitic tuff and ignimbrite) and the soil horizons above, although the clay contents rise from 20 mass-% in the saprolite to about 60 mass-% in the topsoil.

Orientation in the field was based on Russian topographical military maps 1:50,000, Landsat TM 5 scenes and aerial photographs.

Soil samples were taken in fields, house gardens, grape fields, vineyards and orchards from the Ap horizon (0–30 cm) at 10 sites along double-diagonal transects. Eight volume-equivalent cores were taken with an aluminium auger within an area of 4 x 4 m at each site. The fine earth (< 2mm) of the air dried samples, ground in a porcelain mortar, was investigated in the laboratories of the Institute of Soil Science and Soil Conservation of Justus-Liebig University in Giessen, Germany.

The *pH* was determined after DIN 10390 in suspension with 0.01 M $CaCl_2$ with a pH-meter pH90 (WTW).

The amount of *carbonates* was determined by the gas-volumetric method using a calcimeter, following DIN 18129.

Total amounts of carbon (C_t) *and nitrogen* (N_t) were determined on fine ground samples by gas-chromatography using a C-N-S element analyzer (Heraeus). *Anorganic* C was calculated from the carbonate content by using the factor 0.1199, while the amounts of *organic carbon* (C_{org}) resulted from the difference between C_t and anorganic carbon. The amounts of *organic matter* were calculated by $C_{org} \cdot 1.724$.

Particle size distribution was determined by the combined sieving (and fractions) and pipette method (silt and clay) after decomposition of carbonates (HCl) and organic matter (H_2O_2) and dispersion in Na-Pyrophosphate, following DIN 19683.

Amorphous iron (Fe_{ox}) *and manganese oxides* (Mn_{ox}) were determined by extraction with buffered oxalic acid, pH 3.25, under dark conditions as described in [10].

Pedogenic iron (Fe_{DCB}) *and manganese* (Mn_{DCB}) oxides were extracted following the procedure of Mehra & Jackson, as described in [10].

The *mobile and exchangeable fractions of HM*, which are eco-toxicologically relevant because they are potentially plant available and easily leachable, were extracted with NH_4NO_3 according to [11,12]. They are designated in the text as HM_{AN}.

The *total amounts of subsequent deliverable HM*, which is the *supply fraction* and includes the soluble and exchangeable fractions as well as the HM strongly adsorbed to carbonates, oxides and organic substances, were extracted by EDTA. Deviating from the method described by [13], EDTA was dissolved in a buffered solution of ammonium acetate at pH 7. Elements of this fraction are designated in the text as HM_{EDTA}.

The *total amounts of HM* were extracted from finely ground samples by using Aqua Regia following DIN ISO 11466. Elements of this fraction are designated in the text as HM_{AR}.

Element concentrations in the extracts were determined with the atomic adsorption spectrometer FAAS 4100 (Perkin Elmer). For determination of Cd in the NH_4NO_3 extracts a GFAAS SIMAA 6000 spectrometer (Perkin Elmer) was used due to the low concentrations.

Due to the discharge of waste waters from the floatation plant, as well as suspensions and acid mine drainage (AMD) from the mining waste deposits, the water and sediments of the Mashavera and Kazretula rivers and the branching irrigation channels are loaded by HM. pH values < 4 of the flowing waters indicate that HM are temporarily dissolved. The concentration of the river and channel waters greatly exceeds the threshold values of the German Waste Water Regulation and the German Drinking Water Regulation, although a dilution occurs when the mine waters enter the river. The population of the villages along the rivers and channels use the water not only for irrigation but also for washing of fruits and vegetables, clothes and dishes as well as for personal hygiene. Therefore the residents are exposed to a high risk of direct contamination with HM.

Liming of the Kazretula river, sporadically disposed by the Madneuli Mining Company, raised the pH to a weakly alkaline range and diminished the concentration of dissolved HM

Fine-grained sediments of the rivers and the irrigation channels are contaminated to a high degree with HM. They consist of fine ground sulphidic rock fragments and organic matter from bacteria and algae, which absorb a part of the dissolved HM fraction. According to the German Sludge Ordinance [14], the concentrations exceed the threshold values for Cu_{AR} by factors of 10-40, for Zn_{AR} by 3 and for Cd_{AR} by 1.2-3.4. A proportion of 30 (Cu, Zn) to 60 % (Cd) of the total amounts belong to the supply fraction, which is potentially plant available .

In the course of irrigation, suspended fines are distributed with the irrigation water on the fields and gardens. Under the assumption that the irrigation water contains a concentration of suspended fines of 1 g l^{-1}, an irrigation of only 50 mm a^{-1} causes an annual deposition of 500 kg mud ha^{-1} with up to 16 kg Cu ha^{-1}, 3.6 kg Zn ha^{-1} and several 100 g Cd ha^{-1}.

At the end of the annual irrigation period the irrigation channels are cleaned from mud sediments. The mud is excavated and deposited at the field margins along the irrigation channels. Then erosion distributes the fines over the adjacent field areas and causes another source of soil pollution by HM.

Screening investigations of the concentrations of diverse heavy metals in topsoils of the Mashavera valley showed that only Cu, Zn and Cd are problem elements, which accumulated in the course of irrigation and dust deposition, originating from the deposits of mining waste in the Kazreti mountainous area. Other heavy metal

species such as Pb, Cr, V, Co, Ni and Hg occur in negligible low concentrations. The spatial distribution of the problem elements Cu, Zn and Cd shows a soil loading, which depends on the kind of land use. A significant increase of the concentrations follows the sequence: Reference soils (un-irrigated) << arable fields < house gardens << orchards and vineyards. Therefore the presentation and discussion of the results must take the type of land use into account (Fig. 1).

Non-irrigated soils of arable fields used for growing cereals and maize are situated on slopes above the irrigation channels. The concentrations comply with the background values in the study area and are related with the parent material and eolian dust deposition from the mining waste deposits [4]. With respect to the precaution values [15], the HM concentrations in the non-irrigated topsoils are in the uppermost range. For about 30 % of the investigated sites the soil concentration of Cu_{AR} already exceeds the precaution value of 60 mg*kg-1, while the maximum value of the mobile Cu_{AN} fraction lies with 0.31 mg*kg^{-1} clearly below the trigger value for the pathway soil–food crop. The Zn_{AR} and Cd_{AR} concentrations fall below all threshold values and therefore pose no concern at present for the non-irrigated soils.

Most of the irrigated soils under different land use display a strong enrichment of HM that can be traced back to irrigation with polluted water from the Mashavera river over a period of several decades (Fig. 1). The concentrations of Cu_{AR}, Zn_{AR} and Cd_{AR} are narrowly and significantly correlated (r = 0.94–0.96) due to the same source. The span of the HM concentrations in irrigated soils is by far wider than in the non-irrigated topsoils, especially towards higher concentrations above the 75th percentile. Differences in the duration, frequency and amounts of irrigation as well as changes in the type of land use and soil cultivation (e.g. depth of plowing) are the reasons for this high spatial variability. This shows that the degree of soil pollution by HM in the Mashavera valley cannot be estimated but must be investigated for each cultivated field, house garden and orchard in order to evaluate the hazard to the food chain and the population.

In topsoils of arable fields the Cu_{AR} concentrations considerably exceed the regional background values (by a factor of 5), the German precaution values (factor 2–7) and the target values of the Dutch List (factor 2–9). The low concentrations of the mobile fraction, < 1 % of the potentially available supply fraction (HM_{EDTA}), indicates the high sorption capacity of the soils rich in humus, bases and clay for Cu. Nevertheless, at two sites the trigger value is exceeded for the path soil–food crop according to [16].

In 77 % of the investigated sites the Zn_{AR} concentrations exceeded the precaution values [16] and the target values (Dutch List) up to a factor of 3. Due to an obviously strong adsorption of Zn, from which the rather low concentrations of the available fraction (Zn_{AN}) is a result, the trigger value for the path soil–food plant of more than 2 mg*kg^{-1} was reached at one site. Therefore it can be assumed that up to now the growth of food crops is not severely affected by the Zn concentrations.

In 27 % of the investigated topsoils of arable fields, the concentrations of Cd_{AR}, with amounts up to 5.7 mg·kg^{-1} exceeded the precaution value of 1.5 mg*kg^{-1} according to [16] for soils rich in clay. The action value for topsoils of fields used for the cultivation of bread wheat is defined as 0.04 mg Cd_{AN}*kg^{-1}. In three fields the action value was reached or exceeded, but in most sites 50 % or more of the range below the action value was reached. Therefore the growth of bread wheat on irrigated fields of the

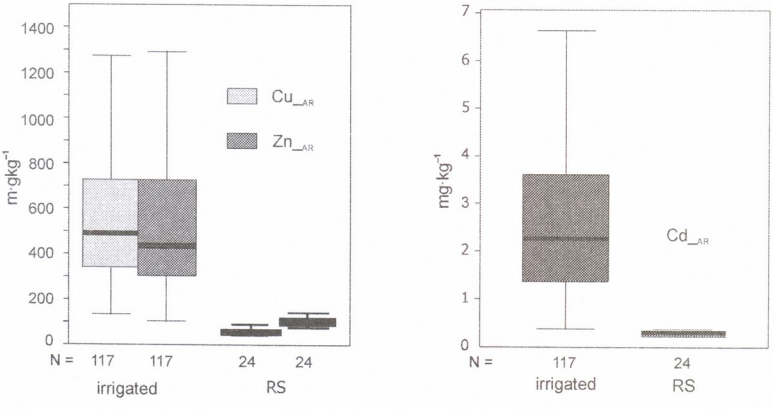


Fig. 1. Total amounts (box plots with medians, 25 and 75 percentiles, minimum and maximum) of heavy metals in topsoils of non-irrigated arable fields (RS = reference soils) in comparison to topsoils of irrigated sites of different land use.

Mashavera valley should be controlled with respect to the Cd pollution of the soils.

The medians of the EDTA fractions amount to 9 (Cu), 13 (Zn) and 6 (Cd) times greater than in the non-irrigated reference soils. About 33 % (Cu), 12 % (Zn) and 67 % (Cd) of the total amounts belong to this supply fraction, which indicates, especially in the case of Cd, a high potential risk of soil pollution with heavy metals.

As Cd is the most mobile heavy metal, an increasing uptake by plants already occurs under slightly acid conditions below pH 6.5. At present the arable soils also contain small amounts of calcium carbonate in the A-horizon, mostly in the range of 2–5 mass%. Because agriculture and irrigation favors decalcification and acidification, an increase of the Cd mobility must be taken into account in the future.

The alluvial floodplains along the Mashavera river are covered by permanent grassland, which becomes occasionally submerged by flood waters of the river. Due to the pollution of the river water by mining wastes, settled fines with a high concentration of HM are distributed on the soil surface and the leaves of vegetation. The meadows are used by extensive grazing of cattle, sheep herds and horses, which also use the river water for drinking. Although the span of the HM concentrations is lower, the medians are in the same range compared to topsoils of arable fields. This shows that due to flooding, the spatial distribution of the HM is more uniform. Due to a sandy texture and partly low pH values, humus and carbonate contents, the alluvial soils display a higher proportion of HM_{AN} in the mobile fractions as compared to irrigated arable soils. The lower sorption capacity of soils, mostly rich in sand, results in the precaution values being lower than for clayey soils [16]. They are 0.4 mg*kg^{-1} for Cd_{AR}, 20 mg*kg^{-1} for Cu_{AR} and 60 mg*kg^{-1} for Zn_{AR}. Eighty per cent of the samples from A-horizons of meadows exceed the precaution value for Cd_{AR}, and 100 % of the samples for Zn_{AR} und Cu_{AR}. On the other hand the action values (for grassland in the AR fraction, [16]) are relatively high (Cu_{AR}: 1,300 mg*kg^{-1}, Cd_{AR}: 20 mg*kg^{-1}, Zn_{AR}: not defined) and not reached by the actual HM_{AR} concentrations, with one exception: for meadows used for sheep grazing, which is frequently the case in the Mashavera valley, the action value for Cu_{AR} is 200 mg*kg^{-1}. More than 60 % of the samples from permanent meadows of the valley floor exceed this action value. This means that the pollution of the meadow soils can be dangerous for sheep, but also other grazing animals are affected through direct uptake of HM with fines adhering to the leaves of the grass vegetation and dispersed in the river water used for drinking. Therefore the meadows of the alluvial plain of the Mashavera river bear a high risk for the transfer of HM into the nutrient chain.

The vegetable gardens are distributed on the lower terrace of the Mashavera valley, adjacent to villages and farms, where they are irrigated manually with water from the river. Other vegetable gardens adjoin the irrigation channels, where they concentrate around settlements. Compared to the arable land, vegetable gardens are irrigated over a longer period in the year, more intensively and with a greater amount of water brought by buckets and pumps or distributed from irrigation channels by furrow irrigation. Therefore the concentrations of the total amounts of HM_{AR} are much higher than in topsoils of arable fields. The same trend is shown by the supply fraction (EDTA). Forty per cent of Cu, 20 % of Zn and 79 % of Cd (Table 1) belong to the supply fraction. This indicates that with increasing transfer of HM with the irrigation water, the supply fraction increased non-proportionally.

At nearly all sites the precaution values are exceeded according to [16] and the target values of the "Dutch List". The median of the soluble Cu_{AN} fraction exceeds the trigger value of 1 mg*kg^{-1} for the pathway soil–food plant. For 10 % of the investigated house gardens, extreme concentrations up to eight times the trigger value were found. Due to toxic effects, frequent growth depressions of the cultivated crop plants must be taken into account.

The Zn_{AR} concentrations in topsoils of house gardens exceed the precaution value of 200 mg*kg^{-1} in nearly all samples and surpasses this value by up to a factor of 12. Fifteen per cent of the samples exceed the trigger values of 2 mg Zn_{AN}*kg^{-1}. Therefore Zn may also cause frequent reductions in growth, yield and quality of the garden crops.

Due to high Cd concentrations, 85 % of the samples exceed the precaution value of 1.5 mg Cd_{AR}*kg^{-1}. Because in vegetable gardens crops with a high affinity for HM accumulation are frequently grown, the action value of 0.1 mg Cd_{AN}*kg^{-1} was reduced to 0.04 mg Cd_{AN}*kg^{-1} according to [16]. In the case of growing non-accumulating species of vegetables, 21 % of the investigated samples from vegetable gardens exceed the action value of 0.1 mg Cd_{AN}*kg^{-1}. Since there is mainly a mixed cropping of accumulating and non-accumulating vegetable species, the lower action value has to be taken into account, which is exceeded in about 40 % of the samples. This means that due to the strong toxicity of Cd and a potentially high transfer into the food chain, further investigations are necessary of all irrigated vegetable gardens in the Mashavera valley. According to [15], restrictions of land use would be necessary in order to protect the population.

An extreme HM load is distributed in topsoils of orchards, grape fields and vineyards (Table 2). Since the 19th century cultivation of fruit trees and wine has a long tradition in the Mashavera valley and its tributaries. These permanent cultures cover most of the area of the Mashavera valley bottom and the lower slopes of the adjacent mountains. Traditional mixed cropping with vegetables between the grape rows requires a high intensity of irrigation and therefore causes a severe HM contamination of the soils. This is clearly indicated by the comparison of the HM concentrations in topsoils of two neighboring, irrigated vineyards, with and without mixed cropping.

Table 1

Concentrations of heavy metal fractions in topsoils of irrigated vegetable gardens of the Mashavera valley and related soil data. *AR* = aqua regia extract, *EDTA* = EDTA extract, *AN* = NH_4NO_3 extract, OM = organic matter; Al/Fe/Mn_{DCB} = free oxides, Al/Fe/Mn_{ox} = amorphous fraction of the free oxides.

		Vegetable Gardens. Irrigated (N = 67)		
		Minimum	Maximum	Median
Cu_{AR}	(mg*kg^{-1})	91.20	2,945.00	553.25
Cu_{EDTA}	(mg*kg^{-1})	15.80	1371.00	227.05
$Cu_{EDTA/AR}$	%	17.00	55.00	39.50
Cu_{AN}	(mg*kg^{-1})	0.27	7.92	1.44
$Cu_{AN/EDTA}$	%	0.00	2.00	1.00
Zn_{AR}	(mg*kg^{-1})	157.70	2,441.00	501.50
Zn_{EDTA}	(mg*kg^{-1})	18.10	585.00	94.45
$Zn_{EDTA/AR}$	%	10.00	60.00	20.00
Zn_{AN}	(mg*kg^{-1})	0.00	31.83	0.33
$Zn_{AN/EDTA}$	%	0.00	21.00	0.00
Cd_{AR}	(mg*kg^{-1})	0.28	14.50	2.55
Cd_{EDTA}	(mg*kg^{-1})	0.13	11.16	2.00
$Cd_{EDTA/AR}$	%	46.00	95.00	79.00
Cd_{AN}	(mg*kg^{-1})	0.00	0.56	0.03
$Cd_{AN/EDTA}$	%	0.00	14.00	1.00
pH (H_2O)		6.67	8.23	7.76
pH ($CaCl_2$)		5.99	7.68	7.37
EC 2.5	(µS*cm^{-1})	83.00	1,297.00	201.50
$CaCO_3$	mass-%	0.00	15.50	1.70
OM	mass-%	0.28	6.50	3.16
Al_{DCB}	mg*g^{-1}	0.50	1.78	1.07
Fe_{DCB}	mg*g^{-1}	5.46	15.30	8.83
Mn_{DCB}	mg*g^{-1}	0.32	1.06	0.67
Al_{ox}	mg*g^{-1}	0.67	3.14	1.82
Fe_{ox}	mg*g^{-1}	0.79	6.78	1.88
Mn_{ox}	mg*g^{-1}	0.04	1.00	0.63

Table 2

Concentrations of heavy metal fractions in topsoils of irrigated orchards and wine gardens of the Mashavera valley and related soil data. *AR* = aqua regia extract, *EDTA* = EDTA extract, *AN* = NH_4NO_3 extract, OM = organic matter; Al/Fe/Mn_{DCB} = free oxides, Al/Fe/Mn_{ox} = amorphous fraction of the free oxides.

		Orchards. Wine Gardens. Irrigated (N = 49)		
		Minimum	Maximum	Median
Cu_{AR}	(mg*kg^{-1})	127.30	2,366.00	605.40
Cu_{EDTA}	(mg*kg^{-1})	56.90	1,006.00	280.10
$Cu_{EDTA/AR}$	%	28.00	79.00	45.00
Cu_{AN}	(mg*kg^{-1})	0.23	6.59	1.22
$Cu_{AN/EDTA}$	%	0.00	1.00	0.00
Zn_{AR}	(mg*kg^{-1})	139.30	3,735.00	614.00
Zn_{EDTA}	(mg*kg^{-1})	8.30	842.00	134.60
$Zn_{EDTA/AR}$	%	4.00	56.00	22.00
Zn_{AN}	(mg*kg^{-1})	0.00	19.30	1.15
$Zn_{AN/EDTA}$	%	0.00	7.00	1.00
Cd_{AR}	(mg*kg^{-1})	0.40	13.80	3.37
Cd_{EDTA}	(mg*kg^{-1})	0.17	7.87	2.76
$Cd_{EDTA/AR}$	%	20.00	88.00	82.00
Cd_{AN}	(mg*kg^{-1})	0.00	0.32	0.05
$Cd_{AN/EDTA}$	%	1.00	6.00	2.00
pH (H_2O)		7.05	8.15	7.65
pH ($CaCl_2$)		6.49	7.54	7.20
EC 2.5	(µS*cm^{-1})	103.00	483.00	197.00
$CaCO_3$	mass-%	0.00	20.00	0.10
OM	mass-%	2.63	5.06	3.46
Al_{DCB}	mg*g^{-1}	0.71	2.33	1.14
Fe_{DCB}	mg*g^{-1}	7.17	12.43	10.00
Mn_{DCB}	mg*g^{-1}	0.45	1.03	0.85
Al_{ox}	mg*g^{-1}	1.39	3.57	1.80
Fe_{ox}	mg*g^{-1}	0.33	4.19	1.65
Mn_{ox}	mg*g^{-1}	0.42	1.00	0.82

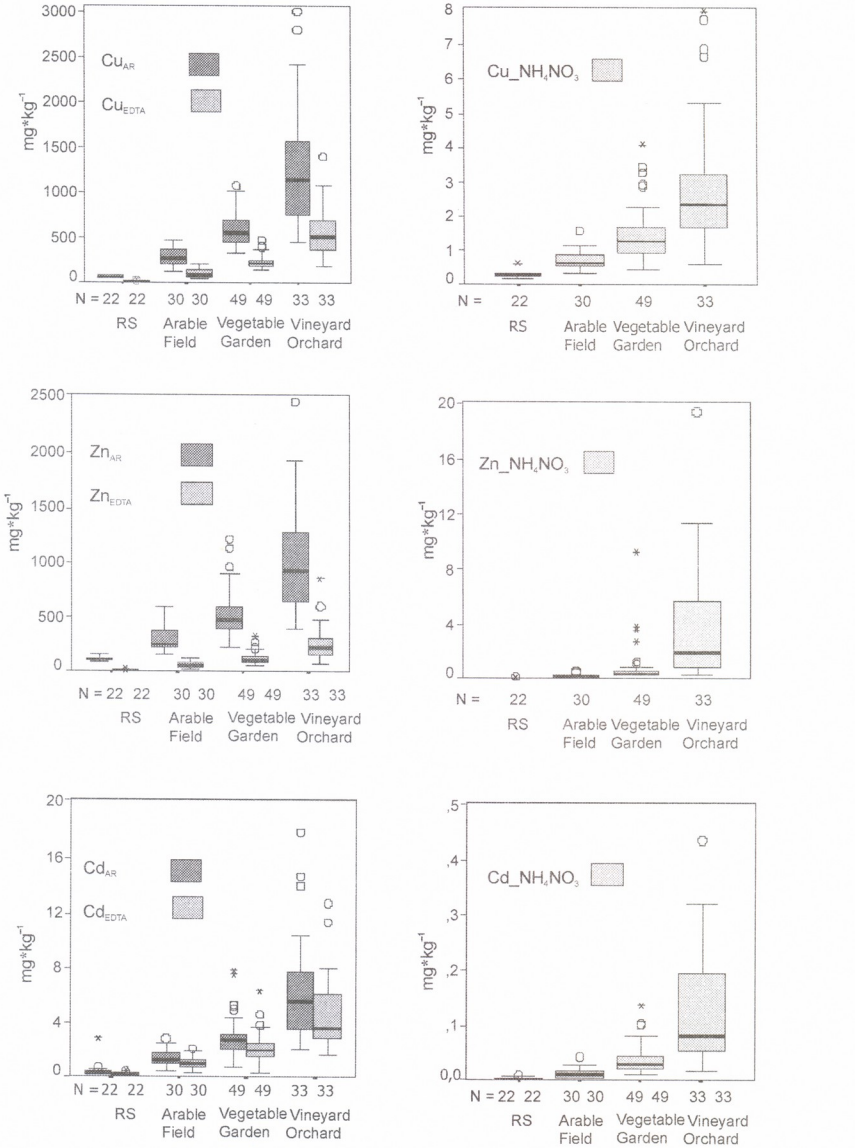


Fig. 2. Total amounts (box plots with medians, 25 and 75 percentiles, minimum and maximum) of heavy metals in topsoils of nonirrigated arable fields (RS = reference soils) in comparison to topsoils of irrigated sites with different land use.

Forty-five percent of Cu_{AR}, 22 % of Zn_{AR} and 82 % of Cd_{AR} belong to the supply fraction (*EDTA*, Table 4) and cause a high and long-term risk potential to the environment and food chain.

Background, precaution and target values of all three HM species are exceeded many times. The amounts of Cu_{AR} exceed the precautions values [16] by between 7 to 49 times. The Zn_{AR} concentrations are up to 18 times higher and the Cd_{AR} concentrations up to 12 times higher than the corresponding precaution value. The amounts of Cu_{AR} exceed the site-specific intervention values of the "Dutch List" (250 - 260 mg $Cu_{AR} \cdot kg^{-1}$) at all sites up to a factor of 11. At 45 % of the investigated sites the amounts of Zn_{AR} exceed the intervention value (210–214 mg $Zn_{AR} \cdot kg^{-1}$). This threatens both the soil functions and growth conditions of the vegetation.

Trigger values for the mobile fraction of Cu (1 mg $Cu_{AN} \cdot kg^{-1}$, [16]) are exceeded in 85 % of the investigated sites and those of Zn (2 mg $Zn_{AN} \cdot kg^{-1}$, [16]) in 45 %.

The evaluation of the Cd concentrations with respect to the relevant action values must take the mixed cropping into account. Grapevine are not strong accumulators of heavy metals. Therefore the higher action value of 0.1 mg $Cd_{AN} \cdot kg^{-1}$ [16] is valid, which is exceeded in about 45 % of the investigated sites. On the other hand, for growing vegetables, which are partly strong accumulators of HM, the lower action value of 0.04 mg $Cd_{AN} \cdot kg^{-1}$ [16] is applicable, which is exceeded in 82 % of the investigated sites.

The high concentrations of HM in topsoils of irrigated soils used for agriculture, house gardens, orchards and vineyards indicate strong contamination due to the deposition of fines with sulphidic metal particles and dissolved HM in acid mine drainage waters. The intra-element correlations show that all metals derive from the pollution of the Mashavera river by sulphidic mining wastes in the form of fines and acid mine drainage. The total amounts of HM in the topsoils depend on the type of land use and are related to the kind of crops and the intensity of irrigation (Fig. 2).

The oxidation of sulphides after deposition at the soil surface or incorporation to the topsoil by infiltration, cultivation and bioturbation leads to a dissolution of HM accompanied by the formation of sulfuric acid. Acid conditions at the microscale favor the mobility of HM ions. Due to the carbonate contents of most topsoils, the sulfuric acid is buffered rapidly and dissolved HM become immobile due to adsorption mainly on the surface of clay minerals, carbonates, iron and aluminum oxides and humic acids. Because the chernozems, phaeozems and kastanozems of the Mashavera valley show high contents of clay and organic matter, moderate contents of iron and aluminum oxides, and low to high contents of carbonates, the buffer capacity of the topsoils for HM is rather high.

The bulk of immobile, specifically adsorbed plus mobile or weakly adsorbed HM form the supply fraction. Highly significant correlations exist between amounts of the supply fraction and the amounts of adsorbents in the topsoils. The important organic substances for the adsorption of the supply fraction increases in the sequence Cd ($r = 0.43$) < Zn ($r = 0.56$) < Cu ($r = 0.58$). The highly significant correlations between the amounts of Fe and Al oxides and the supply fraction increase in the sequence Cu ($r = 0.80$ Fe_{ox}, 0.85 Al_{ox}) < Zn ($r = 0.58$ Fe_{ox}, 0.64 Al_{ox}) < Cd (no significance Fe_{ox}, 0.62 A_{ox}). This shows that the adsorption of Cu and Zn on soil particles is stronger than that of Cd, which therefore has the highest mobility and risk potential to the food chain, also indicated by the lowest threshold values displayed in [16] and the "Dutch List".

While the high affinity of Cu to different fractions of organic substances is well proven [17,18,20], the adsorption of Zn, as reflected by the coefficient of correlation, is something special. Obviously the high adsorption of Zn on organic matter reflects a strong loading situation [4]. Due to a relatively low amount of Fe and Al oxides and the saturation of their adsorption capacity for HM ions, Zn is bound first to humic acids and then migrates with time into the oxide minerals where it is specifically adsorbed. Similar observations were made by [21] and [22,23] referred to the great environmental importance of the supply fraction, due to significant correlations between supply (EDTA) and mobile (NH_4NO_3) fractions. The latter are relevant for the amount of HM uptake in food plants [24,11-13, 25]. In this study the highly significant correlations (r) between the supply fractions (EDTA) and the mobile fractions (NH_4NO_3-extractable) increase in the sequence Cu 0.75 < Zn 0.82 < Cd 0.86.

The supply fraction of HM in topsoils represents a high future risk potential, which is presently still increasing due to continuous irrigation with polluted water from the Mashavera and Kazretula rivers. In case of a slow change of soil properties due to land use, which influences conditions for specific HM adsorption, parts of the supply fraction will become more available. Decomposition of organic matter, decalcification, increasing acidity or intermittent reducing conditions [26], due to excess of irrigation water or logging of surface water, would be the main factors. Due to the recent high adsorption capacity of the soils for HM [27,28], only small amounts of HM in the mobile fraction were found with proportions of less than 1 % of the total amounts for Cu and Zn and a maximum of 1.5 % for Cd.

First results from screening investigation of crops from house gardens as well as field and pot experiments with wheat and spinach indicate a high uptake of Cu, Zn and Cd in cereals and vegetable leaves (manuscript in prep.) that exceed tolerance thresholds for plants, animals and human beings. Although Cu is adsorbed specifically mainly on organic matter and therefore the least mobile element in neutral to weakly alkaline soils (indicated by the results of the NH_4NO_3 extractions), the Cu contents in vegetation indicates a rather strong uptake. According to [29] the mobility of Cu and other metal ions increases in alkaline soils due to the formation of soluble

organic complexes. In future studies the HM uptake in food crops must be investigated systematically. Furthermore it should be established that the characterization of the mobile fraction with the NH_4NO_3 extraction method after [11,12], which was developed for neutral to acid Middle European soils, is also valid for alkaline soils. Due to the possible volatilization of NH_3 in an alkaline environment, the concentration of NH_4 is probably too low for the exchange of weakly adsorbed HM. In a weakly alkaline milieu they could be mobilized due to direct contact with plant roots by acid root exudates.

Nevertheless, the concentrations of the mobile Cd fraction exceed the precaution, trigger and action values of [16] and the target and action values of the "Dutch List" in topsoils of many sites. According to the German Soil Protection Law, land use restrictions and remediation measures would apply to 30 % of the investigated house gardens and to more than 50 % of the grape fields, vineyards and orchards with mixed cropping of vegetables. Therefore the actual risk of Cd transfer into the food chain is proven, which affects the local population as well as people in cities up to the capital Tbilisi, where crops from the Mashavera valley are sold at the open markets.

Acknowledgements

The authors thank the Volkswagen Foundation for the generous funding of the research project I/76 908: *Bergbaubedingte Schwermetallbelastungen von Böden und Nutzpflanzen in einem Bewässerungsgebiet südlich von Tiflis/Georgien – Ausmaß, ökologische Bedeutung, Sanierungsstrategien.* Furthermore we thank the DAAD for a grant for Dr. Beso Kalandadze, with which he was able to make a research exchange at the Institute of Soil Science and Soil Conservation, Giessen, to gain experience with the analytical methods. Dr. Klaus Buschhoff and Dr. Wolfram Kneib from the KfW project *Soil Cadastre* in Tbilisi supported our project by supplying valuable ortho-photographs of the study area. A broad hydrological and climatological data bank as well as maps were supplied by Prof. N. Begalishvilli, Hydrometerological Institute of the Georgian Academy of Sciences and by George Gidziridze of the GIS & RS Consulting Center. The close cooperation with Prof. Lorenz King, Institute of Geography of Justus-Liebig-University Giessen, and his help due to interpretations of satellite images, climatological and hydrological data sets, greatly assisted the project. Wako Narimanidze was a valuable member of the team, due to his careful driving and technical support, and the digging and sampling of many soil profiles. The project was administratively handled at the Centre of International Development and Environmental Research CIDER of the Justus-Liebig-University Giessen. Any problems with the sometimes difficult administrative and budget management were always solved by Dr. Mathias Höher and Waltraud Mirthes.

ნიადაგმცოდნეობა

მდინარე მაშავერას ხეობაში ნიადაგების და სასოფლო-სამეურნეო ნათესების გაჭუჭყიანება მძიმე ლითონებით

პ. ფელიქს-ჰენინგსენი*, თ. ურუშაძე**, ე. ნარიმანიძე£, ლ. ვისმანი£, ს. სტეფენსი§, ბ. კალანდაძე¥

* *ნიადაგმცოდნეობის და ნიადაგის დაცვის ინსტიტუტი, იუსტუს-ლიებიხის უნივერსიტეტი, ჰისენი, გერმანია*
** *აკადემიის წევრი, საქართველოს სახელმწიფო სასოფლო-სამეურნეო უნივერსიტეტი, თბილისი*
£ *საერთაშორისო განვითარების და გარემოსდაცვითი კვლევის ცენტრი, გისენი, გერმანია*
§ *მცენარეთა კვების ინსტიტუტი, იუსტუს-ლიებიხის უნივერსიტეტი, გისენი, გერმანია*
¥ *ივ. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

მდინარე მაშავერას ხეობის (სამხრეთ-აღმოსავლეთი საქართველო) განოყიერებადი, სარწყავი ნიადაგები გამოირჩევა სასოფლო-სამეურნეო მოსავლიანობის მაღალი პოტენციალით. თუმცა, სარწყავად გამოყენებული მდინარის წყლები გაჭუჭყიანებულია სამთო-მოპოვებითი მრეწველობის სპილენძის და ოქროს საბადოების

ნარჩენებით, რომლებიც განლაგებულია მდინარე მაშავერას შუა წელის მთიან რეგიონში. გარდა ამისა, გაჭუჭყიანებული წყლები, რომლებიც ჩამოირეცხება სასოფლო-სამეურნეო სავარგულებიდან, ეროზიული მასალა და სხვა ფაქტორები განსაზღვრავენ გარკვეული ჯგუფის მძიმე ლითონების მაღალ კონცენტრაციას. საირიგაციო არხების ნაფენებსა და მდინარე მაშავერაში სპილენძის, თუთიის და კადმიუმის კონცენტრაცია საკმაოდ მაღალია. საირიგაციო წყლებში განსხვავება მძიმე ლითონების შემცველობაში საკმაოდ არსებითია. განსხვავებები ხანგრძლივობაში, სიხშირესა და ირიგაციის რაოდენობაში, და აგრეთვე მიწათსარგებლობის ტიპი და ნიადაგების კულტივაცია (მაგალითად, ღრმა ხვნა) არის გაჭუჭყიანების მაღალი ვარიაბელობის მიზეზი.

REFERENCES

1. *C.J. Moon, G. Gotsiridze, V. Gugushvili, M. Kekelia, S. Kekelia, R. Migineishvili, Z. Otkhmezuri N. Özgür* (2001), Mineral deposits at the beginning of the 21st century: 309-312.
2. *A.G. Twaltschrelidze* (2002), Erzlagerstätten in Georgien. In: Gambaschidze, I., Hauptmann, A., Slotta, R. & Yalcin, Ü. (Eds.): Georgien - Schätze aus dem Land des Goldenen Vliesses, Bochum.
3. *R. Migineishvili* (2001), Proceedings of the 4th International Symposium on Eastern Mediterranean Geology, Sparta, Turkey, 21-25 May 2001, : 229-233.
4. *K. Burnod-Requia* (2003), Mining, industry and environment. - Environmental Performance Reviews, Georgia, UNECE:107-118.
5. MSPM JSC Madneuli - Technical Highlights. 2000.
6. *G. Nakhutsrishvili* (2000), Biological and landscape diversity of Georgia. WWF Georgia Country Office, Tbilisi.
7. *A. Javakhishvili, G. Gvelesiani* (1962), Soviet Georgia. - Acad. Sci. Georgian SSR, Vakhushti Institute of Geography. Progress Publishers, Moscow (in Russian).
8. *M.F. Purnell* (2000), World Reference Base for Soils. - Report of the correlation tour in Georgia, 20. July to 1 August 2000.
9. *C. Parat, R. Chaussod, J. Leveque, S. Dousset, F. Andreux* (2002), European J. Soil Sci., 53: 663 - 669.
10. *E. Schlichting, H.-P. Blume, K. Stahr.* (1995), Bodenkundliches Praktikum. - 2. Aufl.. Pareys Studientexte 81, Blackwell, Berlin.
11. *H. Zeien, G.W. Brümmer* (1989), Chemische Extraktion zur Bestimmung von Schwermetallbindungsformen in Böden. - Mitt. Deutsch. Bodenkundl. Ges., 59/I: 505 - 510.
12. *H. Zeien, G.W. Brümmer* (1991), Chemische Extraktionen zur Bestimmung der Bindungsformen von Schwermetallen in Böden. - In: Sauerbeck, D. und Lübben, S. (Eds.): Auswirkungen von Siedlungsabfällen auf Böden, Bodenorganismen und Pflanzen. Ber. z. ökol. Forsch. 6: 62 - 91, Jülich.
13. *V. Hornburg, G.W. Brümmer* (1993), J. Plant Nutr. Soil Sci.,156: 467-477.
14. AbfKlärV Klärschlammverordnung vom 15. April 1992. - BGBl. I, S. 912, zuletzt geändert durch Artikel 4 der Verordnung vom 20. Oktober 2006, BGBl. I, S. 2298.
15. BBodSchG, Federal Soil Protection Act. - Federal Law Gazette I, p. 502. 1998.
16. BBodSchV, Federal Soil Protection and Contaminated Sites Ordinance. - Federal Law Gazette I, p. 1554. 1999.
17. *G.W. Brümmer, U. Herms* (1983), Effects of accumulation of air pollutants in forest ecosystems: 233-243, Reidel, Dordrecht.
18. *I. Atanassova* (1995), Environmental Pollution, 87: 17-21.
19. *I. Atanassova* (1999), Water, Air, and Soil Pollution, 113: 115-125.
20. *B.J. Alloway* (1999), Schwermetalle in Böden. Springer-Verlag, Berlin – Heidelberg.
21. *G. Sposito* (1985), Concepts in metal toxicology, Metal ions in biological systems, Dekker, New York.
22. *F.X. Han, A. Banin, G.B. Triplett* (2001), Soil Science, 166: 18-28.
23. *U. Herms, G.W. Brümmer* (1984), J. Plant Nutr. Soil Sci.,147: 400-424.
24. *G.W. Brümmer, J. Gerth, U. Herms* (1986), J. Plant Nutr. Soil Sci., 149: 382 -398.
25. *H.E. Allen, C.P. Huang, G.W. Bailey, A.R. Bowers* (1995), Lewis Publ., Boca Raton, Ann Arbor, London, Tokyo.
26. Charlatchka, R. & Cambier, P. (2000), Water, Air and Soil Pollution, 118:143-167.
27. *K.G. Tiller, J. Gerth, G.W. Brümmer* (1984a), G. Geoderma, 34: 1 - 16.
28. *K.G. Tiller, J. Gerth, G.W. Brümmer* (1984b), G. Geoderma, 34: 17 - 36.
29. *G. Welp, G.W. Brümmer* (1998), J. Plant Nutr. Soil Sci., 162: 155-161.

Received May, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Medical Sciences

Revascularisation of Arterial Prosthesis and Autovenotransplant

Nino A. Javakhishvili*, Mamia E. Komakhidze†**

* *Academy Member, A.Natishvili Institute of Morphology, Tbilisi*
** *Member of the Russian Academy of Medical Sciences. A.Natishvili Institute of Morphology, Tbilisi*

ABSTRACT. After implantation of prosthesis in the artery, incapsulation and formation of neointima, there develops "vasa vasorum", which penetrates into the prosthesis from the outside and from the inner side – from the prosthesis lumen. Internal initiated vessels, contacting with the external, anastomose with each other. In the prosthesis internal lining original capillary network of hexahedral form is produced. This geometric form depends on decusation of longitudinal and transversal filaments of prosthesis fibers.

Practically the importance of neointima vascularization lies in the fragility of capillaries, resulting in small hemorrhages formation. Neointimal capillary topography, their exclusive location to prosthesis lumen, these small hemorrhage foci can simulate a coagulation chain reaction in the prosthesis.

After autovenoplasty in the artery and after restoration of blood flow the autovenotransplant sharply extends under arterial pressure, and there was no case of rupture or aneurism development in the transplanted veins. In some weeks the transplanted vein changes sharply, the wall thickenes and, which is of great importance, vascularization of vein changes and becomes similar to the vascularization of the artery, into which the vein has been transplanted. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *blood vessel alloplasty, autovenoplasty.*

On the basis of large experimental material (dogs) the animal organism's response to vascular plastic reconstruction of arteries and veins has been studied.

Besides various methods of morphological investigation the injection of blood vessels with Indian ink-gelatine mass according to M. Komakhidze's method has been used. The latter added significant originality to the studies carried out.

The organism's response to encapsulation-implantation of different kinds of substitutes: autoveins, hard and soft-porous artificial prosthesis – has been examined.

Testing of hard artificial substitutes of arteries (polyvinyl alcohol and polychlorvinyl tubes) showed that they were suitable only for short-term blood vessel substitution, because of their rapid thrombosing. It is caused by the absence of porosity in the tube walls and their internal surface smoothness, on which fibrin deposit is not retained, as well as significant activation of the blood coagulating system. After hard tubes occlusion and thus their exclusion from the blood circulation, the parameters of blood coagulating system return to the initial level. The problem was solved by the usage of synthetic fibres: nylon, dacron, teflon, terylene, ivalon, orlon, vinon, PTFE-Gore-Teks, kapron, lavsan (Soviet manufacture). A significant step was the usage of goffered artificial tubes, offered by Edwards and Trapp (1955), which facilitated selection of the necessary length of artificial vessels, but not all have taken a positive view of this improvement (Weibel, Szilagyi, 1959; Chvapil, Krajiĉhek, 1966).

Searches and efforts to improve artificial substitutes of an artery continued, each attempt being pronounced as a stage of development. Watanuki (1978) invented felt-like dacron, the firm Meadox (France) created high-perforated, knitted micro-vellur artificial vessels (Perini et al. 1995).

However, long-term testing nevertheless gave preference to factory manufactured goffered dacron porous tubes.

Testing of various soft-pored artery substitutes (capron, lavsan, dacron) has shown that the process of their incorporation principally follows one and the same scheme. This process is subdivided into three consecutive periods or stages: the first one - formation of fibrinous cover, the second - fibrin organization and development of the prosthesis connective tissue capsule, and the third - definitive inner lining - prosthesis neointima formation. All authors studying artificial arteries mark these stages [1-5, 8, 9, 11-15, 19, 21].

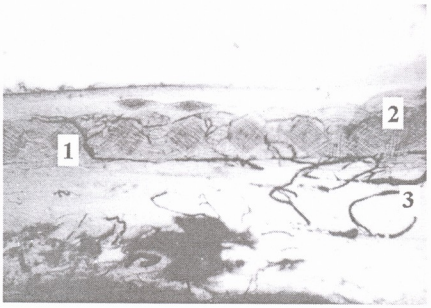


Fig. 1. Prosthetics of the dog's abdominal aorta with smooth capron tube. Duration of experiment (DE) 85 days. Longitudinal cut of a capron tube - 1, injected capillaries in the external prosthesis capsule - 2, capillaries between capron middle fibres - 3, capillaries in the internal lining of prosthesis 4 x 35.

The first stage, i.e. blood coagulation on the prosthesis wall immediately follows after blood circulation starts. In the crimped prosthesis fibrin fills the deepenings and makes the tube surface smoother. Fibrin closely attaches to the prosthesis filaments and contains blood cells. The second stage, i.e. connective tissue germination, occurs in fibrin very quickly after prosthesis inclusion; it initiates from the prosthesis ends and extends to the middle part, and simultaneously - from the surrounding tissues, developing from the outside into the internal part. The whole second stage should be considered as two periods: usually during the first fortnight the fibrinous cover organization process is carried out mainly owing to cell element germination (fibroblasts, epithelioid, plasmatic and lymphoid cells).

Later (2-3 months) the inner and upper cover of the prosthesis takes the form of dense fibrous connective tissue. Fibrous cover of the prosthesis is substituted by connective tissue in two ways: from the ends of the prosthetic vessel and on account of surrounding tissues. That is why demands to prosthesis porosity are determined as the demands to hemostasis, and the necessity of connective tissue germination through pores.

The third stage - formation of the definitive inner lining of the prosthesis, accompanied by formation of neointima endothelial cover is shown to develop gradually. Usually the layer of endothelial cells grows from the anastomosis edges, extending to the middle part of the prosthesis, and ends with total endothelialization of the whole internal surface. This process develops with various intensity in different patients and thus the terms of its completion might be determined relatively from several months to one year. The cells lying on the growth edge are large, of irregular form, polynuclear, close to anastomosis location they do not differ from the ordinary endothelial cells, as well as the cells of completely formed neointima. Its endothelization runs quicker from the central anastomosis edge and a bit slower - from the peripheral anastomoses.

It is rather disputable whether neointimal cover cells are true endothelial cells. But practically it is unimportant from the prosthesis full value, because these cells carry out endothelial function.

Connective tissue having developed round the prosthesis is rich in blood vessels. For the first fortnight prosthesis connective tissue capsule contains a large amount of fissures, which are filled with the contrast mass by vessel injection. Later the wall round the fissures is found to be organized, endotheliolized, and the fissures reduce to vessel capillaries, precapillaries, and their walls become of typical structure corresponding to the vessel gauge.

Small blood vessels like *vasa vasorum* combined with connective tissue fibers germinate inside through prosthesis pores and make the internal lining vascularized (Fig. 1). But another internal lining vascularization source exists. These are vessels initiated from the pros-

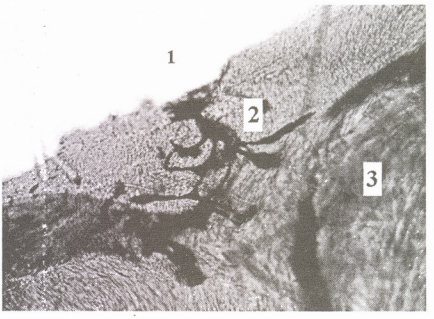


Fig. 2. Prosthetics of the dog's abdominal aorta with goffered lavsan tube. DE 17 days. Inner lining vascularization by the vessel growing from the prosthesis lumen - 1, newly organized vessels are of flask-like form - 2, prosthesis tissue - 3. Vessels injection, x 100.

thesis lumen (Fig. 2). These internal *vasa vasorum* not only provide the internal lining blood supply, but simultaneously are responsible for its endotheliolization (Fig. 3).

Internally initiated vessels penetrate outside through the prosthesis pores, contacting with the external *vasa vasorum*. These and those vessels widely anastomose with each other.

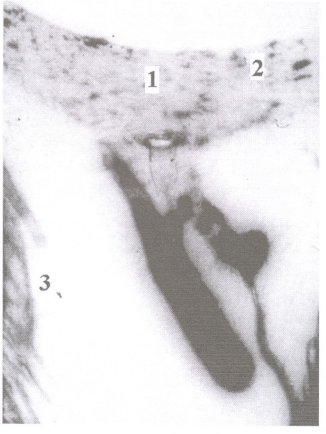


Fig. 3. Prosthetics of the dog's abdominal aorta with goffered lavsan tube. DE 80 days. Inner lining vascularization by the vessel initiated from the prosthesis lumen - 1, neointima - 2, prosthesis tissue - 3. Vessels injection, x 100.

The blood vessels vascularizating the prosthesis internal lining make in its thickness the original capillary network of hexahedral form (Fig. 4). This geometric form of the capillary network internal lining depends on the prosthesis. The prosthesis filaments being closely interwoven form eminences at the longitudinal and transversal filaments decussation spot, and as well as

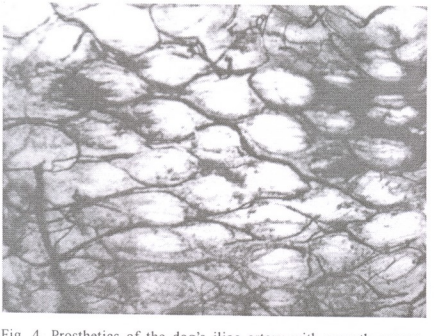


Fig. 4. Prosthetics of the dog's iliac artery with smooth capron tube. DE 201 days. Injected capillary network ansae in the inner prosthesis lining are of hexagonal form, x 50.

Fig. 5. Prosthetics of the dog's abdominal aorta with goffered Dacron tube. DE 127 days. Hexagonal ansae of the capillary network in the inner prosthesis lining. In the centre at the point of injection defect the capillary endothelial lining is observed, x 100.

deepenings-grooves on the decussation edges, where the blood pressure upon them seems to be rather weak and this is the cause of vascular network peculiar form (Fig. 5).

Practical importance of neointima vascularization is obvious in the following: capillaries are the most fragile part of the vascular system and that is why they easily break, especially newly formed, which results in small hemorrhages formation. In our case, owing to neointimal capillary topography, i.e. their exclusively close location to the prosthesis lumen, which does not take place in the arterial and vein intima, these small hemorrhage foci can stimulate blood coagulation chain reaction in the prosthesis. Thus neointima vascularization appears to be one of the causes of arterial prosthesis late thrombosis.

The chance of thrombosis after alloplasty by artificial prosthesis is even greater in an artery of small calibre than an artery of greater calibre (aorta), owing to the arrangement of newly developed capillaries directly under neo-endothelium, which does not take place in normal vessels and in an artery. We decided to test

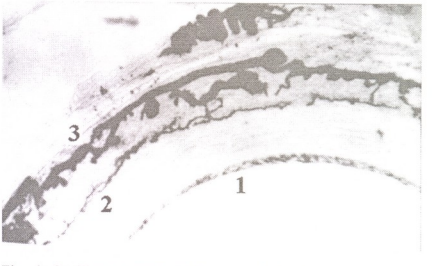


Fig. 6. Capillary network in the wall of the dog's common carotid artery. Arterial lumen - 1, perimuscular capillary network - 2, periadventitial capillary network - 3. x 35.

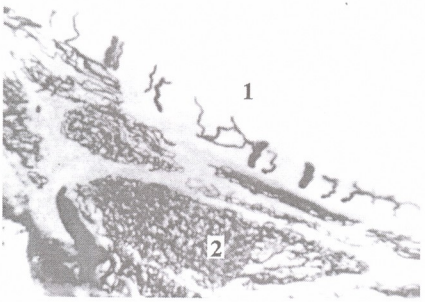


Fig. 7. Capillary network in the dog's v. saphena magna - 1, perivascular network - 2, x 35.

the possibility of replacement of an artery with own vein.

Veins proper (jugularis, saphena magna) had been transplanted into various peripheric arteries (a. a. carotis com., femoralis, renalis) and in all the cases autotransplanted vein wall thickening had been observed, and which is of great importance, preformation of its wall vascularization that became similar to vascularization of artery in which the vein had been included.

This conclusion is made due to the difference between arterial and venous blood supply. The arterial wall has two circular plexuses: perimuscular and periadventitial (Fig. 6), but the vein has a single vascular network, which is located in adventitia (Fig. 7). As soon as vein begins to carry out the arterial function and blood is pulsating in it upon arterial pressure, vascular plexuses in its altered wall develop, which are similar to perimuscular and periadventitial plexuses of arteries (Fig. 8).

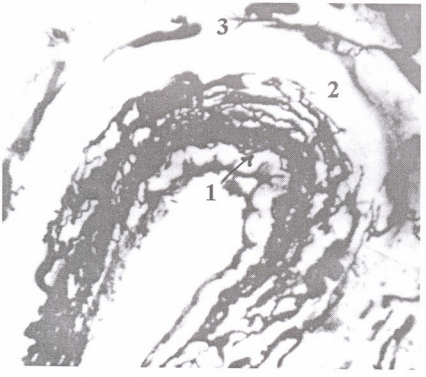


Fig. 8. Subcutaneous jugular vein implanted into the common carotid artery of the dog. DE 91 days. Capillary network in the transplant wall is sharply enlarged, no difference between periintimal - 1, perimuscular - 2 and periadventitial network - 3 gauges, x 35.

The first operations (1954) were made on the common carotid artery in which a segment (4-5 cm) was stitched in a piece of jugular or femoral vein (10 experiments). Following autvenoplasty was made on a hip. A segment of femoral arteries (3-4 cm) was replaced also by a femoral subcutaneous vein.

Each time after restoration of bloodflow, autovenotransplant sharply extended under the influence of arterial pressure and it seemed that it could burst but it never occurred; in one experiment development of aneurysms of a wall of the transplanted vein did not take place either. In some weeks the transplanted vein changed sharply, condensed and externally did not differ any more from the artery in which the stitch was.

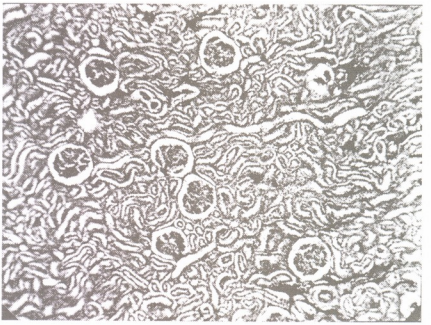


Fig. 9. Dog's right intact kidney after autovenoplast of left renal artery. (DE) 205 days. Glomerules are injected completely. Hematoxylin-eosin. x 80.

Becoming convinced of the stability of the autotransplant venous site, we decided to check the possibility of replacement of kidney arteries by a vein (1963). Pressure of blood in system kidney arteries is selectively high in comparison with other arteries of the same calibre. Renal hypertension less submits to hypotensive therapy, the reason of this pathology is often a change – narrowing, thrombosis of kidney arteries and other cases demanding surgical intervention.

Autovenoplasty of the left kidney was made to 44 dogs. In the beginning the right kidney was not touched. When we became convinced that at successfully executed operation the left kidney externally had not changed, at the second moment, some months later, the right intact kidney was removed. Dogs lived with one autovein transplanted kidney some years (2 dogs - 9 m, 1 dog - 1 y, 2 dogs - 116 m, 1 dog - 3 ys, 6 m, 1 dog - 3 ys, 11 m, 1 dog - 4 ys, 1 dog - 8 ys); we discontinued the experiment to carry out histologic research. One dog pupped, two dogs ran away from an open-air cage. At a

Fig. 10. Dog's left kidney after autovenoplasty of its artery. DE 205 days. *Glomerules* are well injected. *Hematox.-eosin*. x 80.

successful result the pressure of arterial blood kept within the limits of norm, nitrogen in blood 22-35%. At autopsy kidneys did not show signs of great damage of structure (Figs 9, 10), of ultrastructure (Figs 11, 12) as well as the vascular system (Figs. 13, 14, 15).

The duration of deenergizing of a kidney from blood circulation is of great value; 30 minutes are desirable,

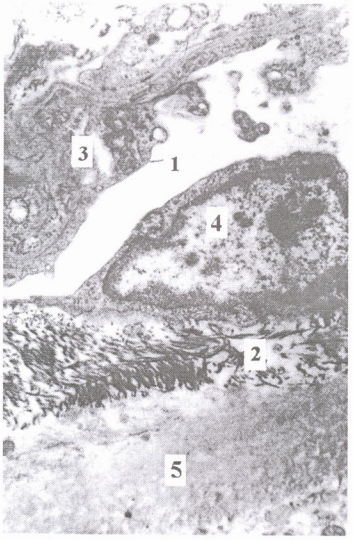


Fig. 12. Dog's left kidney after autovenoplasty of its artery. DE 2 years 9m. peritubular capillary is separated from the urinary tubules by rough collagen fascicles - 2, endotheliocyte is condensed - 3 and has no fenestra, endotheliocyte nucleus is unchanged - 4, basal membrane of the urinary tubule is significantly thickened - 5. x 7000.

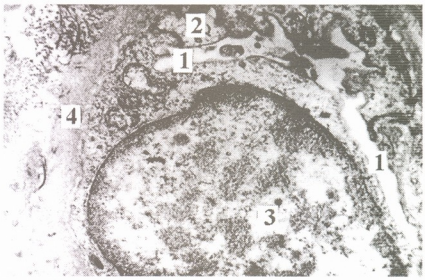


Fig. 11. Dog's left kidney after autovenoplast of its artery. DE 9 months. Peritubular capillary lumen, fissural one - 1, endotheliocyte cytoplasm is endemic - 2, nucleus - 3 protrudes into the lumen, pinocytosis is absent, basal membrane and pericapillary space contain collagenic fibrillae - 4. x 9000.

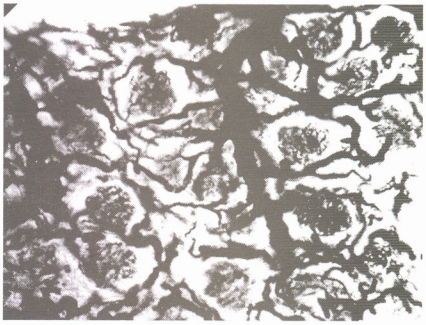


Fig. 13. Dog's left kidney after autovenoplasty of its artery. DE 200 days. All the intraorganic vessels, including glomerules, are well injected, x 120.

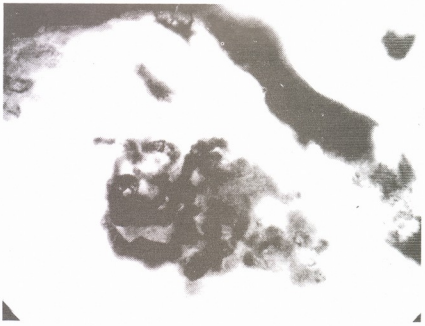


Fig. 14. Dog's left kidney after autovenoplasty of its artery. DE 200 days. Glomerulal vessels are significantly extended and well injected. x 300.

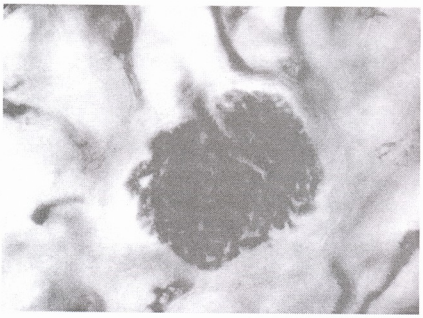


Fig. 15. Dog's right intact kidney. Glomerular vessels are well injected, all the ansae are filled with contrast substance, x 300.

not longer. The calibre of a dog's renal arteries is very small, diameter of 2-3 mm, the form of branching is also important. After operation complications in the system of intestinal vessels and the portal system, and during roentgeno-vasocinematography are frequent.

Especially obvious is the full value of autovenotransplant in the case of renal artery substitution. After successful operations the dogs endure the second kidney ablation and live for several years with a single kidney, whose artery had been replaced by the autovein. The structure and ultrastructure as well as the function of such kind of kidney are preserved.

Autovenoplasty of renal artery is necessary not only at renal hypertension, but also at transplantation of a kidney. Unfortunately, the viability of the replaced vein is not great [7, 10, 16-18, 20].

სამედიცინო მეცნიერებანი

არტერიის პროთეზის და აუტოვენის ტრანსპლანტატის რევასკულარიზაცია

ნ. ჯავახიშვილი*, მ. კომახიძე**†

* *აკადემიის წევრი. ა. ნათიშვილის მორფოლოგიის ინსტიტუტი, თბილისი*
** *რუსთის სამედიცინო მეცნიერებათა აკადემიის წევრი, ა. ნათიშვილის მორფოლოგიის ინსტიტუტი, თბილისი*

ნაშრომში ნაჩვენებია სისხლის ძარღვების ახალგანვითარება და ჩაზრდა არტერიის ხელოვნური შემცვლელის გარშემო განვითარებულ ქსოვილებში, მათ შორის ნეოინტიმაში (შიგნითა გამონაფენში), რასაც არა აქვს ადგილი არც არტერიაში და არც ვენაში. გამოთქმულია აზრი, რომ ნეოინტიმაში ახლად ჩაზრდილი სისხლძარღვები შეიძლება იყოს ერთ-ერთი მიზეზი არტერიის პროთეზის გვიანი თრომბირებისა, ნეოინტიმასთან – სანათურთან ახლო მდებარეობის გამო.

არტერიის საკუთარი ვენით შენაცვლების შემდეგ ვენის კედელი სწრაფად იცვლება, განიცდის „არტერიალიზაციას" და უძლებს მაღალ არტერიულ წნევას, რის გამოც ეს აუტოტრანსპლანტაცია საიმედოა და შეიძლება გაკეთდეს მსხვილ არტერიებშიც.

მონაცემები არტერიების ხელოვნური პროთეზებისა და საკუთარი ვენით არტერიის შეცვლის ვასკულარიზაციის შესახებ მეოცე საუკუნის ბოლო ათწლედების მრავალრიცხოვან ლიტერატურაში არ არის. ამიტომ წერილს ვურთავთ მხოლოდ ჩვენი ზოგიერთი შრომის დასახელებასაც.

REFERENCES

1. *M.V. Anichkov et al.* (1963), Vestnik khirurgii, 2, 41-44 (Russian).
2. *V.A. Zmur et al.* (1959), Vestnik khirurgii, 4, 71-79 (Russian).
3. *M.I. Perel'man, Yu.A. Rabinovich* (1965), Vestnik khirurgii, 3, 7-11 (Russian).
4. *B.V. Petrovoskii* (1963), Vestnik khirurgii, 3, 10-17 (Russian).
5. *Yu.G. Shaposhnikov* (1961), Primenenie autoven, podkreplennykh perlonovymi odnosloinymi obolochkami, dlya plastiki perifericheskikh arterii v eksperimente, v/ch. pp.63032 (Russian).
6. *C.D. Campbell* (1980), J. Surg. 4, 2, 227.
7. *C.D. Campbell et al.* (1979), Ann. Surg. 190, 6, 740-742.
8. *C. D. Campbell et al.* (1976), A preliminary report. Surgery, 79, 5, 485-493.
9. *M. Chvapil, M. Krajicek* (1966), Statni zdravotnistkie nakladet. December 153, 163 (Czech).
10. *W. A. Dale* (1959), Autogenous Vein Grafts and Related Aspects of Peripheral Arterial Disease, Berlin.
11. *W. A. Dale* (1974), Surgery, 76, 6, 849-866.
12. *R. C. Darling* (1980), J. Surgevry, 4, 4, 2, (229).
13. *M. E. De Bakey, G. M. Lawrie* (1979), Ann. Surg, 189, 3, 303-305.
14. *J. Descotes, J. Brudon* (1986), Lyon Chirurg., 82, 3, 145-146 (French).
15. *J. Descotes et al.* (1985), Lyon Chirurg., 81, 1, 59-60.
16. *J. Descotes et al.* (1974), Lyon Chirurg., 70, 4, 259-260.
17. *R. M. Dickerman et al.* (1980), Ann. Surg., 182, 5, 639-644.
18. *J. P. Gamondes et al.* (1981), Lyon Chirurg., 77, 4, 231-235.
19. *J W. Hallett et al.* (1980), Ann. Surg., 191, 4, 430-437.
20. *X. Martin et al.* (1985), Lyon Chirurg., 81, 322-326.
21 *N. S. Martinez et al.* (1957), Surgery, 42, 6.

Received August, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Medical Sciences

Relationship between Depression and Religiousness with the Atherogenic Profile of Lipoproteins in Arterially Hypertensive Patients

Nodar Kipshidze*, Sesili Kabisashvili, Nona Kakauridze****

* *Academy Member, National Centre of Therapy, Tbilisi*
** *National Centre of Therapy, Tbilisi*

ABSTRACT. The research carried out during the last two decades made it possible to confirm from the standpoint of the proof-based medicine the relationship of stress, mental depression and anxiety with the cardiovascular system diseases (CVSD). The majority of depressive persons has been found to manifest a higher percentage of a coronary heart disease (CHD), arterial hypertension and other cardiovascular pathologies. According to an analytical forecast of the World Health Organization, by 2020 CVSD and mental depression will be leaders in reducing full-value life in the whole world. Lately, more obvious has become the interest of science in studying the relationship of religion with depression, stress, and AH, since many studies are known to have corroborated the existence of a direct correlation between the said factors. Proceeding from the above, a comprehensive study of the relationship between depression and CVSD presents a rather topical problem of medicine.

The purpose of the study is to establish a relationship between religiousness, mental depression and lipids in patients with AH.

Elderly patients aged 60 to 74 were subjects of our study. Patients were divided into four groups according to their sex and religion. The depressive state used to be determined by the Fallstein Geriatric Depression Scale. Religiousness was measured according to personal prayers and the frequency and regularity of participation in religious rituals. The studies revealed these criteria to be most sensitive in terms of evaluation of religiousness. Of biochemical parameters, the following were measured in blood serum: TCh, Tg, HDLCh, LDLCh. As a result of an analysis of the study results, the following data have been obtained: a reliably low depression assessment points are observable in the believers against the non-believers; respectively, in the believer males a low correlation of depression, AH and TCh is observed. In the non-believer males a high direct correlation of depression and AH, AH and TCh, LDLCh and Tg is observed. In the believer females, a negative correlation of depression and TCh, LSLCh and Tg is observed, whereas in the non-believer females such correlation is positive. Proceeding from the above, the favorable impact of religiousness on the manifestations of depression, as well as the AH and lipid metabolism can be assumed. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *lipides, depression, religion, arrhythmia.*

The relationship of stress, mental depression and anxiety with the cardiovascular system diseases (CVSD) has long been known, although only during the last two decades such correlation could be confirmed from the standpoint of the proof-based medicine. As a result of long-term prospective studies, it has been revealed that the majority of depressive persons is found to have or to frequently develop an ischemic heart disease (IHD), arterial hypertension (AH), acute myocardial infarction (AMI), and an ischemic stroke [1, 2]. Concurrently, from the standpoint of a CVSD development risk, of importance is not only the clinically apparent "big depression" but also the less mani-

fested depressive semiotics. The epidemiological studies carried out during the last decade have revealed high percentage of spread of anxiety, mental depression and other disorders of the depressive spectrum in the general medical and CVS diseases' treatment practice, which significantly exceeds their prevalence in a population [3, 4]. Both IHD and psychopathological disorders clearly deteriorate the human work-status, social functionality, adaptability and represent a serious problem for the contemporary society in terms of expenditures. According to an analytical forecast of the World Health Organization, by 2020 CVSD and mental depression will become the leading cause of reducing full-value life in the whole world [5]. It is established that upon coexistence of IHD and mental depression in patients, the social functionality indicators are twice lower than in case of existence of any one of these two diseases [4, 5]. Clinical studies have also demonstrated that a mental depression and an anxiety-depressive state worsen the CVSD state, negatively affect the treatment, rehabilitation and prevention processes [4-6]. Lately, more obvious is becoming the interest of science to study the relation of religion with mental depression, stress, and AH. Many studies are known to have corroborated the existence of a direct correlation between the said factors [4, 6-8]. Prayer has been found to reduce depressive semiotics both in females and males [4, 6]. Proceeding from all the above, a comprehensive study of the relationship between mental depression and CVSD presents a rather topical problem of medicine.

The purpose of the study is to establish a relationship between religiousness, mental depression and lipids in patients with AH.

Material and methods

117 elderly patients aged 60 to 74 were subjects of our study. Sixty patients were female, and fifty-six were male. According to religion, the patients were distributed into four groups: group I – 28 believer males; group II – 32 non-believer males; group III – 29 believer females; and group IV – 28 non-believer females. Depressive state used to be determined by the Fallstein geriatric depression scale that being designated to serve as a basic screening measure for functional status in older adults and their adaptation to the environment (1983). The results were assessed by points: 1-22 points indicate low-grade depression; over 22 points – full-blown depression. Religiousness was established according to personal prayers and the frequency and regularity of participation in religious rituals (weekly participation in church rituals), as well as the length of religious affiliation (not less than 10 years). The studies revealed these criteria to be most sensitive in terms of evaluation of religiousness [2, 9].

Of biochemical parameters, the lipid spectrum was determined in blood serum (following a 13-hour fasting on the empty stomach) by means of a spectrophotometer JENWAY-6400, using the enzymatic technique: total cholesterol (TCh) – GPO Kit BOPLABO, France, while high-density lipoprotein cholesterols (HDLCh) were measured following precipitation of low-density lipoprotein cholesterols (LDLCh) and very low-density lipoprotein cholesterols (VLDLCh) - (HDL Cholesterol (Precipitant) SFBC BIOLABO, France). LDLCh were measured according to Friedwald (1972) under the formula: LDLCh = TCh (HDLCh + VLDLCh) (Klimov A.N., Nikulcheva N.G., 1995). The AH degree was measured by the JNS VII classification; HID was diagnosed according to anamnesis, clinical presentations and the loading test. The following methods were used: for statistical processing of the material - M±SD (M-medium, SD-standard deviation); for data analysis – Student's t-test ($p<0.05$), for determining correlation (C) – Pearson method.

Results and discussion

Based on different studies, the relationship of religiousness with the CVSD development, as well as its risk factors: gender, age, dyslipidemia, hypertension, psychocharacterologic peculiarities were established [4, 6-8]. Hypertension occupies a special place in the pathogenesis of the development of such fatal complications as myocardium infarction, stroke, sudden death, especially among elder patients. From this point of view, also of interest is the fact that hypertension is a somatic manifestation of disadaptation which is in close relationship with the psychocharacterologic peculiarities on the one hand, and with the endothelium's dysfunction on the other, which causes the development of atherosclerosis.

We studied elderly people (aged 60-70) of both sexes with arterial hypertension of grade I and grade II, as a contingent with a risk to develop CVSD complications. The contingent was selected so that the groups would not show statistically reliable differences by the systolic and diastolic pressure figures ($p>0.05$). This makes it possible to assess the parameters of other risk factors for CVSD against the background of the arterial pressure constant. The results of the study conducted by us are given in the Table below.

According to our study results, a reliably low indicator (5.6±3.7) was observed in Group I – believer males as compared with non-believer males (12.16±7) $p<9.88$ E8. Similar data were obtained for females as well: in Group III – believer females the depression indicator (6.8±4.3) is reliably low as compared with Group IV – non-believer females (13.5±3.5) $p<4.9$E-0.7. The reason of the above can be probably religious consciousness encompassing components which in each specific situation, including under stress conditions, are directed at self-knowledge and cognition of the environment, the establishment of tolerance and the looking for a conformist way out, which conditions the establishment of psycho-adaptive mechanisms. In addition, an essential factor for religious consciousness is a hopeful attitude to future, which preconditions vitality at all ages. Also important are collectivity

Table 1

Pressure, lipid spectrum and depression indicators by religious division in the patients of both sexes

Group	SAP mmHg	DAP mmHG	TCh mg/dl	LDLCh mg/dl	HDLCh mg/dl	Tg mg/dl	DEP
I	165.9±8.3	86.1± 4.9	192.5±55.4	136.4± 9.3	47.1± 8.9	125.4±46.2	5.6±3.7
II (M±SD)	167.2±12	89.1 ± 10.4	232.1±6.7	140.7±39.9	42.1± 8.9	191.6±72.2	12.16±7
III (M±SD)	163.3 ±9.3	86.2± 4.6	219.9±42.9	143.2±43.6	47.0± 12.7	124.67±45.39	6.8±4.3
IV (M±SD)	166.7±10.6	87.6 ± 10.7	252.1±46.3	171.3±34.1	33.2±14.8	175.07±76.26	13.6±3.5
P I-II	0.6	0.1	0.005	0.6	0.03	0.0001	9.88 E8
P III-IV	0.2	0.5	0.01	0.01	0.2	0.002	4.9E07
P I-III	0.2	0.8	0.04	0.50	0.9	0.9	0.2
P II-IV	0.8	0.6	0.1	0.003	0.5	0.4	0.1

Note: SAP – systolic arterial pressure; DAP – diastolic arterial pressure; TCh – total cholesterol; LDLCh – low density lipoprotein cholesterols; HDLCh – high density lipoprotein cholesterols; Tg – triglycerides; DEP – depression scale indicators; (M±SD): M – medium indicator; SD – standard deviation; P – reliability coefficient ($p>0.05$).

and mutual aid. These assumptions are shared by many authors in whose studies a direct correlation of depression and tolerance is being observed [4]. The practice of praying and the participation in religious rituals are among the most important components of religiousness, which conditions the establishment of orderliness. Many experimental and epidemiological works by various authors, which are focused on the study of a relationship between depression and religiousness, evidence a favorable impact of the regularity of religious practice upon the psycho-emotional state of man. According to some authors, SAP, DAP, and pulse indicators improvement is observable against the background of prayer [4, 6-8]. The central place in the consciousness of a religious person is occupied by the belief in God, which probably represents the guiding force that preconditions overcoming personal weaknesses in conducting an organized religious mode of life and is distinguished thereby from non-religious persons who, as a rule, find more problems in following the healthy mode of life due to weaknesses of the character.

The lipid spectrum analysis revealed the following: low degree of correlation: in Group I -192.5±55.4; in Group II - 232±6.7 ($p<0.005$); LDLCh: in Group I – 136.4±9.3; in Group II – 140.7±39.9 ($p<0.06$); HDLCh: in Group I – 47.1±6.7; in Group II – 42.1± ($p<0.03$), and reliable Tg indicators: in Group I -125.4±46.2; in Group II – 191.6±72.2 ($p<0.0001$). Against Group III and Group IV the following indicators were observed: Correlation: in Group III – 219.9±42.9; in Group IV – 252.1±46.3 ($p<0.01$); LDLCh: in Group III – 143.2±43.6; in Group IV – 171.3±34.1 ($p<0.01$); HDLCh: in Group III – 47.0±12.7; in Group IV – 33.2±14.8 ($p<0.03$); Tg: in Group III -124.67±45.39; in Group IV – 175.06±76.25 ($p<0.002$).

According to numerous study data, the lipid metabolic disorder is one of the principal components in the development of atherosclerosis. At the same time, a concentration of individual fractions of the lipid spectrum in blood serum and a disorder of their correlation are of importance in the endothelium damage pathogenesis. Experimental studies evidence that an increase in the concentration of LDLCh and a decrease in the concentration of HDLCh condition the endothelium damage and impairment of its vasodilative function. There is a difference of opinions regarding an increase in Tg concentration as being an independent pathogenic component, although according to the guidelines of the International Atherosclerosis Society (IAS), a relationship between Tg and the development of atherosclerosis has been established, especially under conditions of an increase of the concentration of Tg and high concentration of LDLCh, or a decrease in the concentration of HDLCh, or in both cases concurrently [10]. As is noted, the level of lipids in blood is characterized by variability according to the individual biological indicators of man. Physiological changes under conditions of a stress represent one of such variations. At this time a prolonged and relatively stable increase of lipoproteins in the serum is being observed [4, 7, 10]. In addition, there exist data evidencing a relationship between a concentration of

atherogenic lipoproteins and psychocharacterologic peculiarities [4, 10]. As our study results indicate, the atherogenic LP concentration in the serum in the group of believers, both male and female, is reliably low against the non-believers, which can be explained by psycho-adaptive peculiarities of the believers and low indicators of the depressive manifestation. There are experimental study data, according to which the concentration of TCh in the serum is found to decrease during praying [3, 4]. Another variation is the character of diet. Diet peculiarities occupy an important place in the mode of life of believers, which is expressed by the consumption of low-calorie, vegetable food in the period of fasting (195 days a year on average), as well as by restriction of the food consumed. The reliably low concentration of Tg in the serum among believers, indicated in our study results, might be conditioned by the above.

As a result of the study, the following correlation has been found: in Group I – light C of DEP and TCh, r=0.33; in Group II – high C of DEP and SAP, r=0.71; DEP and DAP, r=0.42; SAP and Tg, r=0.52; DAP and LDLCh, r=0.59; in Group III: a negative correlation of DEP and TCh, r=0.33; DEP and LDLCh, r=0.35; in Group IV – DEP and SAP, r=0.73; DEP and Tg, r=0.32; DEP and LDLCh, r=0.43.

An analysis of the obtained results enables to state that the lipid metabolism in the group of believers approaches physiological standard, which conditions a lesser risk for the endothelium damage even under high AH conditions among the believers. This is evidenced by a clinical characterization analysis of the groups studied by us: in Group I – AH 0 – 21.7%, AH I – 25%; in Group II – 30%. Of the concomitant diseases: angina – 14.3%, heart insufficiency – 37.8%; arrhythmia – 14.3%; in Group II – AH 0 - 9.7%. AH I – 46.9%, AH 2 – 43.7%, angina – 43.7%, heart insufficiency – 53.1%, arrhythmia - 25%; in Group III - AH 0 – 33.3%. AH I – 66.6%, AH 2 – 7.4%, angina – 19.4%, heart insufficiency – 22.2%, arrhythmia 7.4%; in Group IV - AH 0 - 21.4%. AH I – 39.3%, AH 2 – 39.3%, angina – 34.3%, heart insufficiency – 37.9%, arrhythmia - 25%.

Conclusions

Religiousness represents a potent factor in prevention of CVSD and its risk factors (depression, hypertension, dyslipidemia) in the people of both sexes, since religiousness leads to a reliable reduction of a depressive state during arterial hypertension, which conditions the preservation of adaptability in such components of the homeostasis as the lipid spectrum and the renin-angiotensin system:

– During arterial hypertension less frequent cases (reliably low assessment points) of depression are observed in believers as compared with non-believers.

– Non-believer males show higher direct correlation between depression and AH, as well as between AH and TCh, LDLCh and Tg, whereas believer males, respectively, demonstrate only a weak correlation between depression, AH and TCh.

– High correlation between depression and AH, LDLCh and Tg has been established in non-believer females in contrast to believer females where such relationship is not observed.

სამედიცინო მეცნიერებანი

დეპრესიისა და რელიგიურობის კავშირი ლიპოპროტეიდების ათეროგენულ პროფილთან არტერიული ჰიპერტენზიის მქონე პაციენტებში

ნ. ყიფშიძე[*], ს. კაბისაშვილი[**], ნ. კაკაურიძე[**]

* *აკადემიის წევრი, თერაპიის ეროვნული ცენტრი, თბილისი*
** *თერაპიის ეროვნული ცენტრი, თბილისი*

ბოლო ორი ათული წლის განმავლობაში ჩატარებული კვლევების შედეგად შესაძლებელი გახდა სტრესის, დეპრესიისა და გულსისხლძარღვთა დაავადებების ურთიერთკავშირის დადასტურება მტკიცებაზე დაფუძნებული

მედიცინის პოზიციიდან. აღმოჩნდა, რომ იმ პირებში, ვისაც დეპრესია აღენიშნებოდა, არტერიული ჰიპერტენზია და სხვა გულსისხლძარღვთა პათოლოგია გამოულინდებოდა მაღალი პროცენტით. ჯანმრთელობის დაცვის საერთაშორისო ორგანიზაციის ანალიტიკური პროგნოზის მიხედვით 2020 წ. მთელ მსოფლიოში სრულფასოვანი სიცოცხლის შემცირების თვალსაზრისით გულსისხლძარღვთა დაავადებები და დეპრესია ლიდერები იქნება. ბოლო პერიოდში აღინიშნება მეცნიერების დაინტერესება რელიგიის კავშირის შესწავლით დეპრესიასთან, სტრესთან და არტერიულ ჰიპერტენზიასთან. ჩატარებულია კვლევა, რომელიც ადასტურებს აღნიშნულ ფაქტორებს შორის პირდაპირ კორელაციურ კავშირებს. აღნიშნულ ფაქტორებს შორის კავშირის სრულყოფილი შესწავლა მედიცინის მეტად აქტუალური ამოცანაა.

REFERENCES

1. *E.A. Gogin* (2000), Clinical Medicine, No. 11, pp. 4-7 (Russian).
2. *A.I. Chazov, et al.* (2007), Clinical and epidemiological program for studying depression in cardiology practice, in patients with arterial hypertension and coronary heart disease: (COORDINATA) long-term study results. Cardiology, No. 3, pp. 28-37 (Russian).
3. *Sander G. Robbins* (2001), Correction of lipid abnormalities. Basic principles and practical implementation of therapeutic interventions. Moscow: Medicine, p. 176 (Russian).
4. *E.L. Idler, M.A. Musick, M.A. Ellison, N. Krause, M.G. Ory, et al.* (2003), Measuring multiple dimensions of religion and spirituality for health research: conceptual background and findings from the 1998 General Social Survey. Research on Aging, 25: 327-365
5. Fetzer Institute/National Institute on Aging Working Group. (1999). Multidimensional measurement of religiousness/ spirituality for use in health research.
6. *B.S. Jonas, M.E. Mussolino* (2000), Symptoms of depression as a prospective risk factor for stroke. Psychosomatic Medicine, 62 463-471.
7. *J.W. Hughes, C.M. Stoney* (2000). Depressed mood is related to high-frequency heart rate variability during stressors. Psychosomatic Medicine 28,179-185
8. *V.M. Wilkins* (2005), Religion, Spirituality and Psychological Distress in Cardiovascular Disease, Doctor of Philosophy. Drexel University.
9. *H.G. Koing, D.O. Moberg, J.N. Kvale* (1988), Religious activities and attitudes of older adults in a geriatric assessment clinic. J Am Geriatr Soc., 36 (4): 362-374.
10. *B.S. Jonas, J.F. Lando* (2000), Negative effect as a prospective risk factor for hypertension. Psychosomatic Medicine, 62: 188-19.

Received August, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Medical Sciences

Growth Factors in Blood Serum of the Patients with Benign Prostatic Hyperplasia

Ruben R. Boroyan*

* *St.Grigor Lusavorich Medical Center, Yerevan, Republic of Armenia*

(Presented by Academy Member L. Managadze)

ABSTRACT. The aim of the present research is to define concentration of growth factors in the blood serum of the patients with benign prostatic hyperplasia (BPH). It has been established by means of high-sensitive immune-enzyme analysis that the patients with BPH aged 57-81 have higher basic fibroblast growth factor (bFGF) concentration compared with control group (healthy men). In most cases bFGF is found among 1.6±10.5 to 44.0±10.8 pg/ml, whereas β-1 transforming factor concentration is low and varies from 19.1±2.2 to 3.5±1.2 pg/ml. *© 2007 Bull. Georg. Natl. Acad. Sci.*

***Key words**: basic fibroblast growth factor, β-1 transforming growth factor and benign prostatic hyperplasia.*

I. Introduction

Benign prostatic hyperplasia is one of the vital ones in medical and demographic, social and economic aspects [1] and the problem of its pathogenesis is of topical importance in urology. Prostate is an androgen-dependent organ though androgens have a mediated mitogenic influence on epithelial cells [2] where the interaction between the epithelial and stromal cells through paracrine and autocrine factors play an important role in these processes. Basic fibroblast growth factor (bFGF) plays an important role in mitogenic effects of androgens. bFGF was isolated in the prostate in the 1980s [3-5] and is a representative of different proteins family and has a powerful mitogenic effect on the prostate epithelial and stromal cells [2]. Later some other fibroblast factors (FGF1, FGF3, FGF7 and FGF8) were also isolated in the hyperplastic prostate. These factors posses mitogenic activity regarding prostate stromal and epithelial cells but the most vital factor stimulating prostate growth is bFGF which displays its effect in the prostate through FGFR1 and FGFR2 specific receptors [6]. In the healthy prostate only stromal cells express bFGF receptors. This shows the importance of bFGF for mesenchyme homeostasis through stromal autocrine regulation. There were detected some other prostate growth factors stimulating prostate size growth. These are epidermal growth factors (EGFs) and transforming growth-α factor (TGF-α) [2]. On the other hand, there were also detected growth inhibiting factors whose prototype is a transforming growth factor-β1 (TGF-ß1). TGF-ß1 is a multifunctional polypeptide which affects androgen independently, thus regulating proliferation and differentiation of epithelial cells, apoptosis, extracellular matrix formation and degradation [7]. It is specific that TGF-ß1 is produced by smooth muscle cells but TGF-ß1 receptors are also identified in epithelial and stromal cells. While affecting epithelial cells TGF-ß1 inhibits proliferation and stimulates the differentiation of basal cells into luminal ones but while affecting stromal cells TGF-ß1 stimulates their aggregation. TGF-ß1 is a key factor in stimulating apoptosis in which protease family has the leading role [8].

The basic fibroblast growth factor (bFGF) affects through paracrine/autocrine mechanism through high-affinity transmembrane receptors [9], thus displaying angiogenic and hematopoetic properties and stimulating many cells system proliferation. It has been established that bFGF stimulates stromal growth in the pros-

tate [10] and in the case of benign prostatic hyperplasia (BPH) in the gland tissues bFGF concentration exceeds 2-3 times its content in the normal gland tissues and bFGF receptors genes expression increases from 2 up to 8 times [11, 12].

According to the data by D.Giri and M.Ittmann [13] there is a bFGF double increase in the prostate tissues with BPH and bFGF is a powerful factor for both stromal and epithelial cells growth. bFGF concentration radioimmune definition revealed that in the case of BPH bFGF concentration is 2-3 times higher than in the normal prostate of elderly men [14]. Moreover, it has been revealed that there is no essential bFGF concentration difference in different zones and tissues of the prostate such as periurethral zones, stromal and epithelial tissues. However, according to F.Sciarra and co-authors [15] bFGF concentration in the prostate sections are still different: in periurethral zone tissues it is higher than in peripheral subcapsular zone and androgen concentration (dihydrotestosterone, testosterone) undergoes similar changes. A significant increase in genes expression has been revealed for bFGF with BPH in gland tissues compared with the normal prostate tissues, whereas aFGF transcription has not been observed except for one case. Epidermal growth factor (EGF) in the hypertrophic prostate has not been observed, either. TGF-β2 expression, but not TGF-β1 expression, rises whereas aFGF is not practically seen in the hyperplastic prostate tissues. Hence, it can be asserted that bFGF and TGF-β2 are included in the BPH progression mechanism. bFGF and aFGF expressions were immunohistochemically defined in the BPH patient tissues after transurethral prostatectomy. It turned out that aFGF expression is absent both in basal epithelial cells and in the stromal compartment while bFGF expression was observed in cytoplasm of all the epithelial cells except for luminal ones. It must be pointed out that bFGF expression decreases in the regions with moderate epithelial displasia and intensive nuclear cytoplasmic bFGF expression is revealed in smooth muscle stromal cells [16, 17]. Based on the data at our disposal it can be assumed that bFGF is involved in BPH pathogenesis. In this respect bFGF and TGF-ß1 concentration in blood serum of the patients with BPH has been researched.

II. Material and methods.

208 patients with BPH aged 57-81 have been involved in the research at St.Grigor Lusavorich Medical Center in Yerevan in the period of 2000-2006. Urology examination included anamnesis study, residual urine and the prostate volume dimension by transabdominal ultrasonic scanning and urine and blood laboratory examination. Prostate-specific antigen (PSA) was found in the blood of all the involved patients.

Besides, all the patients underwent uroflowmetric examination. This examination revealed urine maximal flow rate (Qmax) and if Qmax was more than 15ml/s it was considered as a normal one. Urination parameters examination was carried out by comparing the uroflowmetric results data with the ones of healthy people.

For BPH symptoms quantity estimation the International Prostate Symptom Score (IPSS) has been used. IPSS was recommended by the First International Consultative Committee on BPH held under WHO aegis in Paris in 1991.

bFGF quantity definition in blood serum of the patients with BPH was assessed by immune analysis method worked out by R&D Systems Inc. (Minneapolis, USA) using Quantikine High Sensitivity Systems. This made it possible to define bFGF quantity both in blood serum and in urine even if its minimal revealed concentration is 0.05 pg/ml.

TGF-ß1 quantity definition in blood serum of the patients with BPH was carried out by immune analysis method worked out by R&D Systems Inc.(Minneapolis, USA) using Quantikine Human TGF-ß1 Immunoassay. This made it possible to define bFGF quantity both in blood serum and in urine even if its minimal revealed concentration is 3 pg/ml.

Blood serum samples received by the method recommended by National Committee for Clinical Laboratory Standards, USA were kept under td"-20°C before the research. Optical density definition was carried out on Multiscan Data Plasma Photometer (Thermo Electron Corporation, United Kingdom). Concentration quantization was done on a microcomputer Toshiba L10 using a WHO specially worked out program on immune analysis (WHO EISA Data Processing Program, version 5.2).

All the received results were compared with bFGF level in blood serum of healthy aged men without BPH (control group).

The results were also statistically processed by Student's method with *t*-criterion and P values less than 0.05 were not considered as a statistically important difference for the compared values.

III. Results and Discussion.

The results of the research made it possible to establish that bFGF concentration in blood serum of healthy young men is 1.61±0.5 pg/ml. These data correspond to the data received by many other authors who established that bFGF concentration in blood serum of healthy men varies among 0.05 to 2.2 pg/ml [18]. It must be pointed out here that bFGF concentration determination in blood serum of healthy men is rather variable [19]. This fact is apparently conditioned by different research methods with different degrees of sensitivity and exactness.

Table 1

BPH patient groups based on bFGF concentration in blood serum

Patient group	bFGF average concentration in blood serum (pg/ml)
I	1.54 ± 0.6 (n = 45)
II	2.8 ± 0.3 (n = 25)
III	4.2 ± 0.5 (n = 26)
IV	5.9 ± 1.6 (n = 35)
V	16.4 ± 4.2 (n = 39)
VI	44.0 ± 10.8 (n = 38)

At present it is a well-known fact that the most exact and high-sensitive method to determine bFGF concentration in biological liquids, including blood serum as well, is the immune-enzyme analysis with the help of Quantikine HS FGF basic immunoassay kit (Rand D Systems Inc., ASA and United Kingdom) (Quantikine HS. Human FGF basic immunoassay for the quantitative determination of human fibroblast growth factor concentration in serum, plasma and urine. R&D Systems Inc., 2005).

BPH was revealed after a complex examination of the patients and bFGF concentration in blood serum varies at a rather large range among 0.9 and 68.9 pg/ml and the patients were formed into 6 groups according to bFGF concentration in blood serum (see Table 1).

It is worth mentioning that not all the patients with BPH have bFGF higher level in blood serum (Group 1) compared with the control group (healthy men). The second group patients have a high bFGF level in blood serum which is equal to 73.9% ($p \leq 0.001$). The third group patients have 160% more with $p \leq 0.001$. The fourth and fifth groups respectively have 4 and 10 times more bFGF level in blood serum with $p \leq 0.001$. And finally the sixth group has bFGF level in blood serum 30 times more with $p \leq 0.001$.

All these data prove that bFGF concentration in blood serum of the patients with BPH is high in most cases: 153 patients which makes 73.5% of the examined patients, 134 patients (64.4%) have bFGF concentration in blood serum exceeding 4.2 pg/ml, an at last 38 patients (18.2%) with BPH have bFGF concentration in blood serum exceeding 40 pg/ml (See Table 2).

Our research results show that the basic TGF-â1 concentration (pg/ml) in blood serum of the patients with

Table 2

TGF-ß1 average concentration in blood serum (pg/ml) of the patients with BPH (concentration decrease in % compared with the control group is given in the brackets)

Patient group	TGF-ß1 average concentration in blood serum (pg/ml)
I	19.1 ± 2.2 (- 32.3% . $p < 0.05$)
II	12.9 ± 1.9 (- 54.2 % . $p < 0.05$)
III	8.2 ± 1.9 (– 70.9 %. $p < 0.001$)
IV	6.2 ± 1.5 (- 78.0 %. $p < 0.001$)
V	4.1 ± 1.9 (-85.5 %. $p < 0.001$)
VI	3.5 ± 1.2 (-87.6 %. $p < 0.001$)

BPH is rather low compared with its concentration in blood serum of healthy men. For example, average TGF-ß1 concentration in blood serum of healthy men is 28.2±3.4 pg/ml and varies from 22.7 to 42.8 pg/ml. However, it is rather lower for the patients with BPH (see Table 2). It is especially significant and statistically reliable that IV, V and VI group patients who have rather high bFGF concentration exceeding 0.05 pg/ml have a low TGF-β1 concentration in blood serum as well and TGF-β1 concentration reduction in blood serum is respectively 78.0%, 85.5% and 87.6% for the mentioned groups.

Special attention must be drawn to the fact that bFGF and TGF-β1 basic concentrations in blood serum of all the six group patients with BPH are essentially different from each other and if TGF-ß1 concentration decreases bFGF concentration increases (Fig. 1).

So, our research results data revealed that bFGF and TGF-ß1 (bFGF / TGF-ß1) level correlation in the blood serum of the patients with BPH is significantly high compared with the same index of healthy men. This takes place because of both the increase of bFGF concentration and TGF-ß1 concentration reduction.

IV. Conclusion. Our research results data prove that bFGF basic level is essentially high and TGF-ß1 basic level is significantly low in blood serum of the patients with BPH compared with healthy men blood serum.

With respect to the above mentioned it must be pointed out that bFGF, i.e. the decrease of the initial level of one of the key stimulators of stromal fibromuscule proliferation in the prostate is accompanied by a significant fall of ß1 transforming GF (TGF-ß1) which is the proliferation inhibitor and apoptosis inductor.

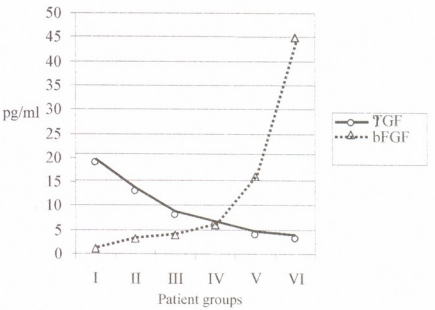


Fig. 1. bFGF and TGF-ß1 basic concentrations in blood serum of the patients with BPH in six groups.

Such a disbalance between the concentrations of the two main factors involved in the stromal proliferation regulation in the prostate is possibly one of the main pathogenetic factors of BPH progression.

Based on the results of our survey it can be assumed that the discovery of peptide regulation role in the prostate growth is a perspective to work out some concepts for BPH antipeptidergic therapy by creating bFGF antagonists.

სამედიცინო მეცნიერებანი

ზრდის ფაქტორები კეთილთვისებიანი პროსტატული ჰიპერპლაზიით დაავადებული პაციენტების სისხლის შრატში

რ. რ. ბოროიანი*

* *წმ. გრიგოლ განმანათლებლის სამედიცინო ცენტრი, ერევანი, სომხეთი*

(წარმოდგენილია აკადემიკოს ლ. მანაგაძის მიერ)

კვლევის მიზანი იყო კეთილთვისებიანი პროსტატული ჰიპერპლაზიით დაავადებული პაციენტების სისხლის შრატში ზრდის ფაქტორების კონცენტრაციის განსაზღვრა. მაღალი მგრძნობელობის იმუნოლოგიური ანალიზით დადგინდა, რომ კეთილთვისებიანი პროსტატული ჰიპერპლაზიით დაავადებულ პაციენტებში

(ასაკი 57-81 წელი) ბაზური ფიბროპლასტის ზრდის ფაქტორის (ბფზფ) კონცენტრაცია უფრო მაღალია, ვიდრე საკონტროლო ჯგუფში (ჯანმრთელი მამაკაცები). ხშირ შემთხვევაში ბფზფ-ის კონცენტრაცია (1,6±10,5) ÷ (44,0±10,8) pg/ml ფარგლებშია, მაშინ როდესაც β-1 გარდაქმნის ფაქტორის კონცენტრაცია დაბალია და იცვლება (19,1±2,2) ÷ (3,5±1,2) pg/ml ფარგლებში.

REFERENCES

1. *N.A. Lopatkin, O.A. Apolikhin* (2004), Urology, **1,** 9-12 (Russian).
2. *C. Lee, A. Cockett, O. Cussenot, et al.* (2001), Regulation of prostate growth. In: Chatelain C., Denis L., Foo JKT. et al. (eds). Fifth International Consultation on Benign Prostatic Hyperplasia, Plymouth, United Kingdom: Health Publications; p. 81-106.
3. *M.T. Story, S.C. Jacobs, R.K. Lawson* (1984), J Urol., **132**(6), 1212-5.
4. *R.K. Lawson, M.T. Story, S.C. Jacobs* (1981), Prog Clin Biol Res., **75A,** 325-36.
5. *J.H. Mydlo, J. Michaeli, W.D. Heston, W.R. Fair* (1988), Prostate, **13**(3), 241-7.
6. *M.T. Story* (1995), World J. Urology, **13**(5), 297-305.
7. *N. Kyprianou* (2003), Urology, **169**(4), 1520-5,.
8. *M. Marcelli, T.C. Shao, X. Li, H. Yin, M. Marani, L. Denner, B. Teng, G.R. Cunningham* (2000), J. Urology, 164(2), 518-25.
9. *J. Faitova* (2004), Cesk Fysiol., **53**(3), 92-101.
10. *C. Saez, A.C. Gonzalez-Baena, M.A. Japon, J. Giraldez, D.I. Segura, J.M. Rodriguez-Vallejo, J. Gonzalez-Esteban, G. Miranda, F. Torrubia* (1999), Prostate, **40**(2), 83-8.
11. *S. Boget, A. Leriche, A. Revol* (2001), Farmaco., **56**(5-7), 467-9.
12. *M.C. Pollan, H.A. Benghuzzi, M. Tucci* (2003), Biomed Sci. Instrum., **39,** 329-34.
13. *D. Giri, M. Ittmann* (2001), American Journal of Pathology, **159,** 139-147.
14. *F.P. Begun, M.T. Story, K.A. Hopp, E. Shapiro, R.K. Lawson* (1995), J. Urol., **153**(3), 839-43.
15. *F. Sciarra, S. Monti, M.V. Adamo, E. Palma, V. Toscano, G. d'Eram, F. di Silverio* (1995), Urol Res., **23**(6), 387-90.
16. *H. Mori, M. Maki, K. Oishi, M. Jaye, K. Igarashi, O. Yoshida, M. Hatanaka* (1990), Prostate, **16**(1), 71-80.
17. *N. Deshmukh, J. Scotson, A.R. Dodson, P.H. Smith, Y. Ke, C.S. Foster* (1997), Br. J. Urol., **80**(6), 869-74.
18. *M. Ii, H. Yoshida, Y. Aramaki, H. Masuya, T. Hada, M. Terada, M. Hatanaka, Y. Ichimori* (1993), Biochem. Biophys. Res. Commun., **193,** 540-545,.
19. *P.I. Hsu, N.H. Chow, K.H. Lai, H.B. Yang, S.H. Chan, X.Z. Lin, J.S. Cheng, J.S. Huang, L.P. Ger, S.M. Huang, M.Y. Yen, Y.F. Yang* (1997), Anticancer Res., **17**(4A), 2803-9.

Received September, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Experimental Medicine

The Association of Blood Pressure Levels with Peritoneal Insulin Absorption in Diabetic Patients on Peritoneal Dialysis

Teona Pantsulaia*, Lavrenti Managadze**

* *A. Tsulukidze National Centre of Urology, Tbilisi*
** *Academy Member, A. Tsulukidze National Centre of Urology, Tbilisi*

ABSTRACT. Intraperitoneal insulin administration is very important for diabetic patients to control blood glucose level and to prevent development of peripheral hyperinsulinemia, which increases the risk of atherosclerosis. Factors affecting intraperitoneal insulin absorption are unclear, and our aim was to study some of the parameters which theoretically can play a role in insulin absorption. One of these parameters is blood pressure, which in itself is associated with the peritoneal transport status. In the results, we give the correlation between insulin absorption and blood pressure; numbers of blood pressure above normal contribute to insulin absorption (p=0.045). This fact is very important in the treatment of diabetic patients, to ensure good control of blood glucose level. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *peritoneal dialysis, membrane transport status, insulin absorption, blood pressure.*

Diabetic nephropathy is the only growing cause of end-stage renal disease (ESRD). Continuous ambulatory peritoneal dialysis (CAPD) has become the preferred mode of dialysis therapy for diabetics with ESRD in many centers [1-3] because CAPD offers the advantages of reduced cardiovascular stress, a steady biochemical state, good control of hypertension, and intraperitoneal administration of insulin [2-5].

Intraperitoneal insulin administration can restore the glucose control to near normal values, better than those obtained with subcutaneous therapy [6].

The degree of hyperinsulinemia is less than with subcutaneous administration. At a given dose of insulin, the amount reaching the periphery is much less in intraperitoneal than in subcutaneous administration. This is especially important because the circulating insulin is directly correlated with the risk of atherosclerosis [7].

Intraperitoneally administered insulin is absorbed more rapidly and evenly than subcutaneously administered insulin. It passes directly into the portal vein system. From the liver, it influences glucose and lipid metabolism [8].

If insulin is instilled into the abdominal cavity along with the dialysis solution, switching from subcutaneous to intraperitoneal administration entails an increase of insulin requirement by approximately 30% [9].

Higher insulin requirements with intraperitoneal delivery, compared to subcutaneous administration, during CAPD might be due to several factors: incomplete peritoneal absorption of insulin, which is concentration- and time-dependent; possible intraperitoneal degradation of insulin by insulinase enzymes; degradation within omental adipocites; and adsorption of insulin to the surface of fluid containers and connecting tubes.

Factors affecting this insulin absorption are unclear, but membrane transport characteristics could be important. During peritonitis, when membrane characteristics change to high transport status, peritoneal insulin requirements fall, suggesting that insulin peritoneal transport is related to membrane transport status. According

Table 1.

Correlation of intraperitoneal insulin absorption with blood pressure levels

Patient group	Blood pressure levels	Insulin absorption	(p)
group I	<120/70 mmHg (1) (n=2)	91.7% ± 11.5%	p_{1-2} = 0.24 (NS)
group II	120/70-140/90 mmHg (2) (n=6)	97.9% ± 2.5%	p_{1-3} = 0.29 (NS)
group III	>140/90 mmHg (3) (n=2)	99.4% ± 0.2%	p_{2-3}= 0.35 (NS)

NS- statistically unreliable

some authors data, the blood pressure level is associated with membrane transport status [10].

Materials and methods. We have studied 10 diabetic (4-Type I, 6-Type II) patients on peritoneal dialysis. There were 7 male and 3 female patients with mean age 42.8±14.0 years.

Duration of diabetes 13-34 years (mean 21.5±1.85 years); 24-hour insulin requirements was 14-56U (mean 34.9±5.9U); and mean HbA_1c was 7.44± 0.5%. None of the patients had peritonitis in the previous 2 months and had no temperature.

Peritoneal membrane solute transport in peritoneal dialysis (PD) patients is assessed by the peritoneal equilibration test (PET), which measures the ratio of creatinine in the dialysate to plasma after a standardized 4-h dwell (D/P_c). Patients then are classified as high (H), high-average (HA), low-average (LA), or low (L) transporters on the basis of this result [8]. In our study we have patients with H, HA and LA membrane transport status. The patient's dose of Actrapid was injected into plastic bags. Dialysate insulin was analyzed using the enzyme-linked immunosorbent assay (ELISA) (DakoCitomation. Denmark) before and after test. Blood pressure was measured before beginning the test.

Results and their discussion. Variation in peritoneal insulin absorption was observed (mean of 90.34%±6.65%). A relationship was found between membrane transport status and percent of intraperitoneal insulin absorption (LA-95.8%±1.3%; H-99.7%; p=0.011).

To determine the role of blood pressure levels in intraperitoneal insulin absorption, in those diabetic patients who were on peritoneal dialysis, we measured blood pressure during the procedures. According to blood pressure levels patients were divided into three groups: low-<120/70 mmHg, normal – 120/70-140/90 mmHg, high->140/90 mmHg (Table 1).

In our results insulin absorption was high, when the level of blood pressure was above 140/90 mmHg, and the same parameter was low when the blood pressure level was 100/70 mmHg. These results were statistically unreliable, there was no relationship between the percentage of intraperitoneal insulin absorbed and blood pressure levels (p_{1-3}=0.29).

To account for these results and small number of patients, we decided to divide them into two groups. (group I - T/A-d"120/70 mmHg, group II -T/A>120.70 mmHg). This attempt was successful, showing a correlation between intraperitoneal insulin absorption and blood pressure levels. This result was statistically significant (p=0.045). (Table 2).

Table 2.

Correlation of intraperitoneal insulin absorption with blood pressure levels

Patient group	Blood pressure levels	Insulin absorption
group I	≤120/70 mmHg (n=6)	95.5% ± 3.0%
group II	>120/70 mmHg (n=4)	99.5% ± 0.3%
p=0.045		

In conclusion, intraperitoneal insulin absorption, as it was expected, is related with membrane transport status. Also, our results demonstrate that high blood pressure level, which is frequent in diabetic patients who are on peritoneal dialysis, contributes to intraperitoneal insulin absorption, probably due to increased peritoneal blood flow. These facts are very important in treatment of diabetic patients, to give them the proper dose of insulin, and ensure good control of blood glucose level.

სამედიცინო მეცნიერებანი

სისხლის არტერიული წნევის კავშირი ინტრაპერიტონეალურად შეყვანილი ინსულინის აბსორბციასთან, პერიტონეალურ დიალიზზე მყოფ დიაბეტიან პაციენტებში

თ. ფანცულაია*, ლ. მანაგაძე**

* *ა. წულუკიძის უროლოგიის ეროვნული ცენტრი, თბილისი*
** *აკადემიის წევრი, ა. წულუკიძის უროლოგიის ეროვნული ცენტრი, თბილისი*

პერიტონეალურ დიალიზზე მყოფ დიაბეტიანი პაციენტებისთვის ინსულინის ინტრაპერიტონეალური დანიშვნა იძლევა გლიკემიის ოპტიმალური კონტროლის საშუალებას და შესაძლებლობას გვაძლევს თავიდან ავიცილოთ პერიფერიალური ჰიპერინსულინემიის განვითარება, რომელიც ათეროსკლეროზის განვითარების წინაპირობაა. ფაქტორები, რომლებიც ინტრაპერიტონეალური ინსულინის აბსორბციის პროცესში მონაწილეობენ, ბოლომდე დადგენილი არ არის. სწორედ ამ მიზნით შესწავლილ იქნა ის ფაქტორები, რომლებიც, თეორიულად, შესაძლებელია გავლენას ახდენდნენ ინტრაპერიტონეალურად შეყვანილი ინსულინის აბსორბციის პროცესზე. ერთ-ერთ ასეთ ფაქტორს წარმოადგენს სისხლის არტერიული წნევა, რომელიც თავის მხრივ პერიტონეუმის ტრანსპორტის ტიპს განაპირობებს. როგორც აღმოჩნდა, სისხლის არტერიული წნევის ნორმაზე მაღალი რიცხვები ზრდის ინტრაპერიტონეალურად შეყვანილი ინსულინის აბსორბციის პროცენტულ მაჩვენებელს, რაც დადასტურდა სტატისტიკური ანალიზით ($p=0.045$) და მნიშვნელოვანია შაქრიანი დიაბეტით დაავადებული პაციენტების მკურნალობის პროცესში ინსულინის დოზის სწორად შერჩევისთვის.

REFERENCES

1. *C.M. Kang, B.K. Bang, H.B. Lee, et al.* (1987), Korean J. Nephrol. 6:S56-S64.
2. *J. Rottembourg* (1988), In: Nolph KD, ed. Peritoneal dialysis, 3rd ed. Dordrecht: Kluwer Academic Publishers; 365-379.
3. *R. Khanna, D.G. Oreopoulos* (1989), J Diabetic Compl. 3:12-17.
4. *J. Passlick, B. Grabensee* (1988), Clin Nephrol. 30:S18-S23.
5. *J.A. Diaz-Buxo* (1986), In: R Khanna, KD Nolph, B Prowant, ZJ Twardowski, DG Oreopoulos, eds. *Advances in continuous ambulatory peritoneal dialysis.* Toronto: Peritoneal Dialysis Bulletin, Inc.; 31-38.
6. *L. Scarpioni, S. Ballocchi, A. Castelli, M. Cecchettin, F. Fontana* (1988), Insulin therapy in diabetic patients on CAPD. Peritoneal dialysis. Milano, 119-24.
7. *E. Quellhorst* (2002), Semin dial, Vol 15, Issue 5: 376-376.
8. *E. Quellhorst* (2002), J Am Soc Nephrol 13:S92-S96.
9. *A. Fine, D. Parry, R. Ariano, and W. Dent* (2000), Peritoneal Dialysis International, **20**: 652-655.
10. *J.J. Lin, N.K. Wadhwa, Suh H, T. Carbalda, C.S. Patlak* (1995), Adv Perit Dial. 11:63-6.

Received January, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Humanities & Social Sciences

History & Philology

The Indo-European Homeland in the Near East: New Evidence

Vyacheslav V. Ivanov*

* *University of California, Los Angeles, USA*

ABSTRACT. The problem of the initial place from which the original Indo- European dialects spread over the West part of Eurasia has been studied by several generations of scholars. Few alternative points of view have been proposed: first an area near the North Sea (in the works of some scholars of the turn of the XIX and XX centuries), then the North coast of the Black Sea (an old idea of Schrader revived by Maria Gimbutas and her followers[1]).

35 years ago the author of the present text together with Tamaz Gamkrelidze suggested first in a talk at a conference, then in a series of articles and in a resulting book that the South-East part of Anatolia, close to North-East Syria and North area of Mesopotamia, may be considered as a possible candidate for the Indo-European homeland, Gamkrelidze, Ivanov 1972; 1995, 1990; see map 1 of the possible migrations suggested in the latter publication and reproduced below.

Since that time many linguists, archeologists and specialists in the other fields of studies bearing on the solution of this question have been discussing the arguments for and against this suggestion. Recent research on these topics has brought additional evidence that seems to prove the Near Eastern hypothesis in a definite way. The article sums up the results achieved in the last decades.

Key words: *Proto-Indo-Europian, Indo-Europian homeland.*

1. Contacts of Proto-Indo-European with other linguistic families. Indo-European elements in Kartvelian. Indirect evidence on the early presence of Indo-Europeans in the areas close to the Near East can be found in the traces of the ancient contacts between linguistic families in this part of Eurasia. Such contacts between Proto-Indo-European and Proto-Kartvelian have been suggested in the work of T.Gamkrelidze and G. Mach'avariani more than 40 years ago. The following studies have established a number of important loanwords from Proto-Indo-European in Proto-Kartvelian. Particularly interesting discoveries in this field were made by the late G.A.Klimov[2]. He has found many new common elements of the two families in addition to a relatively long list in our joint work with Gamkrelidze in which we included also the correspondences earlier noticed by Klimov. The main difficulty in interpreting the results of his investigations is connected to the problem of a possible common Nostratic origin both of Proto-Indo-European and of Proto-Kartvelian. If these two linguistic families were originally cognate, then some part of the correspondences found by Klimov and other scholars might have been traced back to the Proto-Nostratic early period[3] (more

[1] See the chapter on Place and Time in the new textbook: Mallory- Adams 2006, pp. 86-105.

[2] Klimov 1985-1998; Klimov, Khalilov 2003; cf. also Harris 1990, Greppin 1997.

[3] The question was put forward already in : Shevoroshkin 1986.

Map 1. The Indo-European homeland and proposed migrations (after Gamkrelidze and Ivanov 1990).

than 10 000 years ago). Only those words that were not inherited from this ancient time are important as a proof of the later existence of Proto-Indo-European in the area close to the Proto-Kartvelian (to the South-West of the Transcaucasian area in which the latter spread in the historic time, Klimov 1998, pp. IX, XII). In a recent work of the late S.A. Starostin who has tried to select Indo-European elements in Kartvelian distinguishing them from the possible common Nostratic core several definite comparisons have been singled out such as: Proto-Kartvelian **ank'es-*'fishhook' (Georgian *ank'es-i* borrowed in the other Kartvelian languages, Klimov, Khalilov 2003, p. 146) : Indo-European dialectal **ankos-*[4] (Avestan *aka* 'a hook to bind a horse to a vehicle'[5], Sanskrit *aṇká-, aṇkuśa-* 'hook", Greek ἀγκύλος 'crooked, bent', Latin *uncus* 'crook, crooked', Old English *angul* 'fishhook'); Proto-Kartvelian **tel-* 'young pig': Proto- Indo-European **tel-* 'young animal' (this correspondence is supported by another one discovered much earlier: dialectal Kartvelian ** γor* 'pig' has been borrowed from the prototype of the dialectal Indo-European **ĝhoryo-* 'pig', Gamkrelidze, Ivanov 1995); Proto-Kartvelian **usxo-* 'sacrificial ox': Proto-Indo-European **uk^{w}so-* 'ox' (Klimov 1994a; 1998, pp.195-196) and other words[6]. These borrowings point to certain spheres of the economy and farming technology (cf. such clear examples as, for instance, : Kartvelian **berg-* 'to hoe, to crush the earth': dialectal Indo-European **bh(e)r-g-*[=**bherk'-*], Klimov 1994, pp. 49-50; 1998, p. 11) that were more developed in the society from the language of which the nouns had been borrowed.

2. Proto-Indo-European and Proto-North Caucasian. In our book published in 1984 we suggested some common terms shared by these languages explaining them as possible traces of later Indo-European (probably Indo-Iranian) migrations through the Caucasus. The study of this problem has been enriched through the recent research on Proto-North Caucasian. S.L. Nikolaev and S.A. Starostin have compiled a large etymological dictionary of this family (Nikolaev, Starostin 1994) developing the comparative studies started by the Prince N.S.Trubetzkoy[7]. Starostin has gathered a large collection of the terms of material culture common to North Caucasian and Indo-European (Starostin 2007, pp. 310, 312-358, 818). They include many names of domestic animals (Proto North Caucasian **hīnčwe* 'horse'[8]; **?ējžwe*

[4] In Kartvelian this relatively recent borrowing is reflected without a laryngeal which is seen in the other probable Nostratic correspondences of Indo-European stems. If one does not accept the Nostratic hypothesis then it is still necessary to distinguish 2 different strata of Indo-European borrowings: without laryngeal correspondences as in the cited term for 'crook' and with them, cf. forms like Kartvelian ** γweb-* 'to weave, wattle' : Indo-European **Huebh-*, Kartvelian **γwed-*'strap', in Old Georgian and in the modern Dzhavax dialect 'a tie on the yoke, belt' : Proto-Indo-European **[H]wedh* – (Klimov 1994, pp. 74-75; 1994b; 1998, p. 225).

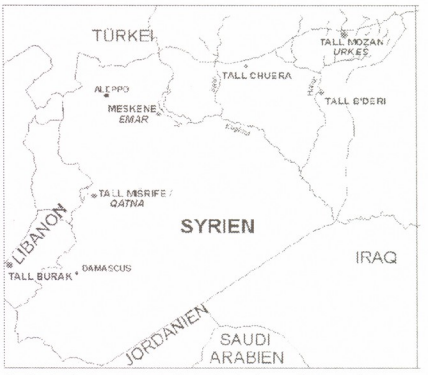
[5] The archaic meaning preserved in Old Iranian might help to understand the reasons for the borrowing of a technical term.

[6] Starostin 2007, pp. 818-819, cf. on other (mostly older) common elements in Indo-European and Karvelian also ib., pp. 807-815.

[7] Trubetzkoy 1987 (with important comments on North Caucasian by S.A.Starostin, pp.437-447; 453-465).

[8] There are at least two more North Caucasian words for *horse* that might be compared to dialectal Indo-European terms (Ivanov 2002). The Indo-European term for 'foal' **polo-s* is found only in some dialects but still belongs to a relatively old part of technical vocabulary since it is attested in Mycenaean

'goat'; *wHārʌʌwɔ'pig'; *päHäʌwɔ'[small-]cattle' corresponding to Proto-Indo-European *(H)ek'wo-'horse'; *(H)aig'- [=*Haik'-]'goat'; *pork'o- '(young) pig'; *pek'u- 'cattle') and parts of the body or products of cattle-breeding (Proto-North East Caucasian *Awāhni 'wool': Indo-European *Hu(ə)l(ə)nā), plants (Proto-North East Caucasian *?qV 'grape, fruit'; *λwin?ī 'seed, flax seed', *HVbVgV/*HVgVbV 'a kind of cereals> oats, rye' : Indo-European *(H)ag-= [*Hak'-] 'berry'; *līno-'flax', *(H)auig-[Hauik'-] 'oats'; a large number of similar names of trees like Proto-North East Caucasian *χwīrk(w)V 'oak': Indo-European *perkʷo- may be interpreted as pointing to an environment that was similar from the botanical point of view); instruments (Proto-North East Caucasian *χIwērV 'mill, grind-stone': Indo-European *gʷerən[=*kʷerən] 'grind-stone'; Proto-North Caucasian *ničo 'knife, sickle': Indo-European *ŋsi-'sword' and many other terms). In a special work on this subject Starostin suggested that all these terms were borrowed in the beginning of the Vth mil. B.C. probably in the area of the Near East to the South of Transcaucasus (Starostin 2007, pp. 357-358). The area seems to be the only possible one. Although we still use the traditional term "North Caucasian" it is not geographically correct even if applied to such living languages as Abkhaz and to the dead Ubykh (spoken originally at the southern part of the South-West Transcaucasian area). Recently it has been established that the Eastern branch of North Caucasian included Hurrian and Urartian (Diakonoff, Starostin 1986; Starostin 2007, pp. 359-406, 629-632, 745-751; Ivanov 2002). In the I mil. BC Urartian was one of the main languages of the Eastern Anatolia. Hurrian was spoken in the Northern Syria as early as the last quarter of the I I I mil.BC in Urkeš (Tell Mozan), see map 2.



Die aktuellen Grabungsprojekte des Altorientalischen Seminars Tübingen in Syrien

Map 2. Tell Mozhan (Urkeš) and the spread of Hurrian in Syria.

According to Giorgio Buccellati who has excavated the city, the Hurrian tradition there goes back earlier than to the middle of the I I I mil. BC. (in Hurrian mythological texts it is a city of the main ancient god). As it is proved by the recent archaeological discoveries, later on Hurrian had been spoken in the same area. It spread much farther to the South in Nuzi, Mari and Arrapha and to the East- in Emar and Qatna (from the recently discovered Qatna tablets one may learn about the end of the Hurrian empire of Mitanni; the tablets were written in Akkadian with a lot of Hurrian forms marked by a *Glossenkeil*). Hurrian diffused also to the southern part of Anatolia (Kizzuwatna); already in the beginning of the II mil.BC Hurrians were present in Kanish (in the centre of Asia Minor). The contacts of speakers of Hurrian with Indo-Iranians had started (earlier than the well-known data on Mesopotamian Aryan in Mitanni[9]) by the beginning of the I I mil.BC: to this time the Hurrian part of a mythological bilingual text can be attributed (Neu 1996) in which the Hittite Ablative form *pa-aḫ-ḫu-e-na-an-za* 'with the fire' is a translation of the Hurrian Ergative *ta-a-re-eš* (from the Iranian *ātar*, Ivanov 2002). A set of correspondences proves that Hattic that was a sacred language of the Hittite empire also belonged

Greek (*po-ro*, Homeric πώλος 'foal'), Albanian *pēlë* 'mare' and in all the ancient Germanic languages: Gothic *fula* 'foal, colt' a. o. (the relation to Hittite *pulla*-'young, son' is not clear) . It seems possible to compare it to the North Caucasian *farnê reconstructed on the base of Avar xʷárni 'horses' (collective form); Khvarshi χaram 'foal'; Lezghian χʷar 'mare'; West Caucasian *χʷara> Adygh *fára* 'thoroughbred horse' (Nikolayev, Starostin 1994, pp. 425-426). The North Caucasian word belongs to a very small class of lexical items beginning with *f-*. From 7 words included in this class in the dictionary by Nikolayev and Starostin two other lexemes (the numeral 'five' and the noun 'fist' related to it) also are shared with Indo-European, Starostin 2007, p. 321, 2.11. The connection between the North Caucasian and Indo-European terms of this class seems beyond doubt. There is a possibility of a distant (and accordingly very old) relation between a North Caucasian group of words and the Hittite *kurka-* 'foal' compared to Greek κῦρνος (Forssman 1980; Melchert 1994, p.132; it is not easy to conclude whether the words should also be compared to Iranian terms like Persian *kurra* 'foal', since the latter is connected to the onomatopoetic appeal to address horses: cf. Vakhan *kurr-kurr* 'Come!' in this function and *kurrás(t)* 'neighing of a horse', Steblin-Kamenskij 1999, p.213; Paria *kurro* 'horse'). The words might represent a common borrowing from the North Caucasian *gwälV 'horse'>Nakh *gile* 'horse, steed' >Chechen *gila*, Proto-Tzez-Khvarshi *guRu> Tzez *gulu* 'stallion, horse'. The root could have been the source of both Greek and Hittite words with a characteristic change of voiced into voiceless consonants depending on the rules of the Hittite phonology. Another borrowing from the same source might be Latin *caballus* 'labouring horse' that has ousted the earlier general term in Romance. This word and such dialectal terms related to it as Albanian *kâl* 'horse' might have been borrowed from this North East Caucasian source.

[9] Recent additions to the known list of these words confirming the especial link to Old Indian are discussed in; Mayerhofer 1996 (on *wadura-nni*).

to the North Caucasian family, probably to its Western branch (Ivanov 1985). In the II mil. BC the language was dead, as in the Hittite archives of Hattušaš and Šapinuwa (Ortaköy)[10] it is found only in the ritual texts (often with Hittite translations). Most of the places of the traditional Hattic cults (like Nerik, Arinna, Zippalanda, Karahna, Tawiniya) can be located in the North Anatolian area. But it is possible that Hattic was spoken also more to the South since it had several prehistoric Semitic loan-words (as *binu* son', *milup* 'ox', *zinir* 'lyre'[11]). Since in the III mil. BC both Hurro-Urartian and Hattic were spoken in the regions to the South of Transcaucasia it becomes possible to reconstruct a homeland of the whole family (which at that time was not "North" Caucasian) to which they belonged in the same area close to the supposed Proto-Indo-European and Proto-Kartvelian homelands.

The fricative *š* in the Hurrian name of horse *ešš ə* and an affricate **č* (>*š*) in the forms of the other North Caucasian dialects correspond to a Proto-Indo-European palatal stop **k'* that has become an affricate ** č* and then a fricative *š /s* in the Indo-European dialects of the *satəm* type. Similar changes are present in the other borrowings discussed by Starostin (2007, p. 310, n. 28; pp.339-358). He supposed that the common words discovered by him were mostly borrowed from Proto-North Caucasian (or from a dialect of it) into Proto-Indo-European. The opposite direction of borrowing can be suggested due to the typologically valid laws of sound change. From the point of view of such general typology a phonetic development of a palatal or palatalized velar stop to an affricate and a fricative is a normal one; but the reverse movement from an affricate of dental type to a velar stop seems quite extraordinary. If Starostin is right and there was a system of regular correspondences at an ancient period it might be that a North Caucasian affricate absent in Proto-Indo-European might have been reinterpreted as an old palatal. But a later dialect of a *satəm* type that should have existed in the III mil. B.C. (and maybe even earlier) might have borrowed forms with affricates and/or fricatives from a language having a similar system of consonants. In that case the direction of borrowing may be only from a Proto-Indo-European dialect of a *satəm* type into a North Caucasian dialect since otherwise the existence of a parallel Proto-Indo-European dialectal *centum* form would have remained mysterious. But no matter which direction of the borrowing should be chosen, the fact of the existence of these loanwords is without doubt. They make the decision on the place of the Indo-European homeland a definite one.

3. Indo-European and (West) Semitic. In our monograph written together with Gamkrelidze we suggested that several words shared by these languages (such as the ancient name of the wine, Hittite *wiyana-*) can be considered to be borrowed from Proto-Indo-European (different from the rest of the most ancient common words usually correctly described as old Semitic or Afro-Asiatic loanwords in Proto-Indo-European, Starostin 2007, p. 817). Accepting this idea S.A.Starostin (in an article written some years ago, but published posthumously quite recently, Starostin 2007, pp. 821-826) suggested that a large number of Semitic words (particularly of those characteristic of West Semitic) that did not have correspondences in the other Afro-Asiatic languages had been borrowed from Proto-Indo-European. He came to the conclusion: "the original Indo-European (Indo-Hittite) homeland was somewhere to the North of the Fertile Crescent from where the descendants of Indo-Hittites could have moved in two directions (starting with early 5th millennium BC) to the South where they came into the contact with the Semites, and indeed could have driven a part of them further to the South, and to the North (North-East) whence they ultimately spread both to Europe and to India" (ib., p. 825-826). In discussing the origin of the Indo-European borrowings in Semitic Starostin remarked that they should have been very early because they were made before the loss of laryngeals. He concluded : "we may deal here either with the loans from proto-Anatolian or from an extinguished branch of early Proto-Hittite " (ib., p. 835). As an example one may analyze the Indo-European word for the 'earth' **dheĝhom*> Hittite *tekan,* Genitive *tagnaš*[12], Tocharian A *tkaṃ*. The word that entered into such Anatolian names of the gods as Hittite *Dagan-zipa* 'the Demon of the Earth" (literally "the Earth personified") probably had been borrowed into (West) Semitic where it is used as a name of the god *Dagan*[13]. In

[10] Süel, Soysal 2007; on the use of Hattic in other places see Soysal 2004.

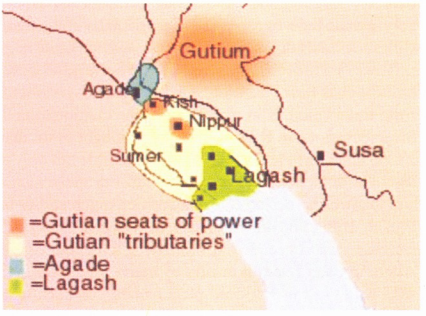
[11] In Hattic the original initial phoneme of the West Semitic **kinor* has been palatalized as also in the Adygh languages and in Old Armenian (Ivanov 1999). In Hattic *m-ilip* 'ox' a prefix *m-* is present which is known in a series of the ancient Semitic terms for animals.

[12] As in Cuneiform Luwian the voiced palatal aspirate **ĝh* disappeared and the word changed to *tiya-*. Hieroglyphic Luwian used mostly in the documents of the Syro-Anatolian princes who would like to pose themselves as those who continue the imperial tradition, had a tendency to imitate the Hittite shape of some important (official, particularly sacred) words; hence in this language the form *takam-i* built contrary to the rules of the Luwian historic phonology.

[13] A hypothesis on this borrowing was put forward by several scholars almost simultaneously and independently of each other, cf. Singer 2000; Ivanov 2004, p. 66 a.o. To the (W) Semitic words that were supposed to be connected to the name of Dagan (cf. on the etymology Renfroe 1992. pp. 91–94). the Semitic name of a 'fish' **dag-* belonged that did not have Afro-Asiatic correspondences and thus also might be borrowed from the Proto-Indo-European **dhĝhu-* (Starostin 2007, p. 823, N 575; on the correspondences see also ib., p.762 [72]). The word had a similar "accessive" structure of the combination of stops ("Brugmann's fricative', see Gamkrelidze, Ivanov 1995, cf. Ivanov 2007, pp. 344-366). A semantic link of the (god of) the earth and a fish might be based on a mythological view of fishes as animals of a lower subterranean world.

Ebla the god was the main one. In an Eblaite text TM 70 75. G.1560 v.VI 14 VI 3 a phrase ^{D}BE *Ḫa-šu-wa-an*^{KI} 'the (divine) Lord (a logogram based on *bēlum* 'lord' and often designating a god of the city) of *Ḫ*ašuwa'[14] occurs which may refer either to Dagan (who is very often designated in this way in Ebla and other neighboring cities as Tuttul) or to another god. The toponym *Ḫaš/zuwa* is identical to the Hittite noun *ḫaššu-* 'king' (cognate to several kinship terms in Hittite and Luwian and to the Hieroglyphic Luwian *hasu* 'family'=Phoenician *šrš* 'roots' in the Karatepe bilingual), the Anatolian city name *Ḫaš(š)uwa* (Tänberg 1994., SS. 54, 70) ='of the King', to the second part of the Old Hittite compound city name *Šal-a-ḫšuwa*(='of the great –King') and of the native Anatolian names of the cities *Ša-la-aḫ-šu-a*[15] and *Ḫa-ra-aḫ-su-wa*[16] in the Old Assyrian tablets from the cities in Asia Minor. From the language of such an Anatolian city a name of an Anatolian god might have been borrowed. It is worth noticing that the same Indo-European term for the 'earth' was borrowed in Proto-Kartvelian where it is represented by words with the meaning of the 'fertile (black) soil' (Klimov 1998, p.41). The interference of the early dialects of Proto-Indo-European, Proto-Semitic and of Proto-Kartvelian to which the early Proto-"North" Caucasian can be added might have led to a formation of a sort of a linguistic zone (*Sprachbund*) that shared not only many words related to new farming economy (Starostin 2007, pp. 256-264, 289-358, 806-826), but had also several phonological and grammatical features in common (Gamkrelidze, Ivanov 1995; Ivanov 2004).

4. Probable Indo-Europeans in the Near East; Henning's hypothesis on Gutians. After we had published our hypothesis on the Near Eastern homeland of the Indo-Europeans several scholars asked us why at the time when writing already existed there were no written documents testifying to the presence of Indo-Europeans in these areas. It seems that now there are several possible answers to the question.The great specialist on Iranian Henning who had worked for many years on the problem of the name of Tocharians in his posthumous article suggested that the early ancestors of Tocharians were Gutians who had invaded Mesopotamia in ca. 2350-2200 BC. In an article written after we finished our book we have developed Henning's idea (based mainly on the etymological links of Near Eastern *Guti* and *Tukri* and Central Asian names of corresponding Indo-European *Kuchean* and *Tocharian* ethnic groups[17]) paying attention also to the possible explanation of some names of Gutian kings preserved in Sumerian texts (Gamkrelidze, Ivanov 1990-1991). Gutians ruled a part of Mesopotamia, see the map 3 below.



Map 3. Mesopotamia at the Gutian period (after the Ancient Near East maps of the Oriental Institute of Chicago)

Probably the further study of those ancient toponyms and hydronyms that might be connected to Gutians may add a necessary material for their definite linguistic identification.

5. "Euphratic": a probable Indo-European dialect reflected in the early cuneiform pre-Sumerian tradition. Recently it has been suggested that an unknown "Pre-Sumerian" language that had been reconstructed on the base of the special phonetic values of many cuneiform signs having several possible phonetic readings was an archaic "Euphratic" Indo-European dialect (Wittaker 1998; cf. Frayne 1992). Its time should be at the second half of the IV mil. BC.: the place is Southern Mesopotamia. According to this hypothesis the phonetic

[14] Pomponio and Xella 1997, p.97 (with a suggestion that the god of the Thunder was meant here; cf. on the god Dagan in Ebla itself ib., pp. 376-377). On Dagan in Syria see Feliu 2003; Crowell 2001. The role of the Indo-European linguistic elements in the texts of Ebla supposed by such scholars as I.Gelb still is not clear; one may cite such personal names as *Zida* (comparable to Luwian forms), but their etymology is controversial. In the correspondence of the Eblaite administration there are letters of the rulers of the neighboring cities with possibly Indo-European names (one of them probably related to Sumerian and Akkadian terms for the 'horse' of an Indo-European origin, Ivanov 2004, pp.50-51).

[15] In the Old Assyrian text of the Golenishcheff's collection GMMI 1554, Yankovskaya 1968, p.165, n 64, line 16, cf. her comments on pp. 27, 166, 225.

[16] Kennedy, Garelli 1960, pp.7-8. In these ancient compounds the root vowel of the second noun had been reduced: *hasu>hsu.* On the Indo-European etymology of this stem comparable to Sanskrit *asu-ra,* Germanic Runic *ansuR* see Gamkrelidze, Ivanov 1995.

[17] See a favorable view of Henning's hypothesis: Thomas 1985, SS.14-17. In his 1965 paper on Tocharians as "The first Indo-Europeans in history" Henning suggested that the ancient pronunciation of the Chinese designation of "Yuezhi" (=Kucheans=Tocharians) could be approximately reconstructed as **Gu(t)-t'i* and possibly related to the ancient name of 'Guti'. Since another probable reconstruction of the Chinese ethnonym 月氏 may be **ngi wāt-tjĕg* that part of Henning's theory may be controversial: Mallory and Mair 2000, pp. 281-2; see further references in: Hill 2004; Liu 2001.

values of approximately one hundred of the early signs that are different from the Sumerian ones go back to the Euphratic words. Some of the etymologies suggested by Wittaker seem plausible as "Euphratic" *hurin* 'eagle' (attested already in the Sumerian lexical list A 11/620 in Ebla) identical to the Hittite *hara-n-*, Ancient Greek ὄρνις[18]. Some grammatical features that may be reconstructed for "Euphratic" nominal stems (such as the use of a final suffix *-t* comparable to the ending of Neuter gender in Lydian and in some other archaic Indo-European forms and the role of the adjectival suffix *–u-* similar to an Anatolian-Baltic isogloss, Puhvel 1982; Erhart 1995; Gusmani 1968) are supported by ancient parallels and seem to speak in favor of this hypothesis.

6. Early Anatolian Indo-European presence in Asia Minor. A large number of Anatolian personal names (of a very archaic Indo-European type, particularly compounds with such final ancient terms as *–nika-* = Old Hittite *nega* 'sister'[19]) studied by Goetze, Laroche and other hittitologists have been found in the Old Assyrian texts from trade colonies in Asia Minor. The continuation of the excavations in Kanish that have brought more than 23000 cuneiform tablets has made it possible not only to discover in them many Anatolian Indo-European names and loanwords (besides those identified much earlier by Balkan, Bilgiç and other scholars, cf. Tischler 1995). J.G. Dercksen has succeeded in identifying a number of Old Assyrian texts relating mostly to native Anatolian persons with typical Indo-European names[20]. In such texts several important Indo-European social terms have been found recently (Dercksen 2004a-2007): Hittite *tuzzi-* 'army' (Indo-European **teu-t-*>Lithuanian *taũta* 'people', Old High German *diota*, Umbrian *totam* 'civitatem', Old Irish *tuath,* the W Indo-European and Balto-Slavic name of 'community, people, land'); Old Assyrian *ubadinnum* "land allotment" : Luwian *ubati-* 'land grant' *(*from the verb *uba-*), Lycian *ube-* 'to offer', Carian *ýbt* (the original meaning 'donation', Adiego 2007, pp. 347, 492)*;* Tocharian B *wepe* 'corral, paddock= Buddhist Hybrid Sanskrit *gocara*" (cf. Indo-European *[*H*]*we-bh-/-dh-* " to weave", metaphoric use "richness, success", Old English *ead,* Gamkrelidze, Ivanov 1995*).* Palmer's old reconstruction of the Indo-European feudal system based on Mycenean data seems to be proved by these new discoveries. A new interpretation of the whole Indo-European socio-economic and military structure may be reinterpreted due to these etymologies. The Anatolian term designating a person who is free of taxes (Old Assyrian name *Arawa*= Hittite *arawa-* 'free from taxes' with an exact Lycian correspondence, cf. Lithuanian *arvas* 'free' and cognate Balto-Slavic terms) refers to neutralization of the semantic opposition of those who can bear weapons and may receive a portion of land according to a general feudal scheme and of those who do not participate in the land-for-the-service system of the relations.

The Old Assyrian documents in Kanish are found in the archaeological levels II and Ib dated ca. 1950–1723 BC (on the base of the recently found lists of eponyms); they precede Old Hittite texts for ca. 250 years. At that time the two Anatolian groups of dialects – a Northern (Hittite) one and a Southern (Luwian) complex – were already completely different from one another[21]. According to dendrochronology that helps to make dates more exact the Old Assyrian colonies existed no less than 2 centuries earlier – in XXII-XXI c. BC (Newton, Kuniholm 2004). No documents from the preceding levels III and IV have been discovered, but it appears that in the half-legendary stories of the earlier revolt of the native Anatolian rulers against Assyrians in the time after Sargon's victorious march through Asia Minor some elements of the historical truth were present. Such ancient compound Indo-European names of the oldest cities as *Puruš-ḫand-a*[22] that had existed at this early period testify for a possibility of tracing this Indo-European tradition much farther back into the prehistory of Asia Minor.

7. A linguistic reconstruction of Indo-European migrations and data of other sciences. The idea of the Indo-European homeland in the Near East from the very beginning was connected to the discovery of a possible link between the appearance of speakers of Indo-European dialects in Europe and the spread of the new farming technology. This trend of thought has been developed in the archeological works of Sir Colin Renfrew (1987;

[18] From this point of view some correspondences have been reinterpreted that earlier we described (in our book with Gamkrelidze) as cultural terms probably common to Indo-European, Sumerian and other neighboring languages, see on Sumerian *gigira* "chariot' Wittaker 1998, p. 144; Frayne 1992, p. 21.

[19] The word (cognate to Luwian *niya-*,cf. above on *tiya-*) has correspondences in the other Nostratic languages (see on the Altaic kinship term **nek*' Starostin, Dybo, Mudrak 2003, p.968; Dravidian matches are added in Starostins Website <starlin@rinet.ru>; a Uralic one was suggested by Čop) and belongs to the oldest type of the Indo-European kinship terms, cf. Ivanov 2001, p. 45. The compounds of this ancient type are not present in the later texts of the historical Hittite period (there are few exceptions as Hittite kings' names like *Šuppiluli-uma* < *'pure'+'pool', cf. above on compound city names).

[20] See a particularly impressive collection of some of such texts and fragments: Dercksen 2004 b: pp. 156–174.

[21] They may continue two different Indo-European dialects; Ivanov 2001. They might have acquired many common Anatolian features due to a long coexistence in Asia Minor. According to another view all of these languages developed from a Proto-Anatolian language(Melchert 1994).

[22] The second element of the name is identical to Hittite – *handa* in the Old Hittite *men-a-ḫḫanda* 'in front of' (an archaic compound), *ḫant-*'forehead" (with a good Indo-European etymology and correspondences in AfroAsiatic). The first part may be compared to the ancient ethnonym used also as a name of the *Prussian* Baltic Indo-European tribe.

2002[23]). The subsequent attempts to support this hypothetical connection have been continued by comparison to genetic data on the time and space characteristics of mitochondrial DNA and Y chromosome of the European population (see a survey in Rexova et al. 2003, with references). Although the genetic counterpart of Sir Renfrew's ideas seems plausible it is not yet definitely proved partly because of the mixed character of the Near Eastern population resulting from the in-breeding (Weng, Sokal 1995). The farming terms common to Indo-European and other linguistic families discussed above show that the innovations were not restricted to one group of languages and were transmitted and exchanged between different ethnic formations (Ivanov 2004). The area of the interference of these families coincides with the kernel of the rising farming in the Near East. That process of a global (multilingual and multicultural) change led to the diffusion of the results of the Neolithic revolution. The main directions of this diffusion (Sherratt 2006) coincide with the trends of the Indo-European migrations, but the new objects might have been introduced earlier than some of their Indo-European names and the latter might precede the coming of those who coined the terms.

The spread of the Near Eastern innovations in Europe developed at the time around the split of Proto-Indo-European (possibly at the beginning of the V mil.BC[24]), but some elements of the new technology and economy might have penetrated it much earlier. The diffusion took several thousand years and probably was all over the whole Europe already ca. 3550BC. At that time Indo-European migrations were only beginning. The speakers of the dialects of Proto-Indo-European living near the kernel of the technological revolution in Anatolia should have acquired the main results of this development. The influence of this civilization already at the V mil.BC was seen in the Balkan area (Merpert 1988). The road through the Balkans for all the newcomers in Europe seems to be the main one (contrary to the still popular concept of the steppes to the North of the Black Sea as the chief source of the development). The growth of farming economy in Europe became more active with the split of the proto-language and the dispersal of the Indo-Europeans. The astonishing scope and speed of that process was made possible due to use of the domesticated horse and wheeled vehicles. D.W.Anthony insists on the role of the North Kazakh steppes for the horse taming (Anthony, Brown 2000). Since according to hydronymy speakers of early Yenisseyan dialects lived there, the partial similarity of the Proto-Yenisseyan **qus* and the *centum* form of the Indo-European horse name deserves a future study[25]. Its *satəm* form has been borrowed into North Caucasian (including Hurrian, see above), Sumerian and Semitic; at the same time another name used in Altaic (Starostin, Dybo, Mudrak 2003, p. 945-946) has been borrowed by Germanic and Celtic, and a Sino-Tibetan form was ousting the ancient name in Balto-Slavic (Ivanov 1998, 2002). This multiplicity of names agrees with the result of genetic studies showing that there were several places of horse domestication in Eurasia (Vilà e.a. 2001; Jansen e.a. 2002). A similar picture of intensive exchange between several Indo-European and neighboring linguistic groups is seen in connection to the invention and spread of the wheeled vehicles and their technical details (Littauer and Crowell 1996; Gökçek 2006; Ivanov 2002; 2007, p. 546-547). Some of their names are common only for Hittite and Tocharian, the first languages the speakers of which went off the common territory probably using the wheeled vehicles called in Hittite *ḫurki,* in Tocharian by a protoform of A *wärk-änt-* "wheel"; the term has North Caucasian (including early Hurrian reflected in Old Assyrian) parallels (Ivanov 2002; Starostin 2007, pp. 309, 332). As in connection to some other earlier objects of the Neolithic revolution in this case also one can not insist that Indo-Europeans were original inventors, but probably they were skillful in spreading other peoples' innovations. For approximately a thousand years and a half a serious preparatory work on horse domestication and the use of the wheeled vehicles have been continuing in different parts of Eurasia. Then almost suddenly the results are seen. On the border of the I I I and I I mil. BC both these important innovations appear together, usually in a context that makes evident the presence of the Indo-Europeans: the traces of the Near East-type chariots and the ritual use of horse are clear in (probably Ancient Iranian) Margiana (Gonur-tepe) discoveries of V.Sarianidi's expedition in South Turkmenia (Kozhin 2004; Ivanov 2004), we see the chariots on the Anatolian type of the seals in Kanish; Hurrian sculptures and other symbols of horse abound in Urkeš as if foretelling the future Mespotamian-Aryan and Hurrian excellent training of horses in Mitanni (as later in Urartu). As we saw already in our book with Gamkrelidze one of the first examples of the sacrificial horses used together with chariots in an archaic ritual was found in Sintashta; the following stud-

[23] In his later publications and public talks C.Renfrew was discussing a possibility of the inclusion of either the South-West Anatolia or Greece into the territory of the Indo-European homeland, cf. Ivanov 2004, p. 53 with references.

[24] Glottochronological calculations by Starostin in a talk at the Workshop on the chronology in linguistics in Santa Fe in 2004 give the date 4670 BC for the split of Hittite that was the first to divide from the proto-language (see for the whole set of dates Blažek 2007). On the theoretical foundation of this calculation cf. Starostin 2007, pp. 827-839, 854-852. This date is close to the one accepted by many linguists. Unfortunately the technique of glottochronological calculation is not at all unified. Thus on the base of Dyen's lexicostatistical lists completely different dates (pointing to the time ca.8700, i.e. 4 thousand years earlier than Starostin's chronology) are given in Gray,Atkinson 2003 with a general conclusion on the acceptance of the Anatolian homeland.

[25] Some other isoglosses as proto-Yenisseyan **karie* 'war' might be a very old Iranian borrowings.

ies of the cities of the Transuralian Sintashta-Arkaim area made it clear that some Indo-European (may be also Iranian) elements were at least partly present there (Grigoryev 2002; Ivanov 2004 with details and literature).

The movement of Indo-Europeans to the North of the Caspian Sea in the North-East direction documented in the Sintashta-Arkaim complex lead them much farther to the Altai-Sayany area where recent genetic investigations found traces of a Caucasoid element (Derenko e.a. 2002). Another Indo-European group moving in a parallel Eastward direction using the South Silk Road caused the presence of a similar anthropological group among the population of Central Asia (Comas e.a. 1998; Perez-Lezaun 1999). It may be supposed that the Caucasoid anthropological type of the Iranian and/or Tocharian population of the Eastern Turkestan attested in the mummies recently found there (Mallory and Mair 2000; Barber 1999) as well as in the contemporary images of the native people (Gabain 1973) is the result of these migrations from the West to the East[26]. The problem whether the boats played a role comparable to that of chariots at the time of early migrations is still to be decided by the sea archaeology[27]; at the historic time they become important only in the Late Hittic empire and in Ahhiyawa (the Ancient Greek state on the border of the latter); it seems that some kind of sea transport was used by the Greeks in their shift between Asia Minor and the continental Greece. It looks possible that before the efficient use of the chariots and horses very long mass movements were hardly possible. The first changes in the geographical position of separate dialects were caused by rather small movements as when Anatolians separated the Greeks from the rest of the East Indo-European group (that included Armenians and Indo-iranians. These short-range changes probably were performed very close to the place of the original homeland in the Near East.

ისტორია და ფილოლოგია

ინდოევროპული საწყისი ახლო აღმოსავლეთში: ახალი მტკიცებულებანი

ვიაჩესლავ ივანოვი*

* *კალიფორნიის უნივერსიტეტი, ლოს-ანჯელესი, აშშ*

საკითხი ინდოევროპელთა თავდაპირველი ადგილსამყოფელისა, საიდანაც ინდოევროპული დიალექტები გავრცელდა ევრაზიის დასავლეთ ნაწილში, მეცნიერთა რამდენიმე თაობის შესწავლის საგანს წარმოადგენს. შემოთავაზებულია რამდენიმე განსხვავებული თვალსაზრისი. ჯერ მოიაზრებოდა ჩრდილოეთის ზღვის ახლო ტერიტორია (რამდენიმე მკვლევარის შრომა XIX-XX საუკუნეების მიჯნაზე), შემდეგ შავი ზღვის ჩრდილო

[26] The opposite directions of the supposed movement of languages from Central Asia has been exaggerated by J.Nickols (1997 a,b) suggesting the Indo-European homeland in Sogdiana (near the Aral Sea where Benveniste and other specialists on Iranian had put *Eran Vēž<Airyanām Vaežo* the Aryan space'). The westward movements of languages from the Central Asia are not so numerous.. Similar objections may be made in connection to several theories on the Indian homeland of Indo-Europeans recently supposed mainly by the authors who suggest the Indo-European character of the Harappan writing (the attempts of its decipherment still do not have the general recognition).

[27] Another task for the sea archaeology may consist in falsifying a suggestion concerning a possible flood causing the end of many cities resulting from a supposed catastrophe that led to the shaping of the Black Sea (from a much smaller fresh lake to which the salted Mediterranean waters had come through the Bosporus and the Dardanelles). As the time of catastrophe may be close to the Indo-European dispersal, there are voices suggesting a link between the two events.

სანაპირო (გერმანელი მეცნიერის *შრედერის* იდეა, გაცოცხლებული ამერიკელი *მარია გიმბუტასის* და მისი მიმდევრების მიერ). ამ 35 წლის წინ ამ სტატიის ავტორმა თამაზ გამყრელიძესთან ერთად წამოაყენა ჰიპოთეზა ინდოევროპული წინარე-სამშობლოს ლოკალიზაციისა ძველი ანატოლიის სამხრეთ-აღმოსავლეთ ნაწილში, რაც საბოლოოდ აისახა მათს ერთობლივ მონოგრაფიაში ინდოევროპული ენისა და ინდოევროპული წინარე-სამშობლოს შესახებ. ამ დროიდან მოყოლებული, ლინგვისტები, არქეოლოგები და მრავალი სხვა დარგის წარმომადგენლები მსჯელობენ წამოყენებული ჰიპოთეზის შესახებ და მოჰყავთ სხვადასხვა მოსაზრებები ამ ჰიპოთეზის სასარგებლოდ ან საწინააღმდეგოდ.

ბოლოდროინდელი გამოკვლევები დამატებით მასალას აძლევს ჩვენს მიერ წამოყენებული დებულების განსამტკიცებლად ინდოევროპელთა წინარე-სამშობლოს ძველ მახლობელ აღმოსავლეთში ლოკალიზაციის თაობაზე, რასაც წინამდებარე სტატიაში წარმოვადგენთ.

REFERENCES

Adiego, I. J. *The Carian Language.* Handbuch der Orientalistik (HdO). Series 1, vol.86, Leiden; Brill, 2007.

Anthony, D.W. and D.R. Brown, D.R. Eneolithic horse exploitation in the Eurasian steppes: diet, ritual and riding, *Antiquity (An)*, 74, 283, 2000, pp.75-86.

Barber, E. *The Mummies of Urumchi*. Norton, 1999

Blažek , V. Review of Mallory-Adams 2006.- *Sbornik praci filozofické fakulty Brněnské univerzity. Studia minora facultatis philosophicae universitatis Brunensis.Linguistica brunensia* a 55, 2007.

Comas, A., Calafell, F., Mateu, E. Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D., Bertranpetit, J. Trading genes along the Silk Road: mtDNA Sequences and the Origin of Central Asian populations./ / *The American Journal of Human Genetics (Am J Hum Gen,)*, 63, 1998, pp. 1824-1838.

Crowell, B.L. The Development of Dagan: A Sketch.- *Journal of Ancient Near Eastern Religions* 1, 2001, pp. 32–83.

Dercksen J. G. *Old Assyrian Institutions* (MOS Studies 4). Uitgavenvan het Nederlands Instituut voor het Nabije Oosten te Leiden. 98. Leiden, 2004a.

Dercksen J. G. Some elements of the Old Anatolian Society in Kaniš.- *Assyria and Beyond.* T.Larsen-Festschrift. Leiden, 2004b, p. 127–178.

Dercksen,J.G.Again on Old Assyrian *tuzinnum*.- *Nouvelles assyriologiques brèves et utilitaires,2*, 2005, pp. 39-39.

Dercksen, J.G.On Anatolian loanwords in Akkadian texts from Kültepe.-*Zeitschrift für Assyriologie und vorderasiatische Archäologie (ZA), 97*, 2007, SS.26-46.

Derenko, M.V., Malyarchik, B.A., Zakharov, I.A., Proiskhozhdenie spetsificheski kavkazoidnykh mitokhondrial'nykh liniy DNK v etnicheskikh gruppakh naseleniya oblastei Altaya-Sayan (Origin of especial Caucasoid DNA mitochondrial lineages in ethnic population groups of the Altai-Sayan areas).- *Genetika (Genetics)*, 38 (9), 2002, pp. 129-197.

Diakonoff I.M., Starostin S.A. *Hurro-Urartian as an Eastern Caucasian Language*. Münchener Studien zur Sprachwissenschaft. Beiheft 12. Neue Folge. München. 1986.

Erhart, A. Archaisch oder konservativ? Das Anatolische und das Baltische.- *Kuryłowicz Memorial volume*. Part 2. Ed.W.Smoczyński. Linguistica Baltica. IV. Kraków, 1995.

Feliu, L. *The God Dagan in Bronze Age Syria.* Leiden, 2003.

Frayne, D.R. Indo-Europeans and Sumerians: Linguistic Contacts.- *Canadian Society for Mesopotamian Studies Bulletin*, 25, 1992.

Gabain, A. von, *Das Leben im uigurischen Königsreich von Qoco /850-1250/*. Veröffentlichungen der Societas Uralo-Altaica, Band 6. I. Textband; II. Tafelband, Wiesbaden: Otto Harrassowitz, 1973.

Gamkrelidze, Tamaz, Ivanov, Vyacheslav V. Problema opredeleniya pervonachal'noj territorii obitaniya i putej migratsii nositelej dialektov obshcheindoevropejskogo iazyka (Problem of the defining the original area of spread and the roads of migration of the speakers of the dialects of Proto-Indo-European). *Conferentsiya po sravnitel'no-istoricheskoj grammatike indoevropejskikh iazykov (12-14 dekabrya). Predvaritel'nye materialy* (Conference on the comparative historical grammar of the Indo-European languages [December 12-14]). A preliminary publication). Moscow, 1972. pp.19-23.

Gamkrelidze, Tamaz, Ivanov, Vyacheslav V. The early history of Indo-European languages. - *Scientific American*, vol.262, 1990, N3, pp. 110-116.

Gamkrelidze,Tamaz V., Ivanov, Vyacheslav V. Les premiers Indo-Européens de l'histoire: les ancêtres des Tokhariens en Asie Mineure ancienne.-*Revue des études caucasiques et géorgiennes,* vol.6-7, 1990-1991, pp. 265-296.

Gamkrelidze, Tamaz, Ivanov, Vyacheslav V. *Indo-European and the Indo-Europeans*. vol.I-II. Berlin-New York; Mouton- de Gruyter, 1995.

Gökçek, L. G. The use of wagons (*eriqqum*) in Ancient Anatolia according to texts from Kültepe.- *ZA,*, Bd. 96, 2, 2006.

Gray, R. D. & Q. D. Atkinson, Q.D. Language-tree divergence times support the Anatolian theory of Indo-European origin, *Nature*, 426, 2003, pp.435-439.

Greppin, John A. C., A Note on Georgian USX- and Indo-Europeanisms in the Kartvelian Languages.- *Journal of Indo-European Studies(JIES)*, vol.25, 1997, p. 383-386.

Grigoriev S. *Ancient Indo-Europeans*. Chelyabinsk, 2002.

Gusmani, R. *Il lessico ittito*. Napoli, 1968.

Harris, Alice C.: *Kartvelian contacts with Indo-European*.- Markey, Thomas; Greppin, John A.C. (eds.), *When Worlds Collide: The Indo-Europeans and the Pre-Indo-Europeans*. 1990. Karoma Press, Ann Arbor, pp.67-100.

Hill, J.E. *The Peoples of the West from the Weilue* 魏略 by Yu Huan 魚豢 A Third Century Chinese Account Composed between 239 and 265 CE Quoted in zhuan 30 of the Sanguozhi Published in 429 CE. Draft English translation, September, 2004.- Internet:< http://hudsoncress.org/html/library/china-japan/>

Ivanov, Vyach.V. Ob otnoshenii khattskogo iazyka k severo-zapadno-kavkazskim (On the relation of Hattic and North Caucasian languages).- *Drevnyaya Anatolia* (Ancient Anatolia). Moscow; Nauka, 1985, pp. 26-59.

Ivanov, Vyach. V. The Horse Symbols and the Hurrian Name of the Horse.- Bibliotheca Mesopotamica, vol.26. Urkesh. Mozan Studies 3. *Studies in Honor of Lloyd Cotsen*. Ed. G. Buccellati and M.Kelly-Buccellati. Malibu: Udena Publications, 1998, pp.145-166.

Ivanov, Vyach.V. An Ancient Name for Lyre // *Arhív Orientální*. Vol. 6, № 4 (Hrozný In Memoriam). Praha, 1999. pp. 585 – 600.

Ivanov, Vyach.Vs. Drevneyshie indoevropeyskie sobstvennye imena iz Anatolii (The most ancient Indo-European proper names from Anatolia.- *Imya: vnutrennyaya struktura, semanticheskaya aura, kontekst* (Name: internal form, semantic atmosphere, context). Part I. Moscow, 2001a, pp. 42-48.

Ivanov, Vyach. V. Southern Anatolian and Northern Anatolian as separate Indo-European dialects and Anatolian as a late linguistic zone. - *Greater Anatolian and the Indo-Hittite Family*.\ ed. R.Drews (JIE S Monograph series, N 38), Washington, D.C., 2001b, pp. 131-183.

Ivanov, Vyach.V. Comparative Notes on Hurro-Urartian, Northern Caucasian and Indo-European.- *Languages and their speakers in Ancient Eurasia*.A.B.Dolgopolsky-Festschrift. Association for the History of Language. Monograph Series 1, vol.5, ed. V.Shevoroshkin, P.Sitwell,Canberra, 2002, pp. 143-234.

Ivanov, Vyach.V. Dvadtsat' let spustya : o dovodakh v pol'zu rasseleniya nositeley indoevropejskikh dialektov iz Drevnego Blizhnego Vostoka (Twenty years later: On the reasons for the Indo-Eropean homeland in the Ancient Near East).- *Near the Sources of Civilization. The Issue in Honor of the 75-Anniversary of Victor Sarianidi (Sar.-F.)*. Moscow: Staryj Sad, 2004, pp. 41-67.

Ivanov, Vyacheslav Vs. *Indoevropeyskie korni v khettskom iazyke. Trudy po etimologii indoevropeyskikh i drevneperedneaziatskikh iazykov* (Indo-European roots in the cuneiform Hittite language and the special features of their structure. Works on the etymology of Indo-European and Ancient Near Eastern languages, vol.I.). Moscow: Iazyki slavyanskoy kul'tury, 2007.

Jansen, Th., Forste, P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. and Olek K. Mitochondrial DNA and the Origins of the Domestic Horse. - *Proceedings of the National Academy of Sciences of the United States of America*, vol.99, 2002, N 16, pp. 10905-10910.

Kennedy, D.A., Garelli, P. Seize tablettes cappadociennes de l'Ashmolean Museum d'Oxford.- *Journal of Cuneiform Studies*, vol. XIV, N1, pp.1-22.

Klimov, G. A.Zu den ältesten indogermanisch-semitisch-kartwelischen Kontakten im Vorderen Asien.-In: *Sprachwissenschaftliche Forschungen. Festschrift für Johann Knobloch,* edited by H. M. Ölberg & G. Schmidt,. Innsbruck: Amos. 1985, SS. 205-210.

Klimov, G. A. *Drevneishie indoevropeizmy kartvel'skikh iazykov (The most ancient Indo-European borrowings in Kartvelian)*. Moscow, 1994.

Klimov, G. A. Kartvel'skoe *usxo-* «byk zhertvennyj» - indoevropeyskoe **uks-on-* (Kartvelian *usxo-* 'sacrificial ox' –Indo-European **uks-on-*). - *Etymologiya* 1990-1993. Moscow : Nauka, 1994a c. 154-158.

Klimov, G. A.Kartvelian evidence for the Indo-European laryngeal? - *Indogermanische Forschungen* , 99, 1994b, pp. 62-71.

Klimov, G.A. *Etymological Dictionary of the Kartvelian Languages* . Berlin/New York: Mouton de Gruyter, 1998.

Klimov, G.A., M.Sh. Khalilov. *Slovar' kavkazskikh iazykov. Sopostavlenie osnovnoj leksiki.* (A Dictionary of Caucasian Languages. Matching of Basic Words). Moscow: Vostochnaya Literatura, 2003.

Kozhin, P.M. Kolesnyi transport preodolevaet pustyni (The wheeled vehicles have trespassed deserts for the first time).-*Sar. – F.*2004, pp. 283- 289.

Liu, Xinru. Migration and Settlement of the Yuezhi-Kushan: Interaction and Interdependence of Nomadic and Sedentary Societies.- *Journal of World History*, Volume 12, Number 2, Fall 2001, pp. 261-292.

Littauer M.A., Crowell J. H. The Origin of the True Chariot. December 1996, pp. 934-939.

Mallory, J.P.- Adams, D.Q. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford - New York: Oxford University Press, 2006.

Mallory, J.P. and Mair, V.H.. *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. New York: Thames and Hudson, 2000.

Mayerhofer M. Ein indo-arischer Rechtsterminus im Mitanni-Brief..- *Historische Sprachforschung*, 109, 1996, SS.161-162.

Melchert ,C. *Anatolian Historical Phonology*. Leiden Studies in Indo-European. 3.
Amsterdam and Atlanta, Georgia.1994.

Merpert, N.Ia. Ob etnokul'turnoy situatsii IV-I I I tysyacheletiy do n.e. v tsirkumpontiyskoy zone (On the ethno-cultural situation of the IV-I I I mil. BC in the Circumpontic area).- *Drevniy Vostok. Etnokul'turnye svyazi*.I. XXX.Moscow, pp. 7-36.

Neu, E. *Das Hurritische Epos der Freilassung*. Untersuchungen zu einem hurritisch-hethitischen Textensemble aus Hattuša. Studien zu den Boğazköy-Texten. Heft 32. Wiesbaden: Harrassowitz Verlag.1996.

Newton, M.W., Kuniholm, P.I. A dendrochronological framework for the Assyrian colony period in Asia Minor.- *TÜBA-AR* VII, 2004, pp.165-173.

Nichols, J. The Epicentre of the Linguistic Spread..- R. Blench and M. Spriggs (eds.) *Archaeology and Language (AL)*.I. London: Routledge. 1997a, pp. 122-148.

Nichols, J. The Eurasian Spread Zone and the Indo-European Dispersal. – R. Blench and M. Spriggs (eds.) *AL*, II. London: Routledge. 1997b.

Nikolaev, S.L., Starostin, S.A. *A North Caucasian Etymological Dictionary*. Moscow: Asterisk, 1994.

Perez-Lezaun A., Calafell F., Comas A., Mateu E. Bosch E., Martinez-Arias R., Clarimon J., Fiori G., Luiselli D., Facchini F.G., Pettener D., Bertranpetitj. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA.- *Am J Hum Gen*, 65 (1), 1999, pp. 208-219.

Pomponio, F. et Xella, P. *Les dieux d'Ebla*. Étude analytique des divinités éblaïtes à l'époque des archives royals du I I Ie millénaire. Alter Orient und Altes Testament. Bd. 245. Münster: Ugarit-Verlag, 1997.

Puhvel, J. Balto-Anatolian Isoglosses.- *Investigationes philologicae et comparativae. Gedenkschrift für H.Kronasser*. Ed.E.Neu. Wiesbaden, 1982.

Renfrew, C. *Archaeology and Language: The Puzzle of Indo-European Origins*. London, 1987.

Renfrew, C.The Indo-European Problem and the Exploitation of the Eurasian Steppes: Questions of Time Depth.- K. Jones-Bley and D. G. Zdanovich, eds., *Complex Societies of Central Eurasia from the 3rd to the 1st Millennium BC*, Vol. 1. Washington, D. C.: Inst. for the Study of Man, 2002, pp. 3-20.

Renfrew, F. *Arabic-Ugaritic Lexical Studies*, Münster, 1992 pp. 91–94.

Rexová, K. , Frynta, D., Zrzavý, J. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data.- *Cladistics*. Volume 19, Issue 2, April 2003, pp. 120-127.

Sherratt, A. The Origins of Farming in South-West Asia. Archaeological Atlas, October 2006, 2nd Edition <http://www.archatlas.org/OriginsFarming/Farming.php>

Shevoroskin, V.V. Inheritance versus Borrowing in Indo-European, Kartvelian and Semitic.- *J IE S*, vol. 14, 1986,.pp. 365-378.

Singer, I. Semitic *dagān* and Indo-European **d^{h}eg^{h}om*: Related Words?.- *The Asia Minor Connexion*, edited by Yoël L. Arbeitman, Louvain, 2000, pp. 221–232.

Soysal, O. *Hattischer Wortschatz in hethitischer Textüberlieferung*. HdO, abt. 1, bd.74. Leiden and Boston,2004.

Starostin, S.A. *Trudy po iazykoznaniyu (Linguistic Works)*. Moscow, 2007.

Starostin, S.A., Dybo, A.V., Mudrak, O. *Etymological Dictionary of the Altaic Languages*. HdO, sect.8, v.8/2. Part 2. Leiden:Brill, 2003.

Steblin-Kamenskij, I.M. *Dictionary of the Wakhi language*.St.Petersburg, 1999.

Süel, A., Soysal, O. The Hattian-Hittite foundation rituals from Ortaköy(I). Fragments to CTH 725 "*Rituel bilingue du consécration d'un temple*".- *Anatolica, XXXIII*, 2007, pp. 1-22.

Tänberg, A. *Der geographische Horizont der Texte aus Ebla. Untersuchungen zur eblaitischen Toponymie*.- Münchener Univesitätsschriften. Philosophische Fakultät. Altertumskunde und Kulturwissenschaften. Arbeiten zu Text und Sprache im Alten Testament. St.Ottilen: Eos Verlag.1994.

Thomas, W.*Die Erforschung des Tocharischen (1960-1984)*, Stuttgart, 1985.

Tischler J. Die kappadokischen Texte als älteste Quelle indogermanischen Sprachgutes.- O. Carruba, M. Giorgieri, C. Mora (eds.). *Atti del II congressso internazionale di Hittitologia*. Pavia, 1995, P. 359–368.

Trubetzkoy, N.S. *Izbrannye trudy po filologii* [*Selected Philological Works*]. Moscow: Progress, 1987.

Vilà C., Leonard, J.A., Götherström, A., Marklund, S., Sandberg, K., Lindén, K., Wayne, R.K., and Elegren, H.. Widespread origin of domestic horse lineages.- *Science*, 291, 2001, pp. 474-477.

Weng, Z. and Sokal, R.. Origins of Indo-Europeans and the spread of agriculture in Europe. Comparison of lexicostatistical and genetic evidence.- *Human Biology*, 67,1995, pp.577-594.

Wittaker, G. Traces of an early Indo-European language in Southern Mesopotamia.-*Göttingener Beiträge zur Sprachwissenschaft*. Heft 1. 1998.

Yankovskaya, N.B. *Klinopisnye teksty iz K'ul'tepe v sobraniyax SSSR*. (Pis'ma i dokumenty torgovogo objedineniya v Maloj Azii XIX v. do n.e. (Cuneiform texts from Kültepe in the USSR collections [Letters and documents of the trade community in Asia Minor in the XIX c. BC]). Pamyatniki pis'mennosti Vostoka XIV. Moscow; Nauka, 1968.

Received August, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

History & Philology

Towards the Identification of Luwian Hieroglyphs Denoting Fauna of Ancient Asia Minor

Nana Khazaradze*

* *Iv.Javakhishvili Institute of History and Ethnology, Tbilisi*

(Presented by Academy Member D. Muskhelishvili)

ABSTRACT. The comparison of data of the Hieroglyphic Luwian script with varied information preserved in Assyrian, Hittite cuneiform and Hieroglyphic Luwian texts leads to the conclusion that the fauna of Asia Minor is reflected in a quite adequate manner in Hieroglyphic Luwian, in spite of the fact that a great part of hieroglyphs bearing representations of animals is not identified so far. The study of Luwian hieroglyphs known to the present day shows that the creators of this script gave priority to those species of fauna (bull, horse, donkey, mule, sheep, pig, dog, lion, hare, deer) which played an especially significant role in the political military, socio-economic (farming, cattle-breeding, hunting, transport) and cultural (religious beliefs, myth, ritual) life of this major area of the Ancient East. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *Hieroglyphic Luwian Script, Late Hittites, The Royal Inscriptions, Hittite Laws.*

Written monuments in the Hieroglyphic Luwian script, discovered at various sites of Asia Minor in the 19^{th}-20^{th} cc., along with Assyrian and Hittite cuneiform texts, are one of the important primary sources for the study of the history of the peoples of Ancient Asia Minor. Hieroglyphic Luwian inscriptions provide significant information concerning natural conditions and resources of Asia Minor of the 2^{nd}-1^{st} millennia BC, as well as political, socio-economic and cultural processes going on there.

The Hieroglyphic Luwian script proper is equally valuable from the scholarly point of view.

The present study deals with Luwian hieroglyphs denoting fauna of Ancient Asia Minor. On the basis of the data available to date, the number of Luwian hieroglyphs containing representations of animals reaches 50. Some hieroglyphs denote a particular animal ("dog", "horse", "sheep"), others reflect both an animal ("donkey", "deer", "bull") and syllables - constituent parts of words (ta_4, rû, kar, sá, u(wa), muwa, mu. i_4, li, etc.), which is clear from the examples below:

"lion", Sumerogram UR.MAH-aš, UR.MAH-iš uncertain pronunciation (Laroche, N97: Meriggi, N88 a, b)

"dog", logogram, pronunciation: suwana – (Laroche, N98; Meriggi, N112-113).

* Examples are cited according to lists compiled by E.Laroche and P.Meriggi (see E.Laroche, Les hiéroglyphes hittites, I, Paris, 1960: P.Meriggi, Hieroglyphisch – Hethitisches Glossar, Wiesbaden, 1962).

"horse", logogram", pronunciation: asuwa – (Laroche, N99: Meriggi, N90 cf. N95).

I."(deity) deer", logogram, pronunciation: Tuwat –
II. rú (Laroche, N102 Sfr. 412; Meriggi, N 94-96).

sá (Laroche, N104; Meriggi, N98).

I. "bull", logogram, pronunciation: wawa- or uwa-.
II. logogram, pronunciation: usapata -, uncertain meaning
III. u (wa) (Laroche, N105; Meriggi, N109).

Ligature of two signs, muwa -, mu -. (Laroche, N107: N105+ N391; Meriggi, N108).

Logogram denoting the name of the city of Malatya, uncertain pronunciation (Laroche, N109; Meriggi, N111).

I. "donkey", logogram, pronunciation: tarkasna -, turlakalisa -.
II. ta_4 (Laroche, $N100_{1,2}$; Meriggi, N94-96).

I . "(deity) deer (transliteration: "$deer_2$"), logogram,
pronunciation: Tuwat- > Ruwat-.
II. rú
III. kar (Laroche, N103; Meriggi, N118-120).

ma (Laroche, N110; Meriggi, N104).

"sheep", logogram, pronunciation: hawa/i – (Laroche, N111; Meriggi, N107).

"hare", logogram and determinative pronunciation: tapa or tapar (Laroche, N116; Meriggi, N121).

Let us see how adequately Luwian hieroglyphs reflect the history of fauna of Asia Minor of the 2^{nd}-1^{st} millennia BC.

Significant reports regarding the fauna of Ancient Asia Minor are furnished by the text of the so-called "Cappadocia tablets", dated to the beginning of the 2^{nd} millennium BC, in which one often finds lexemes and expressions: "bull", "cow", "donkey", "black donkey", "donkey saddle", "black donkey with its harness", "goat", "sheep", "horse"[Khazaradze, 1978, 102-103]. Some documents deal with cattle and small livestock, the road tax price for livestock, e.g. donkey etc.

The existence in the Assyrian economic documents of terms denoting posts , such as: rabi alpâtim ("chief drover of bulls", rabi sîsî ("chief stable-keeper"), rabi rçîm ("chief of shepherds"), sarrîdum ("drover of donkeys"), is also very noteworthy [Cuneiform Texts from Kultepe, 1968, 220].

Significant information concerning the question of my present interest is found in Articles 57-90 from Tablet I of "the Hittite Laws" drafted as early as the period of the Old Hittite Kingdom (1656-1500 BC), which lists quite scrupulously the punishments provided for by the law for the damage caused to the owner of cattle and small livestock in different situations. In this written source of primary importance, the following animals are mentioned: "bull", "yearling", "two-year-old bull", "three-year-old bull", "bull", "plow ox", "cow", "pregnant cow", "horse", "stallion", "draft horse", "mare", "sheep", "goat", "lamb", "mule", "donkey", "pig", "pregnant pig", "dog", "guard dog", "sheepdog", "hound", "wolf", etc..

Thus, § 57 of "the Hittite Laws" reads: "If anyone steals a bull – if it is a weanling, it is not a bull; if it is a yearling, it is not a bull; if it is a two-year-old, that is a bull – they would formerly give 30 bulls. Now 9(the thief of the bull) shall give 15 bulls – 5 two-year-olds, 5 yearlings and 5 weanlings and he shall pledge his estate as security" [Friedrich, 1959, 168; Giorgadze, 1965, 87; Imparati, 1964, 164]. Under the Hittite laws, thieves of a stallion and a ram were also punished ("If anyone steals a stallion – if it is a weanling, it is not a stallion; if it is a yearling, it is not a stallion; if it is a two-year-old, that is a stallion – they would formerly give 30 horses. Now (the thief) shall give 15 horses – 5 two-year-olds, 5 yearlings (and) 5 weanlings and he shall pledge his estate as security"; "If anyone steals a breeding ram, they would formerly give 30 sheep. Now (the thief) shall give 15 sheep – 5 ewes, 5 rams, 5 lambs and he shall pledge his estate as security".

Several articles of the Hittite laws (§ 60, 61, 62) are devoted to persons finding and misappropriating cattle and small livestock: "If anyone finds a bull and removes

the brand (of its owner), and if [its] owner [traces it out], he shall give 7 bulls – 2 two-year-olds, 3 yearlings (and) two weanlings and he shall pledge his estate as security". Under the law, the same compensation must have been made by persons finding and misappropriating a stallion and a ram.

The verdict of the Hittite laws for the thieves of a plow ox and a draft horse is the following: "If anyone steals a plow ox, they would formerly give 15 oxen. Now he shall give 10 oxen – 3 two-year-olds, 3 yearlings (and) 4 weanlings and he shall pledge his estate as security". The same is provided for by the law for the thief of a draft horse ("If anyone steals a draft horse, the compensation is exactly the same for it").

Other, equally significant reports are found in "the Hittite Laws" in connection with the question of my present interest. In this legislative document "sheep", "ram", "lamb", "trained billy-goat", "trained mountain sheep" are mentioned, which provides noteworthy information for the reconstruction of the general image of fauna of Asia Minor of the 2nd-1st millennia BC ("If anyone steals a ewe or a ram, they would formerly give 12 sheep. Now he shall give 6 sheep – 2 ewes, 2 rams (and) 2 lambs and he shall pledge his estate as security"; "If anyone steals a trained (?) billy-goat, if a tame wild goat, if a trained mountain sheep, the compensation is exactly the same, as for a goat").

In "the Hittite Laws" one finds "cow" and "pregnant cow" as well ("If anyone steals a cow, they would formerly give 12 [(head of) cattle]. Now he shall give 6 [(head of) cattle] – 2 two-year-olds, two yearlings (and) 2 weanlings and he shall pledge his estate as security"; "If anyone strikes a pregnant cow and causes it to miscarry, he [shall give] 2 shekels of silver).

Six articles of "the Hittite Laws" (§ 81-86) deal with thieves of "pig", "pregnant pig", "yard pig", "pig fattened on wheat" and "piglet" ("If anyone steals a pig (fattened) on wheat, they would formerly give 1 mina of silver. Now he shall give 12 shekels of silver and he shall pledge his estate as security"; "If anyone steals a pregnant pig, he [shall give] 6 shekels of si[lver] and shall count the piglets as well, (for every) two piglets he shall give (1) measuring jug of wheat [and he shall pledge his estate as security"; "If anyone strikes a pregnant pig and it dies, [the compensation] is exactly [the same]"; "If anyone weans and steals a piglet, he shall give [...measuring jugs of wheat]"; "If a pig [goes] into a grain-heap, a field, (or) a garden, and the owner of the [gra]in-heap, field, (or) garden strikes it and it dies, he shall give it back to its owner. But if he does not give it (back), he shall become a thief").

Only one representation of dog is recorded in the Hieroglyphic Luwian script (see above). Hittite laws furnish more ample information in this regard. Three varieties of dogs are known in them: "sheepdog", "hound" and guard dog". As becomes clear from the respective articles (§ 87-89), the heaviest fine was imposed on the killer of a sheepdog ("If anyone strikes a sheepdog and it dies, he shall give 20 shekels of silver and he shall pledge his estate as security"; "If anyone strikes a hound and it dies, he shall give 12 shekels of silver and he shall pledge his estate as security"; "If anyone strikes a guard dog and it dies, he shall give one shekel of silver").

Fauna of Ancient Asia Minor is reflected in Hattian-Hittite myths and literary written sources as well. In this respect, the Telipinu myth claims attention, which relates the beliefs of the Hittites about the so-called "vanishing" deities. According to the myth, the "withdrawal" and "anger" of the deity Telipinu had grave consequences: "Mist seized the windows. Smoke seized the house. In the hearth the logs were stifled. At the altars the gods were stifled. In the sheepfold the sheep were stifled. In the cow barn the cows were stifled. The ewe rejected her lamb. The cow rejected her calf... barley and wheat do not ripen. Cows, sheep, and humans do not get pregnant. And those who are already pregnant cannot give birth... The mountains and the trees dried up; and the foliage does not come out. The meadows dried".

In this myth a special role is assigned to the sheep fleece. In the final part of the myth, dealing with the "return" of Telipinu, the sheep fleece appears as the symbol of the restoration of life: "There is the evergreen tree Eia in front of Telipinu. On this tree the sheep fleece is hanging, it is the pledge of the fertility of sheep, abundance of grains, abundance of animals, abundance of wine; that of longevity and breeding, it augurs the life, growth, abundance and fatness of sheep; then Telipinu will present the sheep fleece to the king and confer benefits on him".

Interesting information is also found in A, B and C copies of Anitta text, the first of which is dated to the 16th c. BC, whereas B and C - to the period of the New Hittite Kingdom. Lines 53-63 of this text run as follows: "and I (Anitta – N.Kh.) blessed and [cursed (?)...] On the same day [I brought to my city] Neša [two lions, 70 boars (?), 1 wild boar (?)], 120 wild animals, whether [leopards, lions, deer], ibexes, or [. . .]." [Giorgadze, 1965,87]. Some authors think that in this Hittite text hunting is implied, whereas according to others, the text deals with Anitta's initiative to create a zoological park in Neša [Tatishvili, 2001, 106; Alp, 1963, 377; Güterbock, 1938,141].

The nomenclature of animals is recorded in other Hittite written sources as well. Interesting information in this regard is found in hymns and prayers of Hittite kings too. Thus, in the prayer and hymn of Mursilis II to the Sun-goddess of Arinna one reads: "What is this that you, deities, have done? You have let loose the plague (in the interior of the land) and the land of Hatti

has been sorely, greatly oppressed by the plague. There is now constant dying! As a result, no one can offer you bread and wine as sacrifice…no one can give you sacrificial animals – bull and sheep – from cattle stalls (and) sheepfolds. Shepherds of bulls and sheep also died, they left cattle stalls and sheepfolds. Due to this, deities [will remain] without bread, wine and sacrificial animals". The prayer of Muwatallis, the second king of the Hittites, to the Weather-god ^{D}U pihašš-ašši (beginning of the 13th c. BC) contains the following lines: "You, deity of sun from heaven, my lord, you make the law for man, dog (and) pig, wild animals every day" [Tatishvili,2000,101,106].

Animals are mentioned in different contexts in texts reflecting the Hittite building rituals connected with the erection of a palace and a temple. As it turns out, when a king decided to build a palace in any city, the builder, who set off to fell trees on the mountain, received from the palace one bull, three sheep, 3 jugs of wine, 1 jug of *marnuwa*, 10 loaves of bread, 20 round cakes and 50 biscuits. At the second stage of the construction, when the foundation of the palace was laid, one bull, one cow and 10 sheep were driven out from the palace to be offered as sacrifice. The bull was offered to the Weather-god, the cow - to the Sun-goddess of Arinna, and sheep - to other deities (one to each). When a new hearth was placed for a new house, 1 pot of lard, 1 pot of honey, 1 cheese, 1 first stomach, white and black wool, 1 jar of sweet milk, 1 jar of barley malt, shelled walnuts, raisins, almonds (?), ^{GIŠ}*Šunitar*, ox skin and salt were taken from the palace and were placed in the hearth [Tatishvili, 2001, 81-83]. It also becomes clear from Hittite texts relating to building that when a temple was renewed, on the first day *kupti* (a vessel to be sacrificed) was stuck in the foundation, in which 1 lamb, 1 goat kid, 3 ducks, 30 thin loaves of bread, 1 loaf of *tarna mulati* bread, 2 small cheeses were placed, and on the second day the ritual pit was filled with 1 lamb, 1 duck, 30 thin loaves of bread and up to 18 hot loaves of bread baked from various grains [Tatishvili, 1988, 114-115].

Hittite building rituals also contain noteworthy references concerning the placement of miniature representations of animals (lion, bull) linked with various deities (Storm-god, Sun-goddess of Arinna) in the foundation of a temple or a house (e.g. "...on 4 cornerstones too, everywhere, on each cornerstone, he also puts 4 stands. 1 golden lion, (weighing) shekel; 2 pairs of iron bulls joined with a silver yoke, each bull weighning 1 shekel. There are plinths under them. Two bulls are standing on one plinth, and there are two bulls on the other plinth too" [Tatishvili, 1983, 126].

Significant infromation concerning the question of my present interest is found in Hieroglyphic Luwian and Assyrian cuneiform inscriptions of the first half of the 1st millennium BC as well.

"Bulls" and "sheep" are often mentioned in Hieroglyphic Luwian inscriptions in connection with performance of different sacrificial rituals [Khazaradze, 1978, 103]. According to one inscription discovered at Kululu (Kululu I), a vassal of King Tuwati of Tabal, some Ruwa, who built a palace at the site of ruined houses, in token of worship of deity Tarhunt, constructed *artali* (prayer house?) in his name and offered bulls and sheep from year to year" [Meriggi, 1967, 50-51].

A Hieroglyphic Luwian inscription from Sultanhani, quite difficult to interpret, informs us that a servant of Wasusarmas (King Wassurma of Tabal known from Assyrian written sources) laid out a vineyard with the aid of deity Tarhunt. "By the grace of the deity" the vineyrad grew well. In token of gratitude, the servant of Wasusarmas vowed to offer as a sacrifice from 3 to 9 sheep annually [Khazaradze, 2001, 60]. In Hieroglyphic Luwian inscriptions discovered at present-day Kululu, Boðça, Topada and Sultanhani (former territory of Tabal), "horse", "stable", "cavalry" are mentioned [Meriggi, 1967, 116-122; Khazaradze, 2001, 61]. Sheep and bulls are repeatedly mentioned in Hieroglyphic Luwian inscriptions of economic character executed on lead strips [Hawkins,1987,135].

Along with the above-mentioned, Assyrian cuneiform written sources, dated to the 1st millennium BC, also contain noteworthy information. King Tiglath-pileser III of Assyria records that he led 2000 horses away from Tabal. According to Ashurbanipal, the Tabalians brought "big horses" to Assyria as tribute [Luckenbill, 1926-1927,I, №82, 781, 911] and as Sargon II (722-705 BC) records, he captured 1000 war chariots harnessed with horses in the country of Tabal [Luckenbill, 1926-1927, II, №24, 25, 55] etc.

Thus, the comparison of data of the Hieroglyphic Luwian script with varied information preserved in Assyrian, Hittite cuneiform and Hieroglyphic Luwian texts leads to the conclusion that the fauna of Asia Minor is reflected in a quite adequate manner in Hieroglyphic Luwian, in spite of the fact that a great part of hieroglyphs bearing representations of animals is not identified so far. The study of Luwian hieroglyphs known to the present day shows that the creators of this script gave priority to those species of fauna (bull, horse, donkey, mule, sheep, pig, dog, lion, hare, deer), which played an especially significant role in the political (military science), socio-economic (farming, cattle-breeding, hunting, transport) and cultural (religious beliefs, myth, ritual) life of this major area of the Ancient East.

ძველი მცირეაზიური ცხოველთა სამყაროს ამსახველი ლუვიური იეროგლიფების იდენტიფიცირებისათვის

ნ. ხაზარაძე*

* ი.ჯავახიშვილის ისტორიისა და ეთნოლოგიის ინსტიტუტი

(წარმოდგენილია აკადემიკოს დ. მუსხელიშვილის მიერ)

იეროგლიფურ-ლუვიური დამწერლობის მონაცემების შედარებას ასურულ, ხეთურ ლურსმულ და იეროგლიფურ-ლუვიურ ტექსტებში შემონახულ მრავალფეროვან ინფორმაციასთან, ავტორი იმ დასკვნამდე მიჰყავს, რომ იეროგლიფურ-ლუვიურ დამწერლობაში მცირეაზიური ცხოველთა სამყარო საკმაოდ ადეკვატურადაა ასახული, მიუხედავად იმისა, რომ ცხოველთა გამოსახულებიანი იეროგლიფების დიდი ნაწილი იდენტიფიცირებული არ არის. ავტორის მიერ ჩატარებული კვლევა-ძიების შედეგად აღმოჩნდა, რომ ამ დამწერლობის შემქმნელები პრიორიტეტულ მნიშვნელობას მცირეაზიურ ცხოველთა სამყაროს იმ სახეობებს (ხარი, ცხენი, ვირი, ჯორი, ცხვარი, ღორი, ძაღლი, ლომი, კურდღელი, ირემი) ანიჭებდნენ, რომლებიც ძველი აღმოსავლეთის ამ უმსხვილესი რეგიონის პოლიტიკურ (სამხედრო საქმე), სოციალურ-ეკონომიკურ (მიწათმოქმედება, მესაქონლეობა, მონადირეობა, ტრანსპორტი) და კულტურულ (რელიგიური რწმენა-წარმოდგენები, მითი, რიტუალი) ცხოვრებაში განსაკუთრებულად მნიშვნელოვან როლს თამაშობდნენ.

REFERENCES

S. Alp (1963), J. Belleten, 27 (107), Ankara.

Cuneiform Texts from Kultepe in Collections of the USSR (Letters and Documents of the Trade Association in Asia Minor of the 19th c. BC). Autographical copies, transcription, translation, introduction, comments and glossary by N.B. Yankovskaya (1968) edit. I. M. Diakonoff, Moscow(Russian).

J. Friedrich (1959), Die hethitischen Gesetze, Leiden.

G.G. Giorgadze (1965), Anitta Text and Some Questions of the Early History of the Hittites, Vestnik Drevnei Istorii, I. (Russian).

H.G Güterbock (1938), J.Zeitschrift für Assyriologie und Vorderasiatische Archäologie 44, Berlin-Leipzig.

J.D. Hawkins (1987), J. Anatolian Studies 37.

F. Imparati (1964), Le leggi Hittite, Roma.

N. Khazaradze (1978), Ethnic and Political Entities of Eastern Asia Minor in the First Half of the 1st millennium BC, [Aghmosavlet mcire aziis etnik'uri da politik'uri gaertianebebi dzv. ts. I atasc'leulis p'irvel nakhevarshi]. Tbilisi (Georgian).

N. Khazaradze (2001), J. Materials for the Ethnography of Georgia, XXIV, Tbilisi (Georgian).

D. Luckenbill (1926-1927), Ancient Records of Assyria and Babylonia, I-II, Chicago.

P. Meriggi (1967), Di Eteo Geroglifico, Parte II, Testi – I-a Serie, Roma.

I. Tatishvili (1983), J. Matsne (Series of History, Archaeology, Ethnography and Art), Tbilisi, I. (Georgian).

I. Tatishvili (1988), Hattian-Hittite Building Rituals [Khatur-kheturi samsheneblo rit'ualebi], Thesis, Tbilisi (Georgian).

I. Tatishvili (2000), J. Language and Culture I, Tbilisi (Georgian).

I. Tatishvili (2001), The Hittite Religion [Kheturi religia] (Georgian).

Received September, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

History & Philology

Ethnos and Ethnic Self-Consciousness in the Archaic Period

Manana Khidasheli*

* *Iv. Javakhishvili Institute of History and Ethnology, Tbilisi*

(Presented by Academy Member D. Muskhelishvili)

ABSTRACT. Early ethnogenesis may be characterized as a spontaneous process going on within the area of distribution of a common culture defined by collective, unconscious thinking, which is revealed in the anthropological, linguistic, social, and technical unity of human groups.

Ethnoculture took shape under conditions of co-existence of people on a single territory. Their spoken language, culture of everyday life, dressing style, folk medicine, and customs were formed in this environment.

However, the unity of people must have been determined by the belief in their genetic unity, the existence of a common forefather and self-identification with him. On this basis, the archaic society was aware of its belonging to a certain cultural homogeneous unity.

Ethnic self-consciousness was the determining force of social actions of archaic society, it caused the formation of ethnocultural and ethnopsychological systems. In these systems the community of goals, interests, necessary norms of behaviour were recorded, it was understood that they had a common territory and language, shared a common past, had common ancestors, that they belonged to the same people. *© 2007 Bull. Georg. Natl. Acad. Sci.*

***Key words**: ethnos, ethnoculture, ethnic self-consciousness.*

In the archaic period the process of the formation of culture or unity of culture is complex and ambiguous. It was in this period that more or less large and significant human groups took shape. First of all, they united in order to satisfy common requirements. They were bound to their habitat due to economic interests, rather than political ones.

The co-existence of people under similar conditions resulted in the emergence of large, homogeneous, standardized cultures, comprising smaller local ones. At that time this process cannot have been influenced by a political factor. In a traditional situation a minor culture could not have striven to achieve a privileged position and try to disseminate its own culture.

Social groups living under similar cultural conditions in a small geographical area must have been striving for the formation of an anthropological, linguistic, social and technical union. Such relations resulted in the formation of a homogeneous zone of concentration of technical and cultural traits, i.e. an ethnos, undergoing constant changes. In the archaic society the process of the formation of characteristic ethnic features took a long time. The formation of ethnos required political stability, which, undoubtedly, was hard to achieve in the period under discussion [1:326-327].

Ethnic groups, inhabiting vast territories, which found themselves integrated in a comparatively friendly zone, managed to preserve and maintain their identity. But if the balance was upset and strong, unassimilated currents penetrated into the group, the latter lost its individuality and became extinct. Such a death does not mean physical annihilation, but the loss by the people of their originality [2:452-453].

In the archaic period ethnogenesis was primarily revealed in labour and was manifested in such common activities of people as war campaigns, overcoming internal obstacles, taking part in major construction work, etc. In similar cases the unity of people was conditioned by complementary processes. More active forces were selected and in this choice unconscious liking or an-

tipathy played a decisive role [3:18].

The voluntary union of human groups or their joining others must have also been connected with such factors as fear and compulsion, whereas a voluntary union with others was closely interwoven with external factors – both with fear and hope [1:35].

Thus, early ethnogenesis may be characterized as a spontaneous process, going on within the area of distribution of a common culture, conditioned by collective, unconscious thinking, and revealed in the anthropological, linguistic, social and technical union of human groups. The formation of such unions may have both voluntary and compulsory character [4:190].

Within the boundaries of the dissemination of homogeneous culture, a territory permanently populated by a human community took shape from the very beginning. The uniting of people within local territorial borders facilitated ethnic integration, due to which this territory turned into an ethnic space.

It was here, on this territory, where the life of the social group went on with its everyday traditions and customs being developed, that a human microcosm took form. Here man began to interpret and develop the world culturally, to create conceptual images of the powers and objects active in the universe. The cultural development of the world by man begins when he succeeds in the creation of "a conceptual model" of the universe, to record the surrounding phenomena in his "own conceptual vocabulary", i.e. to connect objects with the language [5:138-147].

The native spoken language was a highly significant factor in the process of development of ethnos and ethnic self-consciousness. The spoken language was common to the whole ethnos, clearly differentiating it from other ethnic groups; the language was a major force, it conditioned relations and mutual understanding among the members of the archaic society, guaranteed the preservation of the ethnocultural information and handing it down to the coming generations, and demonstrated the original features of the ethnos [6:90-91].

Human consciousness or myth, as a way of symbolic thinking, a form of the world perception, a great force acting in society and conditioning the way of life of archaic man developed within this traditional environment [7:21-33]. Myth embraced man's entire mentality; it was revealed in everything, be it a living being or an object, in every activity of people and in their speech [8:227].

The world perception of a social group defined the character of the mental field typical of archaic culture in which its characteristic social and ontological categories were taking shape; so were the social-cultural mechanisms which regulated labour processes and labour division, determined the norms of behaviour, justice, sequence of observation of religious feasts and performance of various rituals; these mechanisms also established prohibitions and a complex system of taboos. Myth also determined the archaic man's mentality, which was an integral part of the personified world and a direct participant of the processes going on in nature.

It may be said that in the archaic period culture emerged as a force of major importance, which united people. Culture was the foundation of this union and the most significant feature expressing this unity. Culture formed definite types of peoples and determined their image [4:185].

Each ethnos, like any individual, has its own history. Ethnos, like any human being, is born, grows up, reaches maturity and eventually dies. It follows therefore that both man and ethnos have biographies of their own, impressed in their memory. M.Mamardashvili writes that the essence of the consciousness of archaic man, i.e. myth, first of all, was the remembrance of the past, of the general, rather than concrete events which took place in the past, whereas memory is "the machine" which regulates the ability of having such remembrance. Man living in the mythos world evolved in such a manner that he always remembered his origin and ancestors [9:58].

And, indeed, myth as a form of symbolic thinking of man, was the only means of comprehending and realizing the universe. It also included the history of a social group because it always furnished information about the past, the sacral period of a tremendous significance, when the world had been created and the laws of human life had been established.

In the cosmogonic epoch the mythos ancestors of man, the heroes possessing the culture, were active. They fought against chaos, at the same time being engaged in cultural activities; they created culture, tilled land, farmed livestock, were competent craftsmen and handed down their skills to the coming generations. They introduced social order, spiritual values, built temples, established religious feasts and rituals.

Therefore, the mythos or sacral period was a treasury of all the spiritual powers of man, as the methods of work, knowledge, customs and traditions, justice were created in the sacral epoch and stemmed from the wisdom of the forefathers. Hence, the life of the archaic society was determined by the traditions established in the past and was based on the belief that their ancestors had behaved in that way.

So, the life of man in the profane time did not have its own value. The reality man lived in - everything he did - was determined by the model created in the mythos world. It was necessary to establish permanent relations with the sacral world and man achieved it by means of various rituals.

Through performing rituals man moved from one system of time and space into another. Rituals breathed

life into the mythos epoch, equated the events that had taken place then to a particular moment of man's profane life, in this way establishing links with the sacral epoch [10:116-117].

A ritual always had a sacral model through which man reactivated the mythical time directly touching his past - the holiness of birth and the way of life established in the past. The ritual put in the foreground man's sacral values, confirming that man had adopted all the socially valuable activities established by his forefathers [11:240-294].

This sacral past belonged only to the members of a definite social group and was important only for them; the realization by the social group of the past that was common to everybody also meant the realization of their unity. An individual perceived himself or herself as a member of a definite community of people, because all the members of this community by their origin were linked with their common ancestor who had lived in the sacral epoch. But it is also clear that here they did not mean the genetic links that existed actually but people's notions of them.

It should also be noted that archaic consciousness ruled out any interest in man as an individual or personality. In the archaic society an individual was unable to separate himself from the social group mentality being collective there. Man's accepting the values common to all meant his fusion with the spiritual unity.

Thus, it may be said that in the process of the formation of ethnos and ethnic consciousness a major role was played by the fact that there existed a common territory, spoken language common to all, and customs, traditions and rituals determined by the common mentality.

Ethnoculture evolved among people living together on a common territory, on the basis of common knowledge, experience and a certain social solidarity. Members of the archaic society had a common economy and fought together against common enemies. Various kinds of social relations were established among the people. The way of their everyday life, dressing style, folk medicine, specific features of their cuisine, various customs and traditions took shape under such conditions.

However, the unity of people must have been determined by the belief in their genetic unity, the existence of a common forefather and self-identification with him. On this basis the social group comprehended its belonging to a definite cultural homogenic community.

Ethnic consciousness was the force that determined the social activities of the archaic society; it conditioned the formation of a people's ethnocultural and ethnopsychological systems and fashioned their characteristic features. These systems included the unity of the goals and interests of the ethnos, the obligatory norms of behaviour and common mental features; they were well aware of the fact that they lived on a common territory, had a common language, shared a common past, they had common ancestors and belonged to one and the same people.

This is precisely the ethnic consciousness which, as is believed, is formed in such spheres of spiritual culture as myth, language as well as customs and traditions. Their joint functioning conditions the unity of the way of life, characteristic of the ethnos and the unity of mental processes [6:13].

The great importance of these factors expressing people's spirituality and unity was realized in the remote past. The evidence of Herodotus is noteworthy: during the Graeco-Persian war the Athenians refused to conclude any agreement with the Persians, because the Athenians were Hellenes and "the Hellenes have the same blood and the same language. Their temples of gods, sacrificing rituals, customs and traditions are also common". Here we deal with the criterion of the ethnic-genetic, linguistic and cultural union. The longer and the stronger the unity of these parameters was revealed, the higher was the level of the people's unity and originality [12:20].

The ethnos, its ethnocultural and ethnopsychological features, took shape in the process of constant relationship with other ethnic groups, each ethnic group always trying to preserve its originality and striving for self-assertion.

In order to realize its originality and defend its own interests, it was necessary for the ethnos that opposition "we – they" become active within the group, as only the awareness that there are "them" evokes the wish for self-determination (13.8). That is why it is assumed that the opposition "we – they" provided the basis of relations between various ethnic groups [14:108].

The ethnos' perception of itself as a community implied becoming aware of its own activities, aspirations and interests; it was the foundation of the originality and the ability of preserving its culture. Through protecting and preserving the ethnic union it was possible to save the territory the ethnos resided on, its spoken language, the customs and traditions established by the forefathers, to keep to the traditional ways of life and to maintain the spiritual values making for the originality of the ethnos.

Here I would like to touch upon such an important problem as the self-designation of an ethnic group. The presence of the ethnarch and the ethnonym attests to the ethnic group's quite conscious awareness of its unity and it is one of the most significant features of the ethnos.

In my opinion, such an understanding of the ethnos emerged at a relatively late stage of development. For ethnic processes to take a more rapid course and, what is more important, to acquire a more conscious, target-

oriented character, society has to achieve a higher stage of its development. A state must be formed as a guarantee of the country's political stability. A definite policy should be carried out, which will intensify the people's self-identification, work out a set policy towards the outer world, inspire the people for consolidation and worshipping their common deities, demand respect for their ancestors and observation of the traditions established by them.

The historic memory of ancient peoples always kept their sacral history, which was viewed as "the Golden Age". The Sumerian mythos reflects the beginning of the construction process of Sumerian society, when the institution of "kingship" (*nam-lugal*) descends from heaven to the town of Eridu, alternately continuing its way to different towns. The Sumerian towns were ruled by half-legendary kings. In Ancient Egypt the memory of the sacral epoch, when the country was ruled by demigods and half-heroes, was preserved [16:47, 101]. Even at the peak of their might, the Romans were proud of the fact that their genealogy began from the Trojan Aeneas. The Greeks considered Mycenaean culture to be the symbol of their genetic and cultural unity.

As the Greek sources indicate, the quite differentiated Colchian world was united by the belief that the Colchians were descendants of legendary king Aeëtes [12:29-30]. But in the Hellenistic period, in the Colchian world proper, the myth of the Colchians' genealogy emerged; according to it, the ethnarch of the Colchians', Colchi by name, was a direct descendant of the local deity Phasis.

In *Kartlis Tskhovreba* ("Life of Georgia") *Kartlosiani* is an ethnic term denoting descendants of the eponym *Kartlos*, blood relations, belonging to the same tribe [16:406]. But for Leonti Mroveli blood relationship was no longer of decisive importance. According to the reality of medieval Georgia, anyone who spoke the Georgian language, shared Georgian culture, bore the national Georgian memory and served the welfare of the Georgian people could be considered to be a Georgian. *Kartveli* (Georgian) had a national meaning and first of all referred to the unity of the Kartvelian tribes [17:341-343].

ისტორია და ფილოლოგია

ეთნოსი და ეთნიკური თვითცნობიერება არქაულ ხანაში

მ. ხიდაშელი*

* *ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

(წარმოდგენილია აკადემიკოს დ. მუსხელიშვილის მიერ)

ადრეული ეთნოგენეზი შეიძლება დახასიათდეს როგორც კოლექტიური, არაცნობიერი აზროვნებით განსაზღვრული ერთიანი კულტურის გავრცელების არეალში მიმდინარე სპონტანური პროცესი, რომელიც ხალხთა ჯგუფების ანთროპოლოგიურ, ენობრივ, სოციალურ, ტექნიკურ გაერთიანებაში ვლინდება.

ეთნოკულტურა ჟალიბდებოდა ერთიან ტერიტორიაზე ხალხის თანაცხოვრების პირობებში. ამ გარემოში იქმნებოდა და ვითარდებოდა მათი სალაპარაკო ენა, ყოფითი კულტურა, ჩაცმულობის სტილი, ხალხური მედიცინა, წეს-ჩვეულებები.

მაგრამ ხალხთა ერთობის განმსაზღვრელი უნდა ყოფილიყო რწმენა მათი გენეტიკური ერთიანობისა, საერთო წინაპრის, საერთო-საკრალური წარსულის არსებობა და თვითიდენტიფიკაცია მასთან. ამის საფუძველზე აცნობიერებდა სოციუმი თავის კუთვნილებას გარკვეულ კულტურულ ჰომოგენურ ერთობასთან.

ეთნიკური თვითცნობიერება წარმოადგენდა სოციუმის სოციალურ მოქმედებათა განმსაზღვრელ ძალას, განაპირობებდა ეთნოკულტურული და ეთნოფსიქოლოგიური სისტემების ჩამოყალიბებას. ამ სისტემებში დადასტურებული იყო ეთნოსის მიზნების, ინტერესების ერთიანობა, ქცევის აუცილებელი ნორმები, გაცნობიერებული იყო ის, რომ მათ ჰქონდათ საერთო საცხოვრებელი ტერიტორია, ენა, საერთო წარსული, ჰყავდათ საერთო წინაპრები, რომ ისინი ეკუთვნოდნენ ერთ ხალხს.

REFERENCES

1. *E. Gelnner* (1991), Nations and Nationalism, M. (in Russian).
2. *A. Leroi-Gourhan* (1943), Evolution et technique, l'Homme et la matier, Paris.
3. *L.N. Gumilev* (1991), Geography of Ethnos in the Historic Period, Moscow (in Russian).
4. *M. Khidasheli* (2002), Archaic Culture and Ethnos, Ethnogenesis of the Georgian People, Tbilisi (in Georgian).
5. *A. Leroi-Gourhan* (1964), La Gest et la Parole, II, Paris.
6. *V.Y. Kotinetz* (2000), Ethnic Self-Consciousness, SPB (in Russian).
7. *M. Khidasheli* (2001), The Picture of the World in Archaic Georgia, Tbilisi (in Georgian).
8. *O.M. Freudenberg* (1978), Myth and Literature of the Ancient Times, Moscow (in Russian).
9. *M. Mamardashvili* (2000), My Experience Is Not Typical. SPB (in Russian).
10. *M. Khidasheli* (2005), Ritual and Symbol in Archaic Culture, Tbilisi (in Georgian).
11. *M. Eliade* (2000), "The Sacred and the Secular". The Myth of Eternal Recurrence, Moscow (in Russian).
12. *R. Gordeziani* (1993), The Problem of the Formation of Georgian Self-Consciousness, Tbilisi (in Georgian).
13. *N.B. Porshnev* (1979), Social Ecology and History, Moscow.
14. *N.P. Weinberg* (1986), Man in the Culture of the Ancient Near East, Moscow (in Russian)
15. *E. Avaliani* (2001), The Formation of the Old Civilizations in the Near East and Central Mediterranean World, Tbilisi (in Georgian).
16. *T. Papuashvili* (1993), Foreign and Georgian Terminology Denoting Georgia and the Georgians, Tbilisi (in Georgian).
17. *D. Muskhelishvili* (1993), On the History of the Self-Designation of the Georgians. Foreign and Georgian Terminology Denoting Georgia and the Georgians, Tbilisi (in Georgian).

Received July, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

History & Philology

The Problem of the Origin and Continuity of Statehood in Georgia

Mariam Lordkipanidze*

* *Academy Member, I. Javakhishvili Tbilisi State University*

ABSTRACT. Scientific discussion on "At the Outset of Georgian Statehood" was held on June 2 and 12, 2000 in the I.Javakhishvili Institute of History and Ethnology of the Georgian National Academy of Sciences. Different viewpoints of Georgian scholars on the date of the origin of Georgian statehood were discussed. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *Georgian statehood.*

The problem of state formation is one of the most important issues in the history of all peoples. To give a simple, unified answer to this question is connected with many problems for the countries, Georgia proper, having ancient history: no Georgian written sources of that period existed for the ancient history of Georgia; no direct reference to this issue exists in the oldest foreign sources (Assyria and Urartu cuneiform inscriptions, Greek sources), which give information on the South Caucasus. Study of abundant archaeological material is of great importance in specifying the problem which we have in abundance, but the issue cannot be solved only on the basis of archaeological material.

The 12^{th}-8^{th} centuries B.C. Assyrian and Urartian cuneiform inscriptions mention Daiaeni and Diauhi on historical Georgian territory. Some scholars consider Daiaeni and Diauhi to have been the same formations [1:192-194; 245]. The Diauhi to were connected with the Armenian Taik and the Georgian Tao (Georgian historical-geographic province within present-day Turkey). At the same time, in the opinion of M. Dyakonov and G. Melikishvili, there was population of Hurrian generation there.

Grigol Giorgadze doubts the sameness of the Daiaeni and Diauhi considering them to have been different formations existing in different periods [2]. Neither does he share the supposition on the Daiaeni and Diauhi population belonging to the Hurrian ethnos. According to him "there can be no doubt that the Daiaeni and Diauhi belonged to the Kartvelian world", however, he does not think the existence of a state in Daiaen in the 12^{th} century B.C. fully convincing, considering the Diauhi to have been "an earlier statehood formation". [3: 16-17].

David Muskhelishvili puts the issue in a different way. He thinks that "the problem of the first Georgian state formation should be connected with the activities of Parnavaz," i.e. in the turn of the 4^{th}-3^{rd} cc B.C., when a single eastern Georgian state - Kingdom of Kartli (Iberia) – was formed. The scholar considers the existence of a western Georgian state = the Kingdom of Colchis - in the 1^{st} millennium B.C. real. Sharing the view on the identity of the Daiaeni-Diauhi , he thinks it possible to connect the origin of the Georgian statehood with eastern Georgian [4:3-4].

Guram Qoranashvili categorically rejects the statehood of both Daiaeni-Diauhi and Ancient Colchis. In his opinion,the history of Georgian statehood can start only from the period of King Parnavaz [5: 98-135].

To clarify the issue, consideration should be taken of the opinion of Otar Lordkipanidze, who sharing the view of Simon Janashia, according to which the Kingdom of Colchis took shape in the 6^{th} century B.C. However, on the basis of recently obtained rich archaeological material, Lordkipanidze tends to believe that the Kingdom of Colchis really existed from the 8^{th} century B.C., if not earlier [6:31-97].

In the 1940-1950s, and even later, some foreign (mostly Russian) and Georgian scholars did not acknowledge Simon Janashia's view on the existence of the Kingdom of Colchis (Egrisi) in the 6th century B.C. Mary Inadze, sharing the opinion of S.Janashia, sets the date of the formation of the Kingdom of Colchis in the 6th century B.C. [7: 21-30].

The newspaper "Sakartvelos Respublika" (December 26-27, 1999) published a joint article of Otar Japaridze, Roin Metreveli and mine, in which we raised the question of dating the history of Georgian statehood to 2800 years, connecting it with the Kingdom of Colchis of the 8th century B.C.

The research conducted over the last decade has shed new light on the formation of the Kingdom of Colchis. In the first volume of the eight-volume history of Georgia G.Melikishvili expressed the opinion that in the 12th century B.C. there existed a community at the stage of Colchian tribal order, which assumed the shape of a "statehood formation" with all the features characteristic of a state. [1: 205-208].

Study of the material obtained as a result of long-term joint archaeological excavations conducted by Georgian and foreign scholars in Western Georgia, in my opinion, proves the fact of the existence of the Kingdom of Colchis in the 8th century B.C. The latter emerges as a Georgian (Western Georgian) state with a single Colchian culture, whose area, in addition to Colchis proper, extended to the south, north and east,embracing Shida Kartli as well. Towards the end of the 8th century South Caucasian countries, including Colchis, were raided by the nomadic tribes from the north. As a result the Kingdom of Colchis lost some part of its southern territory and the centre of the state moved from the river Chorokhi valley to that of the river Phasis (Rioni).[1]

I consider it doubtful to start the history of Georgian statehood from the Daiaeni-Diauhi on the basis of the view expressed on the ethnic belonging of its population (Dyakonov, Melikishvili). G.Giorgadze also rules out the identicalness of the Daiaeni-Diauhi. At the same time, the Diauhi are mentioned only in the Urartian cuneiform inscriptions of the 9th century and from the end of the 8th century it is no longer mentioned. It is presumed that its territory was divided by the strengthened Urartu and Colcha [1: 196, 205]. Thus, the state suspended its existence.

Unfortunately, on the territory of the former Diauhi, which at present is within Turkey, no archaeological excavations have been conducted, depriving us of the opportunity to discuss the material culture of the ethnos living there in the period of the Diauhi.

As for Colchis, in the second half of the 2nd millennium B.C. the existence is attested of a highly developed bronze culture, with important density of population, developed agriculture, farming and cattle-breeding. Intensification of agriculture promoted gradual development of the community, reaching considerable progress in the second half of the 2nd millennium B.C. with attendant economic and social differentiation.

In the second half of the 2nd millennium B.C. the area of the prevalence of Colchian culture involves a considerable part of East Anatolia, located to the south, a significant part of Meskheti and Shida Kartli. From the end of the 2nd millennium the influence of Colchian culture increases in Shida Kartli and the influence of eastern Georgian culture grows in Colchis. In the very first century of the 1st millennium B.C. the use of iron expands considerably.

At the beginning of the 1st millennium B.C. Colchian culture reaches a high level of development. And special note should be taken of the fact that the boundary between the local cultures disappears. A single culture is represented in the entire Colchis [8:119-141]. The present discussion would seem to support the assumption on the existence of a state here. However, only archaeological material, though being an indicator of highly developed culture, cannot be sufficient to settle this problem. Special attention here is attached to the Urartian inscriptions, namely, those of King Sardur II (764-735) of 750-748 and 744-742 , which prove the existence of royal autocratic power, administrative governance and other factors in Colchis, giving ground to consider that the state of Colchis (Colcha) was formed from the 12th century B.C. to the 8th century B.C. [1:205-208]. The state was formed within this chronological period, but when? If as a result of the study of the material culture, we take into consideration the evidence cited above on Colchis of the second half of the 2nd millennium B.C. and the data on the inscriptions of Urartian King Sardur II of 750-748 and 744-742, we may assume the end of the 2nd millennium and the beginning of the 1st millennium B.C., the 10th century B.C. to be the date sought.

This, surely, is a tentative assumption, but not quite groundless [9:140]. It seems relevant to point out here that O.Lordkipanidze considers acceptable to determine the existence of the Georgian statehood over the past three thousand years [6:24]. The Kingdom of Colchis had more or less stable territory – mostly Western Georgia, single culture, the Colchians spoke a western Georgian dialect, the present continuation of which is known to us in the form of Megrelian and Laz [10], single ethnic self-consciousness, well-preserved memory of the existence of royal dynasty, known in the ancient sources as the myth of Argonauts, descendants of King Aeetes

[1] I don't think it is necessary to refer to the "materials" of scientific symposiums and the works by O.Lordkipanidze, dedicated to the problem, because all this and his opinion are represented in the volume "At the Outset of Georgian Statehood", which involves his article mentioned above.

- the Aeetids. In the 4th century B.C.the Kingdom of Colchis loses its power. In this period the western Georgian state – the Kingdom of Kartli (Iberia) – was founded, comprising the greater part of Colchis (Egrisi proper). Thus, an uninterrupted line of the existence of the Georgian state continues [9:140-141], which is of major importance for the history of statehood. In the 4th century B.C. Egrisi was within the Kingdom of Kartli. With the administrative system, founded in the period of King Parnavaz, it is one of the principalities of Kartli [11:24]. Kuji, governor of Egrisi, acknowledges his subjection to Parnavaz. He says: "You are our sovereign and I am your slave" [11:22]. Therefore, decline and fall of the Kingdom of Colchis did not suspend the line of statehood, Egrisi joined the Kingdom of Kartli and the state of Georgia continued its existence.

"King" in Georgia is a symbol of statehood. In the 540s A.D. the court of Sassanid Iran abolished the power of the king of Kartli and entrusted the governing of the country to its official – *marzpan*. But soon[1] in revolted Kartli, the local supreme power was assumed by the circle of Eristavis: "Kartli assembled and appointed Guram as an Eristavi" [12: 94]. Simon Janashia calls this process "feudal revolution" [13]. It is true that Guram, elected among the Eristavis, is not "a king" for the Sassanid authorities, but the Georgian public opinion, relations existing in the then Georgian community, recorded in the historical works of that period, recognizes Guram and his heirs as "king". The historian says that Stepanoz, son of Guram, "did not take the name of king, as he was afraid of the Persians and Greeks and called himself head of the Eristavis". [11:229]. Such governors were well aware that they were "kings", i.e. local supreme authorities. Their contemporary Georgian society apprehends them thus. In his opposition to Shio Mghvime monastery, the chief priest, Stepanoz Eristavi says: "I sit in the church of kings [14:256 -257].

In regard to the present problem "the reign" of Archil in the 8th century is most interesting.[2] A descendent of Kartlian kings, governor of Kartli, Archil is mentioned as "king" by the 11th century historian and in his conception he is the "king" of all Georgia. However, he is well aware and this is obvious from his work as well, that in this period Eastern Georgia is ruled by the Arabs and Western Georgia – by the Byzantines. In reality Archil is an official subordinated to the Arabs, head of the principality under the Arabs, who is also owner of Egrisi. According to the same historian, Archil's real activities extended to those parts of Georgia where the power of the Arabs and Greeks was less real, e.g. Kakheti, in which he builds churches and monasteries, promotes the strengthening of Christianity [15:243-244]. Georgia of that period was not united either, having many small administrative-political units, but for the Georgian community, its national consciousness it was a single country.

In the 9th-10th centuries, in the process of unification of the Georgian lands into a single state and ongoing formation of new feudal states, there appear "kings" of these states: king of the western Georgian state, the so-called "King of the Abkhaz", Georgian king of Kartli-Meskheti, king of Kartvelians, kings of Hers, Kakhs. After the formation of the single Georgian feudal monarchy (end of the 10th century), all these "kings" form the titulature of the king of single Georgia [16, 17]. At the end of the 15th and the beginning of the 16th centuries single Georgia disintegrated into three kingdoms of Kartli, Kakheti and Imereti. Starting from the 6th century, after Bagrationi became head of the Kartli state, a non-Bagrationi was not apprehended as king of Georgia. Representatives of the Bagrationi royal family of single Georgia sat on the thrones of all three kingdoms. In addition to these three kingdoms being independent of one another, they kept fighting with one another, the senior among them being the king of Kartli. It is interesting to note that the term "Georgias" was worked out in this period.

Those Muhammadan Bagrationis who were recognized by the Kartli community as their authorities, though in fact being khans (vali) of the Shah of Iran (Rostom (1632-1658), Vakhtang V (1658-1675), Giorgi XI (1703-1709), were called "kings" in Georgian historical works. Beri Egnatashvili [18: 417, 435 ...] and Vakhushti Bagrationi [19: 302, 314, 338 ...] also call them "kings".

In the 1740s, as a result of simultaneous enthroning of father and son Teimuraz II (1744-1762) and Erekle II (1744-1798) East Georgia united and since then senior or superior within the entire Georgia was Erekle II, King of Kartli and Kakheti.

In the most difficult external and internal conditions Georgia managed to defend and preserve its distinctiveness, its own supreme body of power, "king", as a symbol of the existence of the statehood down to the 19th century.

At the beginning of the 19th century, along with the annexation of Georgia, the Russian Empire abolished the kingdom in Georgia, all the representatives of the Georgian royal family were repressed and exiled to Russia.

On May 26, 1918 Georgia gained independence and the Georgian Democratic Republic was formed, which was annexed by Bolshevist Russia on February 25, 1921.The independence of Georgia was declared on April 9, 1991 by the Supreme Council of Georgia, elected in 1990.

[1] Different views have been expressed in regard to the chronology of this event, but they are not essential for our problem, The events develop between the 530s and 590s..See: David Muskhelishvili, Georgia in the 4th -8th centuries, Tbilisi, 2003, pp 225-231.

[2] Various views are expressed about the period of Archil's reigning, but they are not important for the present topic, the main thing is that by all datings his activities fall within the 8th century.

ისტორია და ფილოლოგია

სახელმწიფოებრიობის საწყისისა და უწყვეტლობის საკითხისათვის საქართველოში

მ. ლორთქიფანიძე*

* *აკადემიის წევრი, ი.ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი*

2000 წლის 2 და 12 თვინის საქართველოს მეცნიერებათა აკადემიის ო.ჯ.ჯავახიშვილის სახელობის ისტორიისა და ეთნოლოგიის ინსტიტუტში ჩატარდა სამეცნიერო დისკუსია თემაზე: ქართული სახელმწიფოებრიობის სათავეებთან. დისკუსიის მასალები ამავე სათაურით გამოქვეყნდა 2001 წელს.

გამოითქვა რამდენიმე მოსაზრება. ქართული სახელმწიფოებრიობის ისტორია შეიძლება დაიწყოს: 1. ძვ. წ. XII–VIII საუკუნეების ასურულ და ურარტულ ლურსმულ წარწერებში მოხსენიებულ დაიაენი-დიაუხით, ან საკუთრივ დიაუხით; 2. ძვ. წ. IV–III საუკუნეთა მიჯნაზე ქართლის სამეფოს ჩამოყალიბებით; 3. კოლხეთის (ეგრისის) სამეფოს ჩამოყალიბებით, რომლის შექმნას ზოგი ძვ. წ. VI ს-ში, ზოგი ძვ. წ. VIII ს-ში ვარაუდობს.

ვფიქრობთ, ძვ. წ. II ათასწლეულის მეორე ნახევრის არქეოლოგიური მასალა, ძვ. წ. VIII ს-ის ურარტუს მეფე სარდური II-ის (764–735) 750-748 და 744–742 წლების წარწერები, ენობრივი მონაცემები, უფლებას იძლევა ქართული სახელმწიფოს არსებობა ძვ. წ. II ათასწლეულის დასასრულით და I ათასწლეულის დასაწყისით, პირობით, ძვ. წ. X საუკუნით ვიფარაუდოთ ამ დროიდან საქართველოში, ამა თუ იმ სახით XIX ს-ის დასაწყისამდე სულ არსებობდა სახელმწიფო, რომლის სიმბოლო იყო "მეფე". XIX ს-ის დასაწყისში რუსეთის საიმპერიო ხელისუფლებამ გააუქმა მეფის ინსტიტუტი და ამ ეტაპზე შეწყდა ქართული სახელმწიფოს არსებობა.

1918 წლის 26 მაისს საქართველომ მოიპოვა დამოუკიდებლობა და შეიქმნა საქართველოს დემოკრატიული რესპუბლიკა, რომელმაც იარსება 1921 წლის 25 თებერვალს ბოლშევიკური რუსეთის მიერ მის ანექსიამდე.

1990 წ. 28 ოქტომბერს არჩეულმა საქართველოს უზენაესმა საბჭომ 1991 წლის 9 აპრილის აქტით, გამოაცხადა საქართველოს დამოუკიდებლობა.

REFERENCES

1. *G.A. Melikishvili* (1989), Drevneishie rannefeodal'nye gosudarstva i plemennye soyuzy v yugo-zapadnoi Gruzii (*Ancient early-feudal states and tribal unions in south-east Georgia*). In: Ocherki istorii Gruzii. I, Tbilisi (in Russian).
2. *G. Giorgadze* (2000), "Mushkaia", "mushki" and "Meskheti, History ani" asurul lursmul tekstebshi ("*Mushkaia", "Mushki" and "Daiaen-Diauh" in Assyrian cuneiform texts*). In: Meskheti, istoria da tanamedroveoba. Akhaltsikhe-Tbilisi (in Georgian).
3. *G. Giorgadze* (2001), Daiaen-diaukhi: arakartveluri samqaro? (*Daiaen-Diauh: non-Kartvelian world*?) In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).
4. *D. Muskhelishvili* (2001), Redaktorisagan (*From the editor*). In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).
5. *G. Qoranashvili* (2001), Dzvel kolkhetshi sakhelmtsipoebriobis arsebobis sak'itkhi (*The issue of the existence of the statehood in Ancient Colchis*). In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).
6. *O. Lordkipanidze* (2001), Iqo tu ara sakhelmtsipo kolkhetshi? (*Was there a state in Colchis or not*?) In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).
7. *M. Inadze* (2001), G. Qoranashvilis tsignis "Sakartvelos sakhelmtsipos tssarmoshobis sakitkhis gamo" (*On G. Qoranashvili's book "The Problem of the Origin of the Georgian State"*). In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).

8. *O.M. Japaridze* (1989), Zapadnogruzinskaya kul'tura (*West Georgian culture*). In: Ocherki istorii Gruzii. I (in Russian).
9. *M. Lordkipanidze* (2001), Kartuli sakhelmtsipoebriobis utsqveti khazi (*Continuity of Georgian statehood line*). In: Kartuli sakhelmtsipoebriobis sataveebtan. Tbilisi (in Georgian).
10. *Th. Gamkrelidze* (1998), "Okros satsmisi da dzveli kolkhetis mosakhleobis etnikuri vinaoba, anu ra enaze metqvelebda mepe Aieti?" ("*Golden Fleece" and ethnic identity of old Colchis population, or what language did King Aeetes speak?.* ") In: "Sakartvelos Respublika", March 20-27, Tbilisi (in Georgian).
11. *Leonti Mroveli* (1955), Mepeta tskhovreba (*Life of the Kings*). In: Kartlis Tskhovreba, I, Tbilisi (in Georgian).
12. Moktsevai Kartlisai (1963), In: Dzveli kartuli hagiograpiuli literaturis dzeglebi (*Monuments of old Georgian hagiographical literature).* I, Tbilisi (in Georgian).
13. *S. Janashia* (1949), Peodaluri revolutsia sakartveloshi (*Feudal revolution in Georgia*). In: S. Janashia. Shromebi. I, Tbilisi (in Georgian).
14. Tkhroba sastsaultatvis .... Shiosta. Sakartvelos samotkhe (*The Story of Miracles ... of Shio, Paradise of Georgia*). (1882). M.Sabinin's publication, Petersburg (in Georgian).
15. *Juansheri*. Tskhovreba mepe vakhtangisa. *Life of King Vakhtang*, In: Kartlis Tskhovreba, I, (1955), Tbilisi (in Georgian).
16. *M. Lordkipanidze* (1963), Peodaluri sakartvelos politikuri gaertianeba (*Political unification of feudal Georgia*). Tbilisi (in Georgian).
17. *M. Lordkipanidze* (1994), Essays on Georgian history, Tbilisi.
18. *Beri Egnatashvili* (1959), New Life of Kartli, In: Kartlis Tskhovreba, II, Tbilisi.
19. *Vakhushti Bagrationi* (1973), Description of the Kingdom of Georgia, In: Kartlis Tskhovreba, IV Tbilisi (in Georgian).

Received September, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

History & Philology

On the Personality of the Creator of the Abkhaz Kingdom

Giuli Alasania*

* *International Black Sea University, Tbilisi*

(Presented by Academy Member Th.V.Gamkrelidze)

ABSTRACT. Leon II is known in the scholarly literature as the creator of the Abkhaz Kingdom. One can see some discrepancies in the evidence of the related parts of "Kartis Tskhovreba" ("History of Kartli"), "Matiane Kartlisa" ("Chronicle of Kartli" – the 11th century) and that of the Chronicle of Wakhushti Bagrationi (the 18th century). The first one is more correct. The majority of scholars read the text as follows: " ... at the time, when Greeks became weak, the eristavi of the Abkhaz broke away from them. He was called Leon and was the nephew of Eristavi Leon, to whom Abkhazia was given hereditarily. This second Leon was the grandson of the king of the Khazars and with their aid he broke away from the Greeks and conquered Abkhazia and Egrisi". Both Leons are considered to be dukes of Apkhazeti – Leon I, Leon II. The author of the present article argues that one Leon mentioned by the anonymous author of "Matiane Kartlisai" was Leon of Isauria, the Byzantine Emperor, who appointed his nephew Leon as a ruler of Abkhazia in the late third decade of the 8th century.

Proceeding from an analysis of the related sources, the author of the paper comes to the conclusions that:

1. There was one Leon in Abkhazia in the 8th century;

2. Leon was a family member neither of the Abkhazian, nor of the Egrisi (Western Georgia) rulers;

3. Leon was the nephew of Leon of Isauria, the Byzantine Emperor, who appointed him as a ruler in Abkhazia;

4. Based on his wife's rights, Leon united western Georgia and became its ruler in the late eighth decade of the 8th century. His wife was Gurandukht, daughter of Mir, to whom Egrisi was given as an apanage by Leon of Isauria;

5. Leon seceded from Byzantium and proclaimed himself King of Abkhazia, which covered western Georgia, in the same late 8th century, closer to 786. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *Leon King of the Abkhazians, Georgian historiography.*

In Georgian historiography Leon II is considered to be the creator of the "Abkhaz Kingdom". Leon II is identified with the Leon mentioned in the respective part of the Georgian Chronicles "the Chronicle of Kartli" ("Kartlis Tskhovreba") as well as with the Leon in the same part of Wakhushti Bagrationi's work. According to the ancient Georgian chronicle: "At the time when Greeks lost their power, Leon Eristavi of Abkhazia – the nephew of Eristavi Leon to whom Abkhazia was given hereditarily – broke away from them. This second Leon was the son of the daughter of the Khazar King and with the help of Khazars he ousted the Greeks and conquered Egrisi and Abkhazia as far as the Likhi. And since Iovane was dead and Juansher was old and later passed away, too – the Leon was called "the King of the Abkhazians" (Kartlis Tskhovreba, 1966, p. 206).

Later on, Leon is mentioned in the chronicle again: "at the time Curopalates Ashot launched a campaign. He was helped by Teodosi, the King of Abkhazians and the son of Leon II that was son-in-law of Curopalate Ashot (p. 206), "at that time there came Giorgi, the King of Abkhazians, the brother of Teodosi and Demetre and the son of Leon" (p 208).

In the Armenian translation of the Georgian Chronicles the information is presented in a somewhat different way: There is no evidence of Leon the Eristavi, and the uncle of Leon and Leon King of the Abkhazians are referred to as "a certain Leon, the son of the daugh-

ter of the Khazar King": "Then there came Khosrodisi, Ostikan of Armenia and built Tpilisi that had been destroyed by Khazars. And a certain Leon, the son of the daughter of the Khazar King was crowned to rule over Abkahzia and Egrisi as far as the Likhi. The sons of Archil, Ivane and Juansher, passed away and their place was occupied by Curopalate Ashot. Caesar glorified him. The Agarians got weak. Teodosi took the place of his father Leon when the latter passed away" (Kartlis Tskhovreba, 1953, p. 213).

As compared to the chronicles, Wakhushti presents the same information in a different way, too: "Since the Greeks were weakened by the Agarians, since they had taken overseas – the second Leon, nephew of Leon the Eristavi of Abkhazia conquered Egrisi and was called the King of Abkhazia as Ioane Eristavi passed away and Juansher got old and subsequently died, too".... (Wakhushti Bagrationi, 1973, p. 127). Wakhushti presents the information twice. In the second case, it is given in the chapter: "The description of the country of Egrisi, i. e. Abkhazia, i. e. Imereti" (p. 742). The author gives the names of countries – "Egrisi, the second – Abkhazia, the third – Imereti" and explains why he calls the places these names. "And Abkhazia because of Levan, as after Levan I, Levan II was Eristavi of Abkhazia (in 786). After the death of Khosroans, Leon became King and conquered the whole Egrisi and called his kingdom Abkhazeti and established the name of his saeristavo in Egrisi as a whole".

As we can see, unlike the chronicle where there is no chronology given, Wakhushti Bagrationi indicates the exact date of the renouncing of Abkhazia - 786. At the same time he elucidates whose nephew was Leon - the Eristavi of Abkhazia of those times. In his opinion he was "the nephew of Leon the Eristavi of Abkhazia". I.e., whereas the Chronicle of Kartli refers to Leon as the Eristavi of Abkhazia and to his uncle only as to "Leon the Eristavi", without the indication of the place, Wakhushti does the opposite: he calls the nephew "the second Leon", and his uncle - "Leon the Eristavi of the Abkhazians".

We believe that it is here that an inaccuracy in Georgian historiography was made. According to the prevailing opinion, there were two Leons in Abkhazia in the 8th century – Leon I and Leon II. Yet, some other points of view have also been expressed.

In the French translation of "the Chronicle of Kartli" Marie Brosset ascribes the words: "to whom Abkhazia was given hereditarily" to the uncle" – "Léon eristav d'Apkhazie, fils du frère de Léon, à qui l'eristawat héréditaire de cette contrée avait été confréré précédemment, se détacha d'eux" (M. Brosset, 1850, p.259).

Other researchers interpreted this place in the same manner. In one place C. Toumanoff points out that the uncle of Leon II – Leon I was called "the Eristavi of Abkhazia". In another place, C. Toumanoff translates the information of the "Chronicle" in the following way: "the Duke of Abkhazia, whose name was Leon, to whom Abkhazia had been given as an apanage... The second Leon was the son of the daughter of the King of the Khazars and with their aid he revolted against the Byzantines..." (C. Toumanoff, 1956, p. 75).

In his work devoted to the establishment of the Kingdom of Abkhazia S. Janashia translates the place we are interested in as follows: "When the Greeks got weak, the Eristavi of Abkhazia called Leon, the nephew (the son of the brother) of the Eristavi of Abkhazia, to whom Abkhazia was given as an apanage – renounced from them. This second Leon was the son of the daughter of the Khazar King and he renounced from the Greeks with their assistance. He conquered Abkhazia and Egrisi as far as the Likhi, called himself the King of Abkhazia - since Eristavi Ioane was dead and Juansher was old"(S. Janashia, 1952, p. 325.). Z. Anchabadze translates the extract in question (p. 95): "When the Greeks got weak, the Eristavi of Abkhazia called Leon, the nephew (the son of the brother) of the Eristavi of Abkhazia, to whom Abkhazia was given as an apanage – renounced from the Greeks. This second Leon was the grandson (the son of the daughter) of the Khazar King and he fell away from the Greeks with the assistance of the people. He conquered Abkhazia and Egrisi as far as the Likhi and called himself the King of Abkhazia" (Z. Anchabadze, 1956, p.95). According to P. Golden, Abkhazia was given as an apanage to the uncle of Eristavi Leon – Leon: "But at the time when the Greeks (the Berdzenni) became weak, the eristavi of the Ap'xaz broke away from them. His name was Leon, the nephew of the eristavi Leon, to whom Ap'xazet'i was given hereditarily. This second Leon was the grandson of the king of the Khazars and with their aid he broke away from the Greeks and conquered Ap'xazet'i and Egrisi..."(P. Golden. 1980, p. 63).

During the time that we are investigating, there is only one Leon mentioned in the "the Divan of the Abkhazian Kings" discovered by E. Takaishvili. The Leon, according to the source, governed the country for 45 years. By means of counting back the indicated years of ruling of Abkhaz kings, E. Takaishvili offers the genealogy of Abkhazian kings with the indication of respective years. There is only one Leon mentioned in the 8th c here (746 – 791). Yet, right there the comment of the author comes into conflict with the scheme: "for Leon I Caesar determined Abkhazia as saeristavo from the Klisura to the river of great Khazar. With the help of Khazars Leon II – the nephew of this Leon fell away from the Greeks, i.e., Byzantines, occupied all Egrisi as far as the Likhi Mountains and declared himself the King of Abkhazia in 786" (E. Takaishvili, Tbilisi, 1913, p. 53-54).

I. Javakhishvili agreed to the date of Leon's rule proposed by E. Takaishvili – 746 – 791. He does not mention any other Leon in those years since, seemingly, in his opinion there was only one Leon in the 8th century Abkhazia (I. Javakhishvili, 1983, p. 92-93).

S. Janashia touched upon the issue of the establishment of the Kingdom of Abkhazia. He put forward a noteworthy position on the issue in question. The scholar did a detailed analysis of the sources that refer to Leon. S. Janashia points out that the "Divan of the Abkhazian Kings" is not complete and precise. He argues that the information of Wakhushti is entirely based on the "Chronicle of Kartli", as for the notes that are different – they are invented by Wakhushti. The scholar agrees with Wakhushti on the date of creation of the Kingdom of Abkahzia – 786 and sees the year as more reliable. At the same time he rejects the dating of Leon's rule offered by E. Takaishvil – 746-791. As we specified above, the same position is shared by I. Javakhishvili. S. Janashia gives a detailed picture of the 8th century, the relations of the Byzantines, Arabs, and Khazars. Consequently, he expresses a doubt that the Caesar of Byzantium "Khazar" and Leon the founder of the Kingdom of Abkhazia were cousins - their mothers being sisters, the daughters of the Khagan of the Khazars. "True, the old Georgian historian refers the name of Leon to the founders of the Abkhazian dynasty as well", "S. Janashia doubts the historicism of Leon I and supports his doubt with the "Divan of the Abkhazian Kings", which, as we mentioned above, acknowledges one Leon only". In S. Janashia's opinion, this Leon corresponds to Leon II. The scholar dates the founding of the Kingdom of Abkhazia to the period after 797 (797 – 802) (S. Janashia,1952, p. 323-341). In some cases this same dating is shared by some other scholars (D. Muskhelishvili,1980, p. 140; M. Biro, 1977, p. 256). At the same time, except for a few cases given below, the doubt of S. Janashia regarding the existence of two Leons has not been shared.

P. Ingorokva elaborated on the issue of the Kingdom of Abkhazia in his work "Giorgi Merchule" (1954). In his genealogical scheme of the kings of Abkhazia, the princes of Lazeti (the Patrikios) and afterwards the kings of Western Georgia ("the kings of Abkhazia") – there are both Leon I and Leon II in the 8th century.

Leon I, Teodosi I (730-740), Konstantine II; Leon II (753-798) was crowned in 786-798.

"In the 720-30-s there started a wide-scale movement against Arab conquerors in Western Georgia. This campaign was organized and led by the prince of Lazeti – Leon I. Leon I is an important figure in the history of Georgia of the 8th century". "The merit of Leon I - the organizer and leader of the anti-Arab movement is immense for Georgia....Leon I, after waging a battle with the Arabs, left the inner part of Egrisi where the position of Arabs was pretty strong. Leon I moved to Abkhazia – the north coast of the Lazi countries and consolidated his positions there.... Leon I liberated the territories of Abkhazia: Abkhazia itself, as well as Absileti, Misimianeti, Saniget-Jiketi... He established political relations with Byzantium and acknowledged the suzerainty of the Caesar of Byzantium. At the same time, Byzantium, in its turn, took on some responsibilities. Namely, it recognized the Dynasty of Leon I as the hereditary owner of the territories of Abkhazia that had been liberated from the Arabs – with the title of the Eristavi of Caesar – Archonti....the leaders of the anti-Arab front in Iberia (Kartli), the princes of Iberia: Mir and Archil established relations with Leon I"(P. Ingorokva, 1954, p. 197-198). P. Ingorokva considers that western Georgia was liberated from the "Byzantine wardship" by Leon II who got Abkhazia in the 750s (according to the text of "the Divan of the Abkhazian Kings" - 753). His rule continued to 798. Leon II got the title of "prince" earlier than the 780s. Leon II is mentioned with the title "prince" in the work of Ioane Sabanisdze which gives information about western Georgia of 781-782. The integration with western Georgia took place at the time of Leon II (p. 202). The scholar believes that this territory included "the Trebizond region as well" and Leon established relations with Khazars, too.

While Z. Anchabadze tends to agree with S. Janashia, much of P. Ingorokva's analysis on "the Kingdom of Abkhazia" was unacceptable for him. At the same time, Z. Anchabadze leaves two Leons in the 8th century. According to Z. Anchabadze, "after the ousting of the Arabs, the territory of the owner of Abkhazia ("The Saeristavo of the Abkhazians") included an important region from the River Kelasuri in the South to Nechepsukho in the North. The Emperor of Byzantium Leon of Isauria (717-741) appointed Leon I within these bounds (Z. Anchabadze,1959, p. 60). The scholar believes that Leon I owned a fortress, Sobghisi, based on the notes of the work of Juansher in the respective part of the "Georgian Chronicle": "The Eristavi of Caesar Leon entered the fortress Sobghisi". Z. Anchabadze comments that the Georgian sources refer to Leon as the Eristavi of Caesar because Abkhazia depended on Byzantium (p. 87). Afterwards, Z. Anchabadze talks about the "Divan of the Abkhazian Kings" and asserts that the list of Abkhazian chiefs given there is not complete. He argues that in the "Divan of the Abkhazian Kings" there is an omission of Leon, the uncle of Leon II.

N. Berdzenishvili assumes that in the 8th century there were 2 Leons in Abkhazia: Leon I and Leon II. "In the opinion of Georgian politicians, with the help of Greeks, "Abkhazia" (the western part of the Egrisi heritage) confronted Kartli. "Kartli" contained the remaining part of Egrisi. This "Abkhazia" sometimes bordered

on the Egris-tskali and sometimes included the Inner Egrisi, as well. At times, it only reached Klisura. At the end, with the support of Greeks, "Abkhazia" encompassed the whole territory of Egrisi (presumably, this was in the times of Leon I) and thus there appeared the country of Abkhazia (at the time of Leon I to Scanda-Shorapani and from the time of Leon II – as far as the Likhi) (N. Berdzenishvili, 1990, p. 269). In the context: "as the «Deaf» entered the Klisura and approached the Fortress of Anakopia, there were the kings of Georgia Mir and Archil inside. And the Eristavi of the Caesar entered the fortress on the border of Ovseti" (p. 577). The scholar interprets "the Eristavi of Caesar Leon" as the Eristavi of Abkhazeti that "famed in this way, corroborated by Sabanisdze (the Georgian author of the 8th century). His praise by the king of the Greeks is corroborated in this episode".

D. Muskhelishvili talks about Leon II. "On the initiative of the government of Abkhazia, the Principality of Egrisi and Abkhazia united in about 778. In 797-802 this new political entity got free from the vassal attitude of Byzantium and formed as the "the Kingdom of Abkhazia". According to Wakhushti Bagrationi, Leon II – the founder of this kingdom - carried out a series of significant administrative reforms" (D. Muskhelishvili,1980, p. 140).

In the respective part of his major work on Khazars, M. Artamonov proceeds from the French "Georgian Chronicle" published by Marie Brosset and mentions Lev II (M. Artamonov, 1962, p. 248).

There are also two Leons mentioned in all relatively modern editions (N. Asatiani, M. Sanadze, T. Beradze) including volume II, the index of Ivane Javakhishvili's works, though Ivane Javakhishvili speaks about one Leon in his research (N. Asatiani, 2004, p. 129-130; T. Beradze, M. Sanadze, 2003 p. 100; I. Javakhishvili, 1983, p. 460).

A different position has also been put forward. M. Lordkipanidze thoroughly studied the history of the Abkhaz Kingdom. Although initially she agreed with the prevailing opinion about two Leons in the 8th century (Leon I - from the 740s – to 758, Leon II – 758 – 798), later she made an assumption about one Leon. The scholar reckons that the Leon, who led the anti-Arab front in Western Georgia in the 750s, united all Western Georgia by the 780s and got free from the vassalage of Byzantium and assumed the title of king at the end of the 8th century (M. Lordkpanidze,1973, p.417,429; M. Lordkipanidze,1988, p.281).

Based on appropriate sources and environment, we also consider that in the 8th century there was only one Leon who was first acknowledged as the Eristavi of Abkhazia. Later he took advantage of the domestic situation and the problems in Byzantium and managed to unite Western Georgia with the help of the Khagan of the Khazars. He broke away from the empire, declared independence and was crowned.

In the "Georgian Chronicles" Leon is mentioned by Juansher even before "the Chronicle of Kartli". Leon is mentioned twice in the latter. This name is first mentioned with regard to the campaign of Marwan the Deaf (Marwan ibn-Muhammad). When the Arab conqueror reached Anakopia there were "the kings of Kartli: Mir and Archil there". In the same place there is a reference to "the Eristavi of the Caesar Leon" who "entered the Fortress of Sobghisi, which borders on Ovseti" (p. 193). Later it is said by Juansher: "There came the time of lechery in Kartli, Armenia and Rani... and Mir, Archil and Leon, the Eristavi of Abkhazia, sent an envoy to the King of Greece... The latter produced two crowns and the deed and gave them to Mir and Archil" ...and he wrote to Leon: "I confer the status of the Eristavi of Abkhazia on you and all your successors. May you always respect the kings and the nation of Kartli; may you never be able to harm them and may their borders be on Egrisi". The dying Mir addresses Archil: "since our fathers get married to the daughters of Eristavis, give them my daughters and share the land of Kartli with them. Let half be yours, and the other half - theirs. I am giving the oldest to you: Egrisi, Suaneti, Takueri, Argueti, Guria. Give Shuamtiuleti and Klarjeti to my daughters so that they could be home at the times of sorrow.... (p. 196). "Archil addressed Leon and stated".... "I will go and settle in the Fortress Goji and Kutatisi. Ask me whatever you want"... "Leon responded": "Caesar gave me a land and from now on it is mine from Klisura to Khazareti – where there is the peak of the Caucasus... I do not want anything from you. Let my possession be yours". Then Archil married his niece Guarandukht to Leon and gave him the crown bestowed upon Mir by the King of Greece.... Archil settled in Egrisi as far as Shorapani. 12 years passed and the construction of Kartli started. Mtskheta was a poor place to live in" (Kartlis Tskhovreba,1996, p. 196-198).

None of these extracts, except for the first one, is included in the Armenian translation of 'the Georgian Chronicles". "Leon the Eristavi of the Caesar" mentioned in the first note is considered to be the Eristavi of Abkhazia in scholarly literature. According to C. Tumanoff, Leon I - the uncle of Leon II – is called "Imperial Duke" (C. Toumanoff, 1956, p.75). And following the analysis of Z. Anchabadze, Leon I was called "the Duke of the Caesar" because of Abkhazia's vassalage to Byzantium. M. Biro reflects in the same way: "In the chronicle of Juansher the chief of Abkhazia is presented as "Imperial Prince" and we can deduce from this that in 737 Abkhazeti was controlled by Byzantium" (M. Biro, 1975, p.297). The comment of M. Sanadze and T. Beradze should be specially noted. The scholars proved that "the Duke of Caesar Leon" is Leon of Isauria (717 – 741) and

not "the Duke of Abkhazia". They proceed from the Armenian translation of the "Georgian Chronicle" as well as some other documents. In the Armenian translation of the "Georgian Chronicle" the Leon we are interested in is mentioned as "Leon, the leader of the Greeks" (M. Sanadze, 2001, p. 75-77; M. Sanadze, T. Beradze, 2004, 76-78).

As for the other extracts from the work of Juansher, it is quite challenging to date them. Juansher does not mention the name of the Caesar of Byzantium to whom the Georgian chief sent an envoy. According to Wakhushti Bagrationi, this Caesar is Konstantine (741 – 775). After the attack of Marwan ibn Muhammad and the fight at Anakopia where Mir and Archil participated: "Mir was wounded" and Arabs retreated, they passed Guria and headed for Baghdad;" "the kings sent a person to Caesar Konstantine and notified him on everything. As a response the Caesar sent them a crown and praise" (Wakhushti,1973, p.125)..

It has been noted in the literature that Juansher tends to present events in a compressed way. Though, as a result of comparative studies, it becomes possible to identify a reliable stratum in the source - the material presented here does not suffice for determining the date of the events. When could the Georgian princes have applied to the Caesar of Byzantium? Most probably, that happened when the reign of Marwan ibn Muhammad in Armenia came to an end. As the Armenian author Ghevond, who considers Marwan ibn Muhammad to have been a just ruler, relates: "he punished offenders: thieves, crooks, and other violators of order with deaths on the gallows; he gave orders to cut the legs and arms of the accused" (Ghevond, 1862, p.81).

After Caliph Hisham (724 – 747), there started a power crisis. In a short period of time, several Caliphs replaced each other: Walid II (743 – 744), Yazid III (744), Ibrahim (744). Following them, Caliph Marvan II ibn Muhammad (744 – 750) – the last representative of the Omayad Dynasty took hold of the throne by means of force. However, in a few years he was defeated by the Abasids (St.Lain-Pool, 2004). According to the Syrian author Abu al-Faraj the same Bar Hebraeus, Walid reigned 1 year and 3 months after Hisham. Yazid ruled for 2 months and his brother Ibrahim – for another 2 months. Then there was Marwan (Abu'l Faraj, 1932, p. 742). In these years of turmoil, as Ghevond put it, "all our leaders wanted to overthrow" the rule of Arabs, "to revolt and get rid of tributes". They relied on the help of the Caesar of Byzantium in this respect. There is evidence how Konstantine the Caesar of Byzantium provided refuge for the princes who had escaped from Arabs (Ghevond, 1862, p. 82-83, 93). The Georgian plea to Caesar of Byzantium could be dated back to this time. And the years of Leon's rule (745 – 791), as proposed by E. Takaishvili, seem quite acceptable, which of course includes not only the time of royalty, but such an appeal could have been secretly made right then – after the campaign of Marwan ibn Muhammad – i. e., in 737-739.

After the campaigns of Marwan ibn Muhammad, the necessity to form coalitions against Arabs became evident. Seemingly, as formerly members of such coalitions were Byzantium, the Kingdom of Khazars and Georgia. The union was traditional starting with Heracles the time of the Caesar of Byzantium. It never stopped to exist even at the time of Arabs or the establishment of the independent Khazar state. In respective sources we can trace the association of Leon of Isauria with Khazars. Even before the attack of Marwan ibn Muhammad against Georgia, in 732/3 the Caesar of Byzantium – Leon of Isauria married his son Konstantine to the daughter of the Khagan of Khazars, who after baptizing was named Irene (I. Chichurov. 1980, p. 68). Later on (according to the dating of Theophanes in 750) this marriage resulted in the birth of Leon the IV "Khazar", later – of the Caesar of Byzantium. He took the throne after his father Konstantine.

According to the sources, the Eristavi of Abkhazia was Leon in those years. It is mentioned in the "Chronicle" that his mother, whose name is not mentioned in the source, was the daughter of Khazar Khakan. Who was the father of Leon? According to the "Divan of the Abkhazian Kings", the father of Leon was Thevdore. He is mentioned as the ninth king of Abkhazia. In the above-mentioned list of other predecessors, Leon owns the Kingdom of Abkhazia and rules it hereditarily; the throne goes from father to son (J. Gamakharia, B. Gogia, 1997 p. 198). The marriages were not accidental in those times. This period is significant because the strong Arabic coalition was held. Together with Byzantium's, Khazar's and Kartli's kings the ruler of Abkhazia participates intensively in this coalition. If we believe the evidence provided by the Georgian sources, these two people, Leon "Khazar" and Leon Abkhazian Eristavi were cousins, and their mothers were sisters. S. Janashia supports this standpoint.

There is no chronological order before the governance of Leon in the "Divan of the Abkhazian Kings". The situation changes radically after the ruling period of Leon. All the dates are indicated after his governance. According to P. Ingorokva the governors of Abkhazia are the governors of Egrisi, too. The indication of Barnuki in the list supports this idea. Its connection with Patrikios Lebarnuk and with Sergi, the son of Barnuki, mentioned by the Byzantine chroniclers of the sixties and nineties of the seventh century is clear ("As Patrikios Lebarnuk told us under a dreadful oath"- evidence by Theodore of Gangra, exiled from Byzantium to Lazica, or by Theophanes, in 697 "Patrikios of Lazica, Sergi, son of Barnuk rebelled" – Georgica, IV, 1941, p.

49, 105.). M. Sanadze and T. Beradze share the same standpoint. According to the list presented by the "Divan of the Abkhazian Kings", the link between the Patrikios of Egrisi and the Abkhaz kings is evident. However, how much reliable is the evidence? Are the Abkhaz kings associated with the Patrikios of Egrisi artificially in order to indicate their heir rights over western Georgia? The Abkhaz Kingdom supported this idea in the period of Bagrat the III.

Let us return to the Chronicle of Kartli on the issue of Leon's predecessors. "As the power of Greeks lost strength, Leon the king of Abkhazia, whose kingdom was Abkhazia and owned it hereditarily, and who was the nephew of Leon Eristavi, broke away from Greeks. The second Leon, the son of the daughter of the Khazar King conquered Abkhazia and Egrisi to Likhi, and was called the King of Abkhazians."

There are two Leons in this context. One of them gave the other Abkhazia to rule over. S. Janashia stated the possibility that Leon obtained the rights on the governance of Abkhazia from his uncle, but with some doubts. Still he hardly believes that the uncle, who was considered the Eristavi of Abkhazia, existed in reality (S. Janashia, 1952, p. 326). The words "Leon, who was given Abkhazia, broke away from Greeks" belong to Leon. The same chronicle gives us information about Leon. This time he is in alliance with Curopalates Ashot and Teodosi: "the times when Ashot Curopalates attacked and was supported by Teodosi, the king of Apkhazeti, the son of the second Leon, who was the son-in-law of Ashot Curapalates" (p. 206). In this case the first part of the text determines the last part of it and this means that the son-in-law was Teodosi and not Leon II. According to the Chronicle, Abkhazia was given to Leon, who broke away from Greeks and called himself the King of Abkhazians. Who is Eristavi Leon, whose nephew is Leon the Abkhaz Eristavi? According to the author of Georgian Chronicles, this is the same person as the one mentioned by Juansher "Leon, Eristavi of Caesar" (p.193), the same Leon of Isauria. He was the most honorable person, who appointed the governors of Kartli to the position of Patrikios. But when and why did that happen?

It was noted several times that Juansher did not mention anything about the relations with Arabs in the seventh and eighth centuries. Chroniclers unite several events here. Separating them from each other is only possible if we compare different sources. According to the research of M. Sanadze and T. Beradze, Mir and Archil took the title of Patrikios for participating in the battles against Arabs and supporting the Byzantines. These events took place mostly in the first years of the governance of Leon of Isauria (717). As for Leon Eristavi, the scholars consider him as an honorable person in the 30s of the eighth century (737-739) (p. 76), the time when Marwan the Deaf's invasions ended. This is quite acceptable. There passed quite a long time between the two events and it was quite possible that the Byzantine Caesar appointed his brother's son or one of the relatives to this position. The same standpoint was shared by M. Artamonov. He thought that Abkhaz Eristavi Leon was the nephew of Byzantine Caesar. Nevertheless, M. Artamonov thinks that the Caesar was Leon Khazar (775-780) (M. Artamonov, 1962, p. 248). Probably it was not accidental that according to the evidence of the "Divan of the Abkhazian Kings" Leon's father's, Thevdore's father's as well as Byzantine Caesar's Leon of Isauria's son's name was Konstantine. If we take into account the attitude of Leon of Isauria towards western Georgia's governors, it is impossible for him to have trusted any representative of any branch of the Abkhazia and Egrisi Patrikios. It is hardly believable that Byzantium considered these sides the leading forces against Arabian attacks. In the beginning of the 8th century, when - according to the evidence of Theopanes - "Arabs were governing Abasgia, Lazica and Iberia" (Georgica, 1941, 105), Leon was sent to the Alans to pay them money and to use their force against Abasgians. Despite some obstacles, he partially succeeded in this. On his way back, Leon II named by the Chronicles as Leon the Eristavi, strongly refused to pass the territory of Abkhazia. With the help of Mir, he managed to come out of the territory of western Georgia and returned to Justinian (M. Sanadze, T. Beradze, 2004, p. 77.).

During his relations with Alans he got convinced that it was very profitable to build relations with Khazars. They could help in the battles against Arabs. Later he married his son to the daughter of the Khazar Khakan. According to the Georgian Chronicles, it is not accidental that one of his other relatives or his brother could have an alliance and matrimonial relations with the Khazar Khakan and his son could have been appointed as the Eristavi of Abkhazia. That is, at least, the perception of the author of the "Chronicle of Kartli". The chronicle that mentions the name as the second Leon, does not mean Leon II. There is no other case in the chronicle when the king is numbered. "Second" comes from the context. "Leon who broke away from Greeks and conquered Apkhazeti and Egrisi up to Likhi and who assumed the name the King of Apkhazians" was not Leon Eristavi, who was sent from Byzantium, but another Leon - his nephew. Wakhushti identified this part differently. From his standpoint, Leon, to whom Abkhazia was given, and Eristavi of Caesar Leon were the same person "The title of Abkhaz Eristavii was given to Leon hereditarily for serving the kings (as that Leon I Eristavi appointed by the Caesar)" (Wakhushti, 1973, p. 127). He gave the numbers to the Abkhaz Eristavs, Leon I and Leon II, who, by his wrong perception, ruled in the 8th century.

We think that the title of Eristavi was given to Leon by the Byzantine Caesar Leon of Isauria. This must have

happened before 741. But if we take into consideration that each of the given titles should have been confirmed by each Caesar, the standpoint of Wakhushti that an envoy was sent to the Caesar Konstantine could come from a reliable source.

When could Leon have joined Egrisi, move the capital city to Kutaisi, and secede from Byzantium?

After Archil agreed "to the marriage of Leon and his niece Gurandukht and gave him the crown, which was conferred to Mir by the king of Greeks", he thanked him and said: "I will go and settle down in the Fortress Goji and Kutaisi" (p.197). "Archil settled in Egrisi that extends to Shorapani" (p.198). At the time of Khazars' attacks, in the sixties of the 8th century, Archil was not alive. Kartli and Kakheti were ruled by Juansher, who was taken as hostage. Archil's son Ioane (p. 205) being in western Georgia together with his mother and sisters, was informed by Juansher on the desire of the Khakan to marry their sister. (M. Sanadze, 2001, p. 87-88.). After seven years, at the time when Juansher comes back, there is a rebellion against Arabs. The title of Erismtavari is abolished. Nerse is arrested, though he is freed after several years. The status of Saeristavo is renewed for a while. Later Nerse leaves Kartli. However, after 786 it appears all these people, including Ioane, are dead.

The cooperation of Abkhazia, Byzantium, and Khazar, quite expected in the period of the Byzantine Emperor Leon III, lasted during the governance of Konstantine the V, and his son Leon the IV "Khazar's" periods. The successful development of Byzantium's policy is the result of relations with Khazars, the researchers think (I. Chichurov, 1980, p. 11-12.). The embassy is sent from Byzantium to Khazar in 729. It is considered that Khazars attacks in 730 helped Byzantium to regain the territory of Abkhazia and preserved Lazica. Khazars played a very important role in the trade relations with Byzantium. Byzantium at the same time tried to spread Christianity in Khazaria. The same Byzantines strongly stood the attack of the Arabs. In 740 Leon the III and his son Konstantine won the battle with Acroin, Konstantine the V conquered Germanicia, the territory subjugated by Arabs in Syria, destroyed the Arab fleet near Cyprus. In 751-752 he ruined Melitena (in Mesopotamia) and Theodosiopolis (in Armenia), in 770 he defeated the Arabs in Asia Minor. Leon the IV waged three successful battles against Arabs.

The situation changed after the death of Leon the IV "Khazar" (780). Instead of underage Konstantine the VI the governor of the country was Irene, the mother of the king. In 787 the Christian population rebelled against Khazars in Crimea's Gutia. Byzantium did not take a special part in that battle to protect Christians; instead it was limited to founding only the Gutia Metropolis. In summer of 790 Irene defeated all the enemies and declared herself Empress. However, in December of the same year, the throne was taken by her son Konstantine. In 782 the Arabs started attacks and Byzantium gained peace by sending 70 000 Noems annually. In the eighties, after the defeat in a battle, Konstantine had to pay subsidies to the Bulgar Khan. Irene ordered to blind Konstantine, her son. And in 797, she took the throne once again. These events ended in a coup d'état, and the throne was taken by Caesar Nicephorus. That was the reason why S. Janashia considered that the secession of Abkhazia from Byzantium dates back to 797-802. However, it is quite possible that this fact happened earlier, as it is given in the evidence of Wakhushti, in 786.

Ioane Sabanisdze's "The Martyrdom of Saint Abo of Tiflis" written in the 8th century provides information on Abkhazia. This source proves that the relations between Khazars and Abkhazians were active and well-disposed. The given source "The Martyrdom of Saint Abo Tpileli" does not say that in the eighth decade of the 8th century there is a king in Abkhazia at all. The ruler of Abkhazia is known as *mtavari*, who extended hospitality to Nerse and all his people. According to Ioane Sabanisdze, "there was fear of the Saracens in the country" (Iovane Sabanisdze, 1982, p. 128) .

"In the period of Leon this region is absolutely free from Arabs. It is also possible to suppose that at that time Leon united western Georgia. But it can also be presumed that the source indicates the borders of Christianity, but not Abkhazia. Trebizond and Constantinople are also mentioned: "None of the non-believers can find place in it. The border of it is the Pontus Sea, the place of Christians to the border of Chaldea, where Trebizond, the place of Apsars and the port of Nafsai are included. And there are places and cities ruled by the servant of Christ, the King of Ioans, who resided in Constantinople" (pp. 128-129). "The Martyrdom of Saint Abo Tpileli" was written by the order of Samuel Catholicos (780-790) after the torture of Abo (786), it means no later than 786-790; however, it seems to have been written more immediately after the torture of Abo. It can hardly be presumed that Leon, who received Abkhazia from Leon of Isauria, could declare himself king in the ninth decade of the 8th century. Quite possibly, it happened earlier.

Proceeding from the foregoing, it is possible to assume that Leon united western Georgia in the 80s of the 8th century and declared independence from Byzantium. We do not agree to the information provided by Kevin Alan Brook that the strong and enlarged Abkhazia could be a colony of the Khazars at the end of the eighth century (K. Brook, 2005, p. 258). This attestation doesn't correspond to the earlier time either, when Abkhazia was governed by the grandson of the Khazar Khakan Leon. According to the analysis of the given information, we can conclude the following:

1. There was only one Leon in Abkhazia in the 8th century;

2. Leon was a family member neither of the Abkhazian, nor of the Egrisi (western Georgia) rulers;

3. Leon was the nephew of Leon of Isauria, the Byzantine Emperor, who appointed him as a ruler in Abkhazia;

4. Based on his wife's rights, Leon united Western Georgia and became its ruler in the late eighth decade of the 8th century. His wife was Guranduckht, daughter of Mir, to whom Egrisi was given hereditarily by Leon of Isauria;

5. Leon seceded from Byzantium and proclaimed himself King of Abkhazia that covered western Georgia, in the same late 8th century, closer to 786.

ისტორია & ფილოლოგია

აფხაზთა სამეფოს შემქმნელის ვინაობისათვის

გ. ალასანია*

* შავი ზღვის საერთაშორისო უნივერსიტეტი, თბილისი

(წარმოდგენილია აკადემიკოს თ. გამყრელიძის მიერ)

ქართულ ისტორიოგრაფიაში "აფხაზეთის სამეფოს" შემქმნელად მიჩნეულია ლეონ II. "ქართლის ცხოვრების" შესაბამისი მონაკვეთის, "მატიანე ქართლისაის", და ვახუშტის ჩვენებებს შორის ჩვენთვის საინტერესო შემთხვევაში შეინიშნება ზოგიერთი სხვაობა. პირველი, როგორც ჩანს, უფრო ზუსტია. მკვლევართა უმრავლესობა ცნობას ასე კითხულობს: "რაჟამს მოუძლურდეს ბერძენნი, გადგა მათგან ერისთავი აფხაზთა, სახელით ლეონ, ძმისწული ლეონ ერისთავისა, რომლისად მიეცა სამკვიდროდ აფხაზეთი. ესე მეორე ლეონ ასულის წული იყო ხაზართა მეფისა და ძალითა მათითა გაადგა ბერძენთა, დაიპყრა აფხაზეთი და ეგრისი". ორიე ლეონი მიჩნეულია აფხაზეთის მთავრებად: ლეონ I-დ და ლეონ II-დ. წინამდებარე სტატიის ავტორი სათანადო წყაროების განხილვით მიდის დასკვნამდე, რომ "მატიანე ქართლისაში" მოხსენიებული ერთ-ერთი ლეონი არის ლეონ ისავრიელი, ბიზანტიის იმპერატორი, რომელმაც დანიშნა თავისი ძმისშვილი ლეონი აფხაზეთის გამგებლად მე-8 საუკუნის 30-იან წლებში.

ვფიქრობთ, შესაძლებელია შემდეგი ვარაუდების გამოთქმა:

1. მე-8 საუკუნეში აფხაზეთში მმართველად მხოლოდ ერთი ლეონი იყო;

2. ლეონი არ ეკუთვნოდა არც ეგრისის, და არც აფხაზთა გამგებლების ოჯახს;

3. ლეონი იყო ბიზანტიის იმპერატორის ლეონ ისავრიელის ძმისშვილი, რომელიც ბიძამისმა ურჩი აფხაზეთის გამგებლად დანიშნა;

4. ლეონი აერთიანებს დასავლეთ საქართველოს და ხდება მისი გამგებელი მე-8 საუკუნის გვიან 80-იან წლებში – ცოლის უფლებებზე დაყრდნობით. ეს უკანასკნელი იყო გურანდუხტი, მირის ქალიშვილი, რომელსაც ეგრისის ტახტი სამემკვიდრეოდ იმავე ლეონ ისავრიელისაგან ჰქონდა მიცემული;

5. ლეონი განუდგა ბიზანტიას და გამოაცხადა თავი დასავლეთ საქართველოს მეფედ, ხაზართა ძალებზე დაყრდნობით, მე-8 საუკუნის იმავე გვიან 80-იან წლებში, ახლოს 786 წელთან, როდესაც ქართლის სამეფო ოჯახის წევრები გარდაცვლილი იყვნენ.

REFERENCES

Z. Anchabadze (1959), From the history of medieval Abkhazia, Sukhumi (in Georgian).
M. Artamonov (1962), History of Khazaria. Leningrad (in Georgian).
N. Asatiani (2004), History of Georgia, Tbilisi (in Georgian).
T. Beradze (2003), History of Georgia, Tbilisi (in Georgian).
N. Berdzenishvili (1990), Issues of Georgian History, Tbilisi (in Georgian).
M. B. Biro (1975), Marwan ibn Muhammad's Georgian Campaign. – Acta Orientalia Academiae Scientiarum Hungaricae. Tomus XXIX (3).
M. Biro (1977), Abo's Georgian vita. – Acta Orientalia Academiae Scientiarum Hungaricae, Tomus XXXI (2).
K.A. Brook (1999), The Jews of Khazaria, L.
M. Brosset (1850), Histoire de la Géorgie, I, S.-Petersbourg, p. 259.
I.S. Chichurov (1980), Byzantine historic chronicles. "Chronography" by Theophanes, "Breviarium" by Niciphorus. Text, translation, comments, M. (in Russian).
The Chronography of Gregory Abu'l Faraj commonly known as Bar Hebraeus. Translated from the Syriac by E. A. W. Budge, vol. I, Oxford, 1932.
From the "Memoirs" by Theodore of Gangra.- Georgica, Evidence of Byzantine Writers on Georgia, IV, p.I, Tbilisi 1941 (in Georgian).
J. Gamakharia, B. Gogia (1997), Abkhazia – Historic region of Georgia. Tbilisi (in Russian).
P. Golden (1980), Khazar Studies, vol. I, Akademiai Kiado. Budapest.
History of Caliphs by Vardapet Ghevond, writer of the 8th century, S.-Petersburg, 1862 (in Russian).
P. Ingorokva (1954), Giorgi Merchule. Tbilisi (in Georgian)
S. Janashia (1952), Concerning the time and conditions of the emergence of the Abkhaz kingdom. – Proceedings of TSU, II, Tbilisi (in Georgian).
I. Javakhishvili History of the Georgian Nation. – Proceedings, II (in Georgian).
"Kartlis Tskhovreba" (Georgian Chronicles), Ancient Armenian translation. Text edited, supplied with Introduction and Comments by I. Abuladze, Tbilisi, 1955 (in Georgian).
Kartlis Tskhovreba (Old Georgian Chronicles). Edited, supplied with Introduction, Explanations of Terminology by I. Antelava, N Shoshiashvili, (Tbilisi, 1966 (in Georgian)
M. Lordkipanidze (1973), Emerging new feudal principalities. – Essays on Georgian History, II, Tbilisi (in Georgian).
M. Lordkipanidze (1988), The Abkhaz kingdom. – Essays on Georgian History, II, Tbilisi (in Russian).
D. Muskhelishvili (1980), The main issues of the historic geography of Georgia. II, Tbilisi (in Georgian).
M. Sanadze (2001), The Erismtavaris Mir and Archil and "Kartlis Tskhovreba". – Meskheti, IV.
M. Sanadze, T. Beradze (2004), On the political history of Kartli and Egrisi in the first half of the 8th century. – Georgian Source Studies, X, Tbilisi (in Georgian).
Stanley Lain-Pool. Muslim dynasties, M., 2004 (in Russian).
E.Takaishvili (1983), What was the content of the so-called Abkhaz history of King Bagrat that was used by Jerusalem Patriarch Dositheos in his History of Jerusalem Patriarchs. – Ancient Georgia, Tbilisi, II, ed. E. Takaishvili (in Georgian).
Theophanes. Chronography. Georgica, IV, p. I (in Georgian).
C. Toumanoff (1956), Chronology of kings of Abasgia and other problems. – Le Muséon, XIX, 1-2, Louvain.
Wakhushti Bagrationi (1973), Description of the Georgian Kingdom. – Kartlis Tskhovreba, IV, Text edited according to the main manuscripts by S. Kaukhchishvili, Tbilisi (in Georgian).

Received September, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Semiotics & Linguistics

From Tokens to Writing: the Pursuit of Abstraction

Denise Schmandt-Besserat*

* *The University of Texas at Austin, USA*

(Presented by Academy Member Th.V.Gamkrelidze)

ABSTRACT. The development of civilization – the stage of cultural development at which writing is attained - required the acquisition of complex cognitive processes such as abstraction. In this paper I analyze the development of the capacity of abstraction in the ancient Near East between 7500-3000 BC as reflected by tokens and writing.

Key words: *tokens and writing, tokens and abstraction, tokens and cognition.*

COUNTING AND THE REDISTRIBUTION OF ECONOMY IN THE ANCIENT NEAR EAST

In the ancient Near East, agriculture was associated with the formation of a redistribution economy. Based on studies of modern archaic societies and of the pre- and proto-historic Near East, it can be assumed that the neolithic community leaders collected at regular intervals a share of the farmers' flocks and harvests. In turn, the accumulated communal goods were redistributed for the benefit of the group. Some were allocated to those who could not support themselves, but the greatest part was earmarked for the performance of rituals and festivals in honor of the gods. In other words, with agriculture came the need for counting and accounting in order to control and manage community surpluses.

Before analyzing their cognitive significance, I briefly describe tokens and writing, the two accounting systems created to compute entries and expenditures of goods in kind during the first four millennia of the Near Eastern redistribution economy. The earliest system with

Fig. 1. Plain tokens, Mesopotamia, present day Iraq, ca. 4000 B.C. The cone, spheres and disk represented various grain measures; the tetrahedron stood for a unit of labor. Courtesy Denise Schmandt-Besserat, The University of Texas at Austin.

tokens, appeared about 7500 BC. It consisted of counters, about 1-2 cm across, modeled in clay in multiple, often geometric shapes such as cones, spheres, disks, cylinders, tetrahedrons, ovoids, triangles and quadrangles (Fig.1). Some of them bore markings in the form of incised lines and impressed dots. In 7000 BC, because the system exclusively recorded goods, there were only some 10 token shapes, each representing one of the farm products levied at the time, such as grain, oil and domesticated animals.

The prehistoric administration used tokens to record three types of information. The shape and markings of the artifacts indicated, first, the type of goods computed, and second, the quantity thereof. For example, both the cone and the sphere stood for grain but in two different quantities probably equivalent to a "small" and a "large" basket of grain, and an ovoid with a circular incision represented a jar of oil. It should be well understood that, at the time, measures were not yet calibrated. They consisted of the traditional containers used to handle goods in everyday life, such as different-sized baskets, jars, juglets, bowls or cups. It is even conceivable that the tokens represented such casual units as an "armful" or a "handful." The system, therefore, only dealt with approximate quantities comparable to to-day's "carafe" of wine or "cup" of coffee. Lastly, the tokens recorded the number of units of goods received or dispensed in one-to-one correspondence. In other words, two small units of grain were shown with two cones, three cones stood for three small units of grain, and so on.

There can be no doubt that an unceasing cross-fertilization took place between the economy's increasing demands and the development of counting and accounting. For example, the number of token shapes increased to about 350 around 3500 BC, when urban workshops started contributing to the redistribution economy. Some of the new tokens, stood for raw materials such as wool and metal, and others for finished products among which textiles, garments, jewelry, bread, beer and honey. (Fig. 2) These so-called "complex" tokens sometimes assumed the shapes of the items they symbolized such as garments, miniature vessels, tools and furniture. These artifacts took far more skill to model compared to the former geometric shapes such as cones and spheres, suggesting that specialists were then manufacturing them. (Schmandt-Besserat 1992).

By 3300 BC, tokens were still the only accounting device to run the redistribution economy now administered at the temple by priestly rulers. The communal offerings in kind continued but the types of goods, their amount and the frequency of delivery to the temple became regulated, which meant that non-compliance was penalized. The response to the new challenge was the invention of envelopes where tokens representing a delinquent account could be kept safely until the debt was paid. The tokens standing for the amounts due were placed in hollow clay balls and, in order to show the content of the envelopes, the accountants created markings by impressing the tokens on the wet clay surface before enclosing them. (Fig. 3) The cones and spheres symbolizing the measures of grain became wedge-shaped and circular impressed signs. (Fig. 4) Within a century, about 3200 BC, the envelopes filled with counters and their corresponding signs were replaced by solid clay

Fig. 2. Complex tokens representing (above, from right to left) one sheep, one unit of a particular textile, one measure of honey and, one jar of oil. (Below, from right to left,) (?), one fleece of wool, one ingot of metal, from Susa, Iran, ca. 3300 B.C. Courtesy Musée du Louvre, Département des Antiquités Orientales, Paris.

Fig. 3. Envelope showing the imprint of three ovoid tokens with an incised line representing jars of oil, from Habuba Kabira, Syria, ca. 3300 B.C. Courtesy Museum für vor- und Fruhgeschichte, Schloss Charlottenburg, Berlin.

Fig. 4. Tablet showing the impression of spheres and cones representing measures of grain, from Susa, Iran, ca. 3300 B.C. Courtesy Musée du Louvre, Département des Antiquités Orientales, Paris.

tablets which continued the system of signs impressed with tokens. By innovating a new way of keeping records of goods with signs, the envelopes created the bridge between tokens and writing.

With the formation of city states, ca. 3200-3100 BC, the redistribution economy reached a regional scale. The unprecedented volume of goods to administer challenged writing to evolve in form, content and, as will be discussed later, in cognitive ability. First, about 3100 BC, the form of the signs changed with the use of a pointed stylus that sketched more accurately the shape of the most intricate tokens and their particular markings. The sign for oil, for example, clearly reproduced the ovoid token with a circular line (Fig. 5).

Fig. 5. Pictographic tablet featuring an account of 33 measures of oil, (circular = 10, wedges = 1) from Godin Tepe, Iran, ca. 3100 B.C. Courtesy Dr. T. Cuyler Young, Royal Ontario Museum, Toronto, CANADA

Second, plurality was no longer indicated by one-to-one correspondence. Numbers of jars of oil were not shown by repeating the sign for "jar of oil" as many times as the number of units to record. The sign for jar of oil was preceded by numerals – signs indicating numbers. Surprisingly, no new signs were created to symbolize the numerals but the impressed signs for grain took on a numerical value. The wedge that formerly represented a small measure of grain came to mean "1" and the circular sign, formerly representing a large measure of grain meant "10."

Third, about 3000 BC, the state bureaucracy required to enter on the tablets the names of the recipients or donors of the goods listed. And to record the personal name of these individuals, new signs were created that stood for sounds – phonograms. The phonograms were sketches of things easy to draw that stood for the sound of the word they evoked. The syllables or words composing an individual's name were written like a rebus. The drawing of a man stood for the sound "lu" and that of the mouth for "ka," that were the sounds of the words for "man" and "mouth" in the Sumerian language. For example, the modern name Lucas, could have been written with the two signs mentioned above "lu - ka" (Fig. 6).

The state administration could no longer deal with the approximate quantities of informal containers and this prompted the standardization of measures. The resulting adjustment in accounting was to assign new signs for the standard measures of grains (*ban, bariga* etc…), liquids (*sila*), and surface areas (*ikus, eshe*3, *bur*,

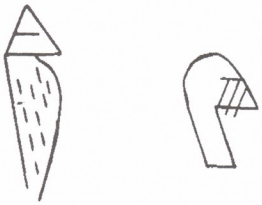


Fig. 6. Example of the rebus principle used to record names.

etc…). The standardization of measures brought accounting to an unprecedented precision, while putting an end to dealing with informal hand-manufactured containers (Fig. 7).

During four millennia and a half, from 7500 to 3000 BC, tokens and writing constituted the back-bone of the Near Eastern redistribution economy. Both recording systems were closely related in material, form, and function. They shared clay as a raw material; the token shapes were perpetuated by the written signs; both kept track of similar quantities of the same types of agricultural and industrial goods for an identical socio-economic function. The difference between the systems was cognitive, namely the degree of abstraction used to manipulate data.

TOKENS AND ABSTRACTION.

The cognitive principle at work in the token system was abstraction – i.e. the dissociation of one feature from a collection. Each token abstracted goods in two ways. First, their shape abstracted one of the types of merchandise levied; for example animals, grain or oil. Second, the shape abstracted the quantity of merchandise. The cone abstracted the unit corresponding to a *small* basket vs. the sphere that abstracted a *large* basket.

Otherwise the token system remained concrete in form, content and in the representation of plurality. (Malafouris)

- The tokens were solid and tangible artifacts. They could be grasped with the fingers and held in the hand.
- Each token stood for a concrete entity, namely, one unit of staple goods. Note that an ovoid token stood for a "jar of oil," and a sphere for a "large basket of grain" which means that, as in daily life, the product (oil, grain) and their usual container (jar, basket), were fused into a single concept.
- The token system dealt with plurality concretely, in one-to-one correspondence - as it is in the real world. In nature, a forest consists of a multitude of single trees; a flock is a set of single animals.

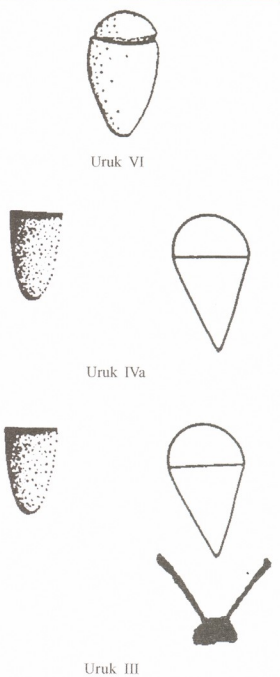


Fig. 7. The representation of "one *sila* of oil" in 3500 BC, 3100 BC and 3000 BC.

TOKENS AND COGNITION

The true cognitive significance of the token system was to foster the manipulation of data. Compared to oral information passed on from one individual to the other, tokens were extra-somatic, that is outside the human mind. As a result, the Neolithic accountants were no longer the passive recipients of someone else's knowledge, but they took an active part in encoding and decoding data.

The token system substituted miniature counters for the real goods which eliminated their bulk and weight and allowed dealing with them in abstraction by. As a result, heavy baskets of grains and animals difficult to control could be easily counted and recounted. The accountants could add, subtract, multiply and divide by manually moving and removing counters.

Patterning, the presentation of data in particular configurations, also promoted the abstraction of particular features. For example, the tokens representing the budget for a festival could be ordered in columns abstracting the merchandise according to its types, do-

nors, entries and expenditures, and intended use, i.e. for particular rituals. The relative value of merchandise could be abstracted by lining up units of greater value above those of lesser value. For instance, spheres, standing for large measures of grain could be placed above the cones, representing small measures of grain. It is well possible that the geometric lay out of operations such as adding two tokens to two tokens, and three tokens to three tokens, and so on, helped the conceptualization of abstract numbers (Justus 1999, 56, 64. Hoyrup 1994, 70).

Finally, because the clay tokens could be manufactured at will and stored indefinitely they abstracted goods from time. Consequently, accountants could manage merchandise independently of their current status. For instance, quantities of grain could be accounted for whether they were still in the fields or harvested, stored in granaries or in transit, delivered or promised.

In sum, the immense value of the token system was in promoting the acquisition of new cognitive skills that capitalized upon the visualization and physical manipulation of data. Computing with tokens in ever greater volume of more complex data paved the way to writing.

WRITING AND ABSTRACTION

Archaeology can interpret the technological innovations of the token system, such as the creation of new shapes and envelopes. But the cognitive dynamics that led writing to create logograms, numerals, standard units of measure and phonograms are far beyond the scope of traditional archaeology. These remarkable leaps in abstraction can be identified and dated to the early fourth millennium BC but their interpretation will have to wait until cognitive archaeology comes of age (Malafouris).

The early logograms, i.e. signs in the form of tokens standing for a unit of merchandise, represented a second degree of abstraction. The signs impressed or traced with a stylus, abstracted tokens, which were themselves abstracting actual goods. A circular marking replaced the round token, which means that the written signs kept the outline of the counters and their symbolic significance but did away with their volume. Intangible written signs replaced the awkward piles of three-dimensional tokens.

Written numerals abstracted the common denominator between sets such as *three* baskets of grain and *three* jars of oil. As a result, "*three*" became a concept that could be expressed by a sign. The invention of abstract numerals had extraordinary consequences. First, it put to an end dealing with goods in one-to-one correspondence. Second, numeral signs made obsolete the use of different counters or numerations (different number words) to count different products. Finally, with the abstraction of numbers counting had no limit.

About 3000 BC the abstraction of numbers (how many) was followed by that of quantity (how much.) Thereafter writing abstracted each of the concepts involved in for instance "one" "sila" of "oil," requiring a sequence of 3 signs for notation. Instead, a century earlier, in 3100 BC, two signs were sufficient to record a similar amount, namely, "one" "jar of oil," and in 3500BC, a single token fused the three concepts together "one jar of oil."

Finally, the invention of phonograms, that abstracted the sounds of speech, removed writing from the concrete world of real goods. The signs no longer referred to concrete objects, but instead to the sound of a word. This was the beginning of a phonetic script when, by emulating speech, writing was no longer confined to the recording of goods.

Of course, all these processes of abstraction innovated by writing, in particular that of numbers, were to take many steps to be fully realized. (Justus 1999a) It is clear that for many centuries the commodity counted still determined the arithmetical value of numerical signs. For example, when animals were being counted the circular sign signified "10" whereas it was to be read "6" when it referred to measures of grain. Also, as long as the cuneiform script existed, one-to-one correspondence continued to express the number of units such as "1" and "10." For example 33 jars of oil were expressed by three tens (three circular signs), three ones (three wedges) followed by the sign for "jar of oil" (Fig. 5).

The standardization of measures also progressed at a slow pace and, for a long time, the relation between units continued to vary with the kind of entities dealt with. For example, the units of grain (*ban, bariga* etc…) followed a sequence of factors: 5, 6, 10, 3, (Fig. 8) compared to 6, 3, 10, and 6 for the units of area measures (*ikus, eshe*3, *bur*, etc…). (Nissen, Damerow and Englund 64-65)

CONCLUSION

Between 7500-3000 BC, tokens and writing processed the data of the growing Near Eastern redistribution economy in ever greater abstraction. Each of the two accounting technologies, tokens and writing, documents one stage of the manipulation of data in abstraction. By abstracting units of real goods, the tokens could manage, one by one, a limited number of casual measures of selected staples. With the abstraction of tokens, numbers and measures, writing raised data management to limitless quantities of any possible unit of goods. Moreover, by abstracting sounds writing reached beyond accounting to take on new functions in communication.

სიმბოლოებიდან დამწერლობამდე: განზოგადების მცდელობა

დენიზ შმანდტ-ბესერატი*

* ოსტინის უნივერსიტეტი, ტექსასი, აშშ

(წარმოდგენილია აკადემიკოს თ. გამყრელიძის მიერ)

ცივილიზაციის განვითარება — კულტურის განვითარების ის საფეხური, როდესაც დამწერლობა ჩნდება — საჭიროებდა კომპლექსური შემეცნებითი პროცესების ათვისებას, როგორიცაა, მაგალითად, „განზოგადება". წინამდებარე სტატიაში ავტორი აანალიზებს „განზოგადების" უნარის განვითარებას ძველ ახლო აღმოსავლეთში ძვ.წ. 7500–3000 წ.წ., როგორც ეს აისახება ნიშან-სიმბოლოებისა და დამწერლობის წარმოშობაში.

REFERENCES

Hoyrup J. 1994. *In Measure, Number, and Weight*. New York: State University of New York Press.
Justus, C., 2004. On Language and the Rise of a Base for Counting. *General Linguistics* 42, 17- 43.
Justus, C. 1999a. Pre-decimal Structures in Counting and Metrology. In Gvozdanovic J. (ed), *Numeral Types and Changes Worldwide*: 55-79. Berlin: Mouton de Gruyter, 55-79.
Justus C., 1999b. Can a Counting System be an Index of Linguistic Relationship? *Journal of Indo-European Studies*, Monograph 33, 219-240.
Justus C. 1996. Numeracy and the Germanic Upper Decades. *Journal of Indo-European Studies* 23: 45-80.
Malafouris, L. forthcoming, Grasping the concept of number: How did the sapiens mind move beyond approximation, in Colin Renfrew, ed. *Measuring the World and Beyond*, The McDonald Institute for Archaeological Research, Cambridge, England.
Nissen, H. J., Damerow, P., Englund, R.K., 1993. *Archaic Book-Keeping*, Chicago: The University of Chicago Press.
Schmandt-Besserat D. 1996. *How Writing Came About*. Austin: The University of Texas Press.
Schmandt-Besserat D. 1992. *Before Writing*. 2vols. Austin: The University of Texas Press.

Received September, 2007

საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 175, №3, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 175, no. 3, 2007

Semiotics & Linguistics

The "Principle of Complementarity" and the Nature of the Linguistic Sign

Thomas V. Gamkrelidze*

* *Academy Member, G. Tsereteli Intitute of Oriental Studies, Tbilisi; Georgian National Academy of Sciences*

ABSTRACT. The nature of the verbal sign, as of the sign in any semiotic system, must be specified through the relationship between the «signans» and «signatum», as well as through the relations of the given sign to the other signs of the system, both at the level of the «signata» and at that of the corresponding «signantia». In defining a sign of a semiotic system, then, we must consider not only the 'vertical' relationship between the two components of a sign taken in isolation, but also the twofold 'horizontal' relations existing between the respective components of the interrelated signs. The Saussurean thesis of the arbitrariness of the sign is partial and incomplete in that it specifies only the 'vertical' relations, disregarding the nature of the 'horizontal' ones. On the other hand, the opposite views on the motivated, iconic nature of the bond between the «signans» and «signatum», as maintained by the adversaries of the Saussurean thesis, involve exclusively the sphere of the 'horizontal' relations and cannot refer to the 'vertical' relationship, which is characterized in principle by arbitrariness and conventionality. If the verbal sign is conceived of as a unity of the 'vertical' and 'horizontal' relations, the opposed propositions concerning the nature of the relations between the «signans» and «signatum» present themselves not as contradictory, but as complementary to each other (in Niels Bohr's sense of the term), specifying with necessary completeness the essence of the verbal sign. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *linguistic sign, principle of complementarity, signans~signatum.*

The question of the nature of the linguistic sign and of the character of relations which constitute its essence is one of the crucial problems of linguistics and of the general theory of sign systems – contemporary **semiotic** – whose main principles were outlined and advocated by the philosopher *Charles Sanders Peirce* and the linguist *Ferdinand de Saussure*. The data of linguistics, as one of the main and most advanced fields of semiotic, prove to be of vital importance in this respect for an overall characterization of sign systems and for the elucidation of the essence and governing laws of signs in general.

In his general classification of signs, Peirce distinguishes sharply between the material qualities of a sign – its «signans» – and an immediate interpretant of the sign – its «signatum».[1] The differences seen in the relationships between the «signans» and «signatum» give

[1] The concept of the sign as an indissoluble unity of two constituents – the signifier (*semainon*) and signified (*semainomenon*) – was current in Stoic theory, where the former was defined as 'perceptible' (*aistheton*) and the latter as 'intelligible' (*noeton*). In St. Augustine's writings – presenting, with Latinized terms, an adaptation and further development of the Stoic doctrine of the action of signs (semeiosis) – the signum (sign) was conceived of as an entity consisting of the signans (signifier) and signatum (signified). This formalized dichotomy between form and meaning, distinguishing 'the signifier' and 'the signified' in language, underlies the medieval philosophy of language, as well as the sign theories of Peirce and, in particular, of Saussure, whose terminologies are strikingly reminiscent of the Stoic doctrine (cf. Jakobson 1965:22-3, Robins 1967:16).

a basis for distinguishing among three cardinal sign-types:

(a) **icons**, i.e. signs in which the «signans» is characterized by a certain degree of factual similarity with the «signatum», being its formal reflection or repetition (e.g. the picture of an animal as an iconic sign of the animal itself);

(b) **indices**, i.e. signs in which the «signans» is attached to the «signatum» by a causal or any other relationship (e.g. factual contiguity, sequence in space and time), according to the principle 'aliquid stat pro aliquo'– something that serves to imply the existence of something else (e.g. smoke as a sign-index of a fire);

(c) **symbols**, i.e. signs in which the «signans» is linked with the «signatum» by convention, where this connection does not depend on the presence or absence of any resemblance or physical contiguity, being arbitrary and realized because of a certain rule or convention (Peirce 1931-35, vol. 2).[2]

The problem of the nature of the connection or relationship between the «signans» and «signatum», between the form of a word and its meaning, is one of the oldest problems of linguistics, and evoked fervent discussion even at the dawn of scientific linguistic thought: is the link between the form and the meaning of a word established 'by nature' (*phusei*) or by convention (*thesei*)? Essentially the same questions are involved, with an emphasis on the functional rather than genetical aspects of the problem, in attempts to define the character of the verbal sign in modern linguistic science (cf. Büchner 1936, Schneider 1930); and depending on which solution is adopted, there are, in the ensuing scholarly controversy, opposing views on the nature and essence of the verbal sign.

Among linguistic scholars who upheld the conventional character of the verbal sign, the American linguist William Dwight Whitney (1827-1894) should be mentioned first; his doctrine was later adopted and developed by Saussure (cf. Godel 1957:193), who declared that 'the bond uniting the signifiant with the signifie is arbitrary'. The principle of the arbitrariness of the sign (*l'arbitraire du signe*) is the first and one of the basic principles in the linguistic framework of Saussure. According to Saussure (1967:152), 'the principle stated dominates the science of language and its consequences are innumerable'.[3]

Since the appearance of Saussure's posthumous *Cours* in 1916, the thesis of the arbitrary sign–of the conventional character of the link uniting the «signans» (Saussure's signifiant) with the «signatum» (signifie), of the absence of any internal motivation between the sound form of a word and its meaning–has dominated the views of most linguists.[4] Agreement with this view on the nature of the verbal sign is, however, far from unanimous; and contemporary linguistic science includes rejoinders from the adversaries of the Saussurean position (cf. Engler 1962, 1964; Spang-Hanssen 1954; cf. also 'Materialy' 1969). Special mention should be made in this connection of the objections to the Saussurean thesis raised by such scholars as Benveniste (1939) and Jakobson (1965).

Contemporary theory of signs recognizes three different levels, representing different types of abstraction; *semantics* (the study of the relations between the «signans» and «signatum»), *syntactics* (the study of the relations between signs), and *pragmatics* the study of signs and the relations between signs), and pragmatics (the study of signs in relation to their users).[5] These may be fully applied to the study of linguistic systems and verbal signs (cf. Morris 1964:60-2; Mel'chuk 1968:426-7). At the same time, the inferences from a study of the system in its semantic and syntactic aspects may not coincide fully with the results of its study in the pragmatic aspect. The considerations advanced by Benveniste, as to the existence of an intimate, natural, and necessary link for a native speaker between the «signans» and «signatum», involve a transfer of the problem onto the pragmatic plane – evaluating a sign system as perceived and appreciated by its user, taking into account his concrete psychological associations. Hence the objections to the arbitrary sign put forth by Benveniste, referring wholly to the sphere of *pragmatics*, do not in principle affect the character of the relationship between the «signans» and «signatum» as studied in *semantics* and *syntactics*.

The nature of the verbal sign, as of the sign in any semiotic system, must obviously be specified not only through the relationships between its two components, but also through the relations of the given sign to the other signs of the system, both at the level of the «signata» and at that of the corresponding «signantia». In defining a sign of a semiotic system, then, we must consider not only the 'vertical' relations between the two components of a sign

[2] For further details, with a wide comparison of diverse semiotic systems, see Jakobson, 1970 a, b.

[3] For probable influences on the formation of Saussure's views, see Coseriu (1967), who gives a comprehensive critical account of the origin and historical developments, in diverse philosophical and linguistic theories, of the thesis of the arbitrary and conventional linguistic sign. According to Coseriu, this thesis, which was fairly current in philosophy and pre-Saussurean linguistics, should be traced back, directly or indirectly, to the Aristotelean doctrine of the historically established functional connection, 'kata syntheken', between sound and meaning in nomina.

[4] For a critical review of the relevant literature, with a utilization of Saussure's manuscript sources, see Engler 1962, 1964; cf. also 'Zeichen und System' 1961-62, Čikobava (1959:113-15) and Koerner 1972. The arbitrariness of the sign is regarded by some scholars as a linguistic universal: 'The relation between a meaningful element in language and its denotation is independent of any physical or geometrical resemblance between them' (Hockett 1963:8).

[5] Cf. Morris 1938, Cherry (1957:221).

taken in isolation, but also the 'horizontal' relations between the respective components of interrelated signs– i.e. the relations between the «signata» on the one hand, and the relations between the corresponding «signantia» on the other. The 'horizontal' relations, in contradistinction to the 'vertical' ones, are distinguished by their twofold character – implying, as it were, parallel relationships at the respective levels between the two components of the interrelated signs: the relations between the «signata» (on the plane of content) and the relations between the corresponding «signantia» (on the plane of expression). Schematically, the twofold nature of the 'horizontal' relationship may be represented as in Figure 1.

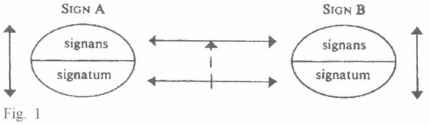


Fig. 1

Such a differentiated approach to the verbal sign and to the relations which constitute its essence, envisaging the linguistic sign jointly in the semantic and syntactic aspects, eliminates a number of objections put forth against the thesis of the arbitrariness of the sign.

The Saussurean thesis is incomplete in the sense that it specifies only the 'vertical' relations between the two components of the sign, disregarding the nature of the 'horizontal' relations. Since the verbal sign is a linguistic entity determined by both these types of relationships, any specification of the sign solely by one type of relation, leaving out of account the other type, is incomplete.[6]

At the same time, the opposite thesis of the motivated, iconic nature of the ties between the «signans» and its «signatum», as advanced against the Saussurean doctrine of arbitrariness, involves exclusively the sphere of the 'horizontal' relations between the signs. It cannot refer to the 'vertical' relationship, which is characterized in principle by the conventional, arbitrary nature of the bond between the components of the sign: a concrete «signatum» may be expressed by any phonemic sequence admissible by the phonotactic rules of a given language.[7]

This specific property of the 'vertical' relationship between the components of the sign is one of the factors of multilingualism–furthering, among other things, phonetic variability of language in diachrony.

Attempts at discovering in natural languages certain correspondences between sound and meaning, between the phonetic symbolism and the phonemic form of words, do not in general yield positive results. It is possible to establish, under experimental conditions, definite universal patterns of correspondences of sounds to meanings;[8] but these sounds are not distributed in the concrete words of natural languages according to the symbolic meanings expressed by them in isolation, under experimental conditions.[9]

In this respect especially, the 'vertical' relations are different in principle from the 'horizontal' relations between the components of the verbal sign. A specific feature of the 'horizontal' relations, as outlined above, is their twofold character implying parallel series of interdependencies, i.e. the relationships at the level of the «signata» as well as at the level of the corresponding «signantia» of the interrelated signs. Between these two series of 'horizontal' relations a certain correlation exists, finding its expression in the fact that specific relations at the level of the «signata» are reflected in the character of the relationships between the corresponding «signantia».

Thus, diverse relations between the «signata» on the *plane of content* (e.g. the relation of quantity, 'one' vs. 'many'; of proximity in space and time, 'near' vs. 'far','early'vs. 'late'; of size, 'large' vs. 'small'; the relation of similarity-dissimilarity; the relations of contiguity, kinship etc.[10]) are expressed in the corresponding

[6] In the part of the *Cours* dealing with linguistic value (valeur linguistique), Saussure does introduce the notion of mutual relations among signs which symbolize their signification; but his 'linguistic value' does not concern the nature of the relations between the components of the verbal sign, which he considers as a sign taken in isolation (cf. Engler 1962:46-9, 62).

[7] This does not concern the onomatopoeic forms (phonetically motivated) which constitute peripheral vocabulary, nor words with morphological or semantic motivation (cf. Ullman 1963:175-6). Only so-called primary signs are meant here, the ultimate constituents of the words – the 'prota onomata', according to the terminology of Plato's Cratylos – ' l'arbitraire absolu' according to Saussure (1967:297-303), in contradistinction to ' l'arbitraire relatif.

[8] Thus the vocoids which on the vocalic scale are close to a (compact vowels) tend to denote 'large', whereas the vocoids close to / (diffuse vowels) usually convey the meaning 'small'. In pairs of voiced ~ voiceless consonants under experimental conditions, the voiced is perceived as 'large' (cf. Sapir 1929, Chastaing 1964).

[9] Cf. Eng. big ~ small, Russ. velikij ~ malyj, Georgian didi ~ pat'ara etc. (cf. Ullman, 179). For sound symbolism, see also Gudava 1958, Ultan 1970, Nichols 1971.

[10] Such basic relations among the signata constitute the principal conceptual relationships of the content plane of the language, and make up its deep structure. They are characterized by generality and universality, in the sense that such semantic relations are present on the content plane of all languages, reflecting their basic, 'deep' relations.

Human beings, regardless of their language and culture, share a common meaning system, and organize their experience along similar symbolic dimensions (cf. Osgood 1963:244). One might draw up a typology of these basic conceptual relations which make up the deep structure of the plane of content and which underlie the diverse semantic relationships manifested in individual linguistic systems. Such a typology could serve as a semantic metasystem for investigating the transformation of the basic conceptual relations onto the plane of content of individual linguistic systems, and of their manifestation in the corresponding entities on the expression plane.

«signantia» by specific correlations of phonemic similarity, by phonemic alternations (juxtapositions), by phonemic length of the interrelated words, by certain syntactic features pertaining to the order of the elements, and by other universal properties of the *plane of expression* of language. Such relations at the level of the «signata» give rise to the specific character of the relations between the corresponding «signantia». These properties of the plane of expression show the dependence of the formal relations between the «signantia» on the relations between the corresponding «signata»; we may accordingly speak of a motivation of one series of relationships as being motivated by the other in the content plane of language. Such a dependence of the relations between the «signantia» on the relations between the corresponding «signata» does not, however, have the character of a diagrammatic correlation in Peirce's sense, in which the relations between the «signata» are iconically reflected in the nature of relations between the corresponding «signantia». It is rather manifested in the shape of conventional ties between these two series of relationships, which is expressed in a motivated appearance, caused by the character of the relations at the level of the «signata», of certain formal features specific for the level of «signantia» and not in principle characteristic of semantic entities (such as phonetic similarity, phonemic length, phonemic alternations etc.). The relations at the level of the «signata» are thus specifically projected onto the level of the corresponding «signantia».[11]

It is just in this sense that we must interpret the linguistic data adduced by Jakobson and other scholars (cf. Wescott 1971) maintaining the motivated character of the linguistic sign illustrating the existence of inner, iconic connections between the «signans» and «signatum»–in particular, between grammatical concepts and their phonological expression–as opposed to the Saussurean thesis of the arbitrariness of the sign. Thus the phonetic similarity of such paired numerals as Russ. *devjat'* 'nine' ~ *desjat'* 'ten', Svan *woštxw* 'four' ~ *woxwišd* 'five', Ger. *zwei* 'two' ~ *drei* 'three', Tigrina *šo'atte* 'seven' ~ *šommonte* 'eight', which came about as a result of an assimilation of one form to the other, is caused by the existence of a certain relationship at the level of the «signata»: the relation of immediate neighborhood between numerals is reflected in a partial phonetic resemblance between the corresponding «signantia».

The phonetic closeness of such kinship terms as Eng. *father* ~ *mother* ~ *brother,* or Fr. *père* ~ *mère* ~ *frère,* is a phonemic allusion to the semantic proximity of the corresponding «signata» standing in definite relationships to each other. Thus certain relations at the level of the «signata» are manifested at the level of the corresponding «signantia» by definite relationships of phonetic character specific to the plane of expression; the *plane of «signata»* calls forth a specific character of the formal relations on the *plane of «signantia».* In this sense alone can we speak of the motivation of one series of relations through the other, of the dependence of the relations between the «signantia» on those between the corresponding «signata». Very characteristic in this respect are groups of words with assonance pertaining to a definite semantic field, e.g. Germ. *Bube, Bursche, Bengel, Baby; Balg, Blage,* in which certain semantic relations among «signata» are exhibited in the corresponding «signantia» in a monotonous repetition of concrete phonemic units, establishing a specific phonetic relationship at the level of the «signantia».[12]

Linguistic facts from the realm of morphology, as dealt with by Jakobson (1965), must be interpreted in the same sense. Of particular interest in this respect is the correlation between singular and plural forms. There are languages in which the plural forms are distinguished from the corresponding singular by an additional morpheme, whereas there is scarcely a language with a reversed correlation, i.e. with a specific morpheme for singular vs. absence of an affix in the plural forms (Greenberg 1963b: 74). This correlation in phonemic length between singular and plural forms (shorter singular vs. longer plural forms) is a specific manifestation of the relationship of quantity ('one' vs. 'many') on the *plane of signata.* Note also the correlation of the short vs. longer verbal forms in the singular and plural, respectively, of French (1) *je finis* ~ *nous finissons,* (2) *tu finis* ~ *vouz finissez,* (3) *il finit* ~ *ils finissent;* or of Pol-

[11] Such a dependence of the relations between the signantia on the relations between the corresponding signata (this being a universal property of the interrelated signs) is not eliminated in the process of diachronic change. Phonetic changes in the interrelated verbal signs are apparently realized not as arbitrary and isolated sound shifts, but as a regulated process of interdependent phonetic transformations aimed at preserving the general character of the formal relationship between the signantia motivated by the relations between the corresponding signata. The general character of the formal relationship between the signantia remains, on the whole, invariant under such phonetic transformations, whereas the specific forms of phonemic expression of this relationship may vary considerably in the process of diachronic change. The principle of motivation of the relations between the signantia by the relations between the corresponding signata thus imposes certain regulating limits on the phonetic changes in the interrelated signs. (Cf., in contradistinction to this thesis, Saussure's view on the unlimited character of phonetic changes – this being a corollary to the principle of the arbitrariness of the sign; cf. Saussure 1968:344).

[12] Cf. Deeters (1955:31). For such constellations of words having similar meanings tied to similar sounds (e.g. Eng. *twist, twirl, tweak, twill, tweed, tweeze, twiddle, twine, twinge*), see Bolinger (1950:117-36). Interesting linguistic data are to be found also in Syromjatnikov 1972.

ish (1) *znam* ‘ I know’ ~ *znamy,* (2) *znasz* ~ *znacie,* (3) *zna* ~ *znają.* Such correlations between the «signata» and corresponding «signantia» underlie the whole morphological structure of language (cf. the correlation of different degrees of adjectives in Indo-European and other languages, as outlined by Jakobson.) In such forms, the relation of the growing intensity of a quality is reflected at the level of the «signantia» by a gradual increase in phonemic length of the corresponding adjectives: Lat. *altus* ~ *altior* ~ *altissimus;* Eng. *high* ~ *higher* ~ *highest;* Georg. *maγali* ‘high’ ~ *umaγlesi* ‘highest’, etc. (cf. Jakobson 1965:30).

These and similar examples, the number of which could be considerably increased by instances from diverse languages, testify to the existence of a specific dependence in the sphere of the *‘horizontal’ relations*, i.e. the dependence of the relations between the «signantia» on those between the corresponding «signata». Such a dependence is, however, not of an iconic nature (iconicity is characteristic of various types of diagrammatic representations). It is in principle conventional, in the sense that the specific formal relations between the «signantia», though conditioned and motivated by the character of the relations between the corresponding «signata», do not reflect them iconically; on the one hand, we have *relations of quantity*, of the *intensity of a quality*, of *immediate neighborhood,* etc.; on the other, we have relations of *phonemic length* and *phonetic similarity* between the corresponding «signantia».[13]

We may consequently conclude that the dependence of the *plane of expression* on the *plane of content* as regards linguistic signs is confined exclusively to the sphere of ‘horizontal’ relations, and does not extend to the sphere of ‘vertical’ relations, which are characterized in principle by arbitrariness and lack of motivation.

Along these lines, apparently, we must interpret the famous experiments with figures designated by the ‘words’ *takete* and *maluma* (Köhler 1947:224-5).

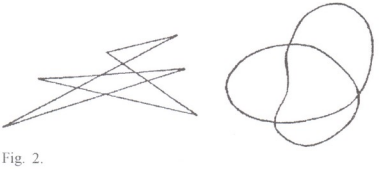


Fig. 2.

The motivated character of the link between the «signans» and «signatum», as revealed in this and similar psychological experiments, does not involve the ‘vertical’ relationship; it must be referred exclusively to the ‘horizontal’ relations between the signs:

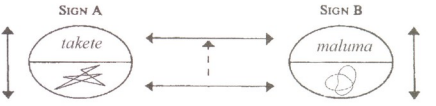


Fig. 3.

In pathology such *horizontal relations* between signs may be disrupted and, as a result, we get *“diagonal”* relations out of the normal *“horizontal”* ones, upon the elimination of *“vertical”* relations (Buachidze, 1984):

Fig. 4.

Such *“diagonal” relations* may illustrate also semantic shifts in historical development of language. What is manifested as pathology in synchrony may reflect the process of horizontal semantic development in diachrony.

The *thesei-phusei* problem of the verbal sign, when envisaged separately on the plane of the ‘vertical’ or on the plane of the’ horizontal’ relations, is characterized by incompleteness, and leads necessarily to contradictory conclusions as to the nature of the linguistic sign and the character of the relations between the «signans» and «signatum». When the linguistic sign is conceived of as a unity of the ‘vertical’ and the ‘horizontal’ relations, the opposed propositions concerning the nature of the relations between the «signans» and «signatum» present themselves not as contradictory but as complementary to each other, specifying with necessary completeness the essence of the verbal sign. These propositions – each true, but partial when considered in isolation with respect to the entity under consideration – are in the type of correlation called ‘*complementarity*’ by Niels Bohr.[14]

[13] Elements of a diagrammatic correspondence in linguistic signs may be seen in the realm of syntax, where the linear sequence of the members of a syntactic group may to a certain extent be in iconic dependence on the relations of succession or rank of their referents. E.g., the chain of verbs *vēni, vīdi, vīci*, in Julius Caesar’s famous aphorism, reproduces iconically the succession of interrelated events; or, in *Le Président et le Ministre prirent part à la réunion*, the sequence of the syntactic members *Président* ~ *Ministre* reflects the hierarchical order of the referents (Jakobson 1965:27; cf., however, Bolinger 1968:16).

[14] For the principle of complementarity, see Bohr 1948, 1958. For the application of the principle to linguistic problems, cf. Jakobson 1971:689-90.

სემიოტიკა და ენათმეცნიერება

"დამატებითობის პრინციპი" და ენობრივი ნიშანი

თამაზ გამყრელიძე

აკადემიის წევრი, გ. წერეთლის აღმოსავლეთმცოდნეობის ინსტიტუტი

ენობრივი ნიშნის ბუნება, ისევე, როგორც რომელიმე სხვა სემიოტიკური სისტემის ნიშნისა, უნდა განისაზღვროს „აღმნიშვნელსა" და „აღსანიშნს" შორის არსებული მიმართებებით, აგრეთვე მოცემული ნიშნის ამავე სისტემის სხვა ნიშნებთან მიმართებებით როგორც „აღსანიშნთა", ასევე შესაბამის „აღმნიშვნელთა" დონეებზე. ამგვარად, სემიოტიკური სისტემის ნებისმიერი ნიშნის განსაზღვრისას ჩვენ უნდა მივიღოთ მხედველობაში არა მხოლოდ „ვერტიკალური" მიმართებები, რომლებიც არსებობს ცალკე აღებული ნიშნის სათანადო კომპონენტებს შორის, არამედ უნდა გავითვალისწინოთ აგრეთვე ორმაგი „ჰორიზონტალური" მიმართებები, რომლებიც არსებობს სისტემის ურთიერთდაკავშირებული ნიშნების სათანადო კომპონენტებს შორის.

ფერდინანდ დე სოსიურის თეზისი ენობრივი ნიშნის „პირობითობის" შესახებ ნაწილობრივია და არასრული, რამდენადაც იგი ითვალისწინებს მხოლოდ „ვერტიკალურ" მიმართებებს და მთლიანად უგულებელყოფს ნიშნებს შორის არსებულ „ჰორიზონტალურ" დამოკიდებულებებს. მეორე მხრივ, „პირობითობის" საწინააღმდეგო თვალსაზრისი „აღსანიშნსა" და „აღმნიშვნელს" შორის არსებული „მოტივირებულობისა" და „იკონიკური" დამოკიდებულების შესახებ, რაზედაც *სოსიურის* თეზისის მოწინააღმდეგენი ლაპარაკობენ, გულისხმობს მხოლოდ „ჰორიზონტალურ" მიმართებათა სფეროს და ვერ შეეხება „აღსანიშნსა" და „აღმნიშვნელს" შორის არსებულ „ვერტიკალურ" მიმართებებს, რომლებიც ხასიათდებიან პრინციპში სრული „პირობითობით".

რამდენადაც ენობრივი ნიშანი უნდა განიხილებოდეს როგორც სისტემის ნიშნებს შორის არსებულ „ვერტიკალურ" და „ჰორიზონტალურ" მიმართებათა ერთობლიობა, საწინააღმდეგო დებულებები იმ მიმართებათა შესახებ, რომლებიც არსებობს „აღმნიშვნელსა" და „აღსანიშნს" შორის წარმოგვიდგება არა როგორც „კონტრადიქტორული" პროპოზიციები, არამედ როგორც „დამატებითი" მტკიცებულებანი („დამატებითობის" *ნილს ბორისეული* გაგებით), რომლებიც აუცილებელი სისრულით ახასიათებენ ენობრივი ნიშნის ბუნებას.

REFERENCES

Benveniste, Émile. 1939. Nature du signe linguistique. Acta Linguistica 1.23-9. (Reprinted in Readings in Linguistics, vol. 2, ed. by Eric P. Hamp, Fred W. Householder, and Robert Austerlitz, 104-8. Chicago, 1966.)

Bohr, Niels. 1948. On the notions of causality and complementarity. Dialectica 2.312-19.

Bohr, Niels. 1958. Atomic physics and human knowledge. New York: Wiley.

Bolinger, Dwight. 1950. Rime, assonance, and morpheme analysis. Word 6.117-36. – 1968. Aspects of Language. New York: Harcourt, Brace & World.

Buachidze, M. (1984), Iskaženije otnošenij meždu 'oznachajuščim' i 'označaemym" v lingvističeskix znakax pri aphazii. Bulletin of the Academy of Sciences of the Georgian SSR, №2: 429-432.

Büchner, K. 1936. Platons Kratylus und die moderne Sprachphilosophie. Berlin.

Chastaing, M., 1964. Nouvelles recherches sur le symbolisme des voyelles. Journal de Psychologie Normale et Pathologique 1.75-88.

Cherry, Colin. 1957. On human communication: a review, a survey, and a criticism. Cambridge, Mass.: MIT Press.
Čikobava, Arnold. 1959. Problema jazyka kak predmeta jazykoznanija. Moscow.
Coseriu, Eugenio. 1967. L'arbitraire du signe: zur Spätgeschichte eines aristotelischen Begriffes. Archiv für das Studium der Neueren Sprachen und Literaturen 204.81-112.
Deeters, Gerhardt. 1955. Gab es Nominalklassen in allen kaukasischen Sprachen? Corolla Linguistica, Festschrift Ferdinand Sommer, ed. by H. Krahe, 26-33. Wiesbaden: Harrassowitz.
Engler, Rudolf. 1962. Théorie et critique d'un principe saussurien: l'arbitraire du signe. Cahiers Ferdinand de Saussure 19.5-66.
—— 1964. Compléments à l'arbitraire. Cahiers Ferdinand de Saussure 21.25-32.
Godel, Robert. 1957. Les sources manuscrits du Cours de linguistique générale de F. de Saussure. Genève: Droz.
Greenberg, Joseph H. 1963a. Universals of language, ed. by ... Cambridge, Mass: MIT Press.
—— 1963b. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, 1963a: 58-90.
Gudava, Togo. 1958. Ob odnom vide zvukopodražanija v megrel'skom dialekte zanskogo jazyka. Tezisy Dokladov XVI Naučnoj Sessii Instituta Jazykoznanija AN GSSR. Tbilisi.
Hockett, Charles F. 1963. The problem of universals in language. In Greenberg, 1963a:1-22.
Jakobson, Roman. 1965. A la recherche de l'essence du langage. Diogène 51.22-38.
—— 1970a. Language in relation to other communication systems. Linguaggi nella Società e nella Tecnica, 3-16. Milano. (Reprinted in Jakobson's Selected Writings 2.697-708, 1971.)
—— 1970b. Linguistics. Main trends of research in the social and human sciences, 1: social sciences, 419-63. Paris & The Hague: Mouton/Unesco.
—— 1971. Linguistics in relation to other sciences. Selected Writings, 2.655-95. The Hague: Mouton.
Köhler, Wolfgang. 1947. Gestalt psychology: an introduction to new concepts in modern psychology. New York: Liveright.
Koerner, E. F. K. 1972. Contribution au débat post-saussurien sur le signe linguistique: introduction générale et bibliographie annotée. The Hague: Mouton.
Materialy ... 1969. Materialy seminara po probleme motivirovannosti jazykovogo znaka. Leningrad.
Mel'čuk, Igor'. 1968. Stroenie jazykovyx znakov i vozmožnye formal'no-smyslovye otnošenija meždu nimi. Izvestija Akademii Nauk SSSR, serija literatury i jazyka, 28.426-38.
Morris, Charles. 1938. Foundations of the theory of signs. International Encyclopaedia of Unified Science Series, 1:2. Chicago.
—— 1964. Signification and significance. A study of the relations of signs and values. Cambridge, Mass.: MIT Press.
Nichols, Johanna. 1971. Diminutive consonant symbolism in Western North America. Lg. 47.826-48.
Osgood, Charles E. 1963. Language universals and psycholinguistics. In Greenberg, 1963a:236-54.
Peirce, Charles Sanders. 1931-35. Collected Papers. Cambridge, Mass.: Harvard University Press.
Robins, R. H. 1967. A short history of linguistics. London: Longmans, Green.
Sapir, Edward. 1929. A study in phonetic symbolism. Journal of Experimental Psychology 12.225-39.
Saussure, Ferdinand de. 1967-68. Cours de linguistique générale. Edition critique par Rudolf Engler. Wiesbaden: Harrassowitz. (Fasc. 1-2, 1967; fasc. 3, 1968.)
Schneider, W. 1930. Über die Lautbedeutsamkeit. Zeitschrift für Deutsche Philologie 63.138-79.
Spang-Hanssen, Henning. 1954. Recent theories on the nature of the language sign. Travaux du Cercle Linguistique de Copenhague, 9.
Syromjatnikov, N. A. 1972. Opredelenie rodstvennosti kornej. Voprosy Jazykoznanija, 2, 109-23.
Ullmann, Stephen. 1963. Semantic universals. In Greenberg 1963a: 172-207.
Ultan, Russell. 1970. Size-sound symbolism. Working Papers on Language Universals 3.1-31 (Stanford University).
Wescott, Roger W. 1971. Linguistic iconism. Lg. 47.416-28.
Zeichen und System der Sprache. 1961-62. Veröffentlichungen des I. International Symposions 'Zeichen und System der Sprache' vom 28.9 bis 2.10.1959 in Erfurt. 2 vols. Berlin.

CORRECTION:

Semiotics & Linguistics. In the article "***On Agreement in Udi***" by Alice C. Harris that appeared in the issue vol. 175, no.1, 2007, some paragraphs contained misprints in citing examples, which were introduced during the production of the article. The correct version is reproduced here (cf. also the online version of the article):

Page 195:

It has been known that the PM may occur (i) between morphemes in the verb, as in *ač-es-**ne**-b-e* [lose-INF-3SG-AORII] 'he lost it', (ii) word-finally in the verb, as in *eč-al-**zu*** [bring-FUT**II**-1SG] 'I will bring', (iii) enclitic to other words in the sentences, as in *turin nik'o-**ne** ači-p-e* [foot ball-3SG play-LV-AORII] 'he played football', or even inside the verb root, as in *e-**ne**-f-esa* 'he keeps it', where the PM *-ne-* [3SG] splits the morpheme *ef-* 'keep'.

...Pančvizɛ 1974,...

(1) sa ait xabar-aq'-al-**zu**
one word ask-take-FUTII-1SG[3]
'I will ask one word.'

The clitic *-zu* 'I' is enclitic to the verb in (1).

Page 196:

Negative particles and question words are always in focus.

(2) juɣab-a te-**ne** ta-d-e.
answer-DAT NEG-3SG give-LV-AORII
'He did not give an answer.'

(3) ma-**z** ta-ɣ-o?
WHERE-1SG go-LV-FUTI
'Where shall I go?'

(4) gölö-**q'un** ta-ɣ-o, k'ic'i-**q'un** ta-ɣ-o.
much-3PL go-LV-FUTI little-3PL go-LV-FUTI
'They traveled a lot, they traveled a little.'

In these examples the PMs *-ne* 'he, she, it', *-z* '**I**', and *-q'un* 'they', respectively, are enclitic to the focused constituents - the negative *te,* the question word *ma* 'where', and the words *gölö* 'much' and *k'ic'i* 'little'. *(Ta-* in all of these verbs means 'thither', but I have glossed it according to the meaning of the whole lexeme.)

(5) uk'-a-i-n p'uran mand-a nu irazi bak-a-**n**
say-SUBJVI-PAST-2SG again stay-PURP NEG agreement.NOM be-SUBJVI-2SG
'You should say you will not agree to stay again.'.

(6) gorox-**ne** me čubux.
pitiful-3SG this woman
'This woman [is] pitiful.'

Example (7) illustrates rule 4a, and example (8) 4b.

(7) ...nišan-ax-al e-č-er-i ak'-es-**nan**-d-o.
sign-DAT-AND bring-LV-R-PTCPL see-INF-2PL-CAUS-FUTI
'...And having brought the sign [of it], you will show it to me.'

(8) čubux čomo[x] t'oɣol čur-p-i imux-**ne**-lax-o.
woman.NOM door near stand-LV-PTCPL ear-3SG-put-FUTI
'The woman, standing near the door, listened'.

In (8) we have the light verb *lax*; the agreement clitic *-ne* occurs between the incorporated element, *imux* 'ear', and this light verb.

The conditions on rule 4b are met by the complex verb *käi-b-es* 'to spend the night', containing the light verb *b-* 'do' and the incorporated element *käi* 'dawn'. The conditions on rule 4a are met by the light verb *-t'* and its incorporated element, the infinitive *käi-b-es.*

(9) aginam ba-ne-k-o me šu käi-b-es-t'-a-**nan**
if be$_1$-3SG-be$_2$-FUTI this night dawn-LV-INF-CAUS-SUBJVI-2PL
'If it is possible that you let me spend the night tonight...'

In Udi, most verbs are complex, formed with a light verb; there are only about forty verbs with monomorphemic stems (Harris 2002: 64-93). (10) illustrates the application of rule 7 with the monomorphemic root *bey-* 'look'. The third person singular agreement clitic, *-ne,* is situated inside this morpheme.

(10)	šel	be-**ne**-y-o	čubɣ-o	laxo
	well	look$_1$-3SG-look$_2$-FUTI	woman-DAT	on
	'He looked carefully at the woman.'			

Page 197:

...*ak'-es-d-* 'show, cause to see')
Rule 6. b- ... *č'e-re-ne* 'she went out'...

Notes ...Tsatso Čik'vaidze...

Subscription Information:

Bulletin of the Georgian National Academy of Sciences
is published quarterly

Correspondence regarding subscriptions and manuscript publications should be sent to:

Georgian National Academy of Sciences,
52, Rustaveli ave, 0108 Tbilisi, Georgia
Phone:+995-32 99-75-93
Fax/Phone: +995-32 99-88-91
E-mail: Bulletin@science.org.ge

The online edition of the *BULLETIN* is available
from the website of the Academy at:
http://www.science.org.ge

რედაქციის მისამართი: 0108, თბილისი-8, რუსთაველის პრ. 52. ტელ.: 99-75-93

INDEX 76181

GEORGIAN NATIONAL ACADEMY OF SCIENCES

0108 Tbilisi, Georgia
52 Rustaveli ave
Tel.: (+995 32) 99 88 91;
Fax/Tel.: (+995 32) 99 88 23
E-mail: academy@science.org.ge

საქართველოს მეცნიერებათა ეროვნული აკადემია

0108 თბილისი,
რუსთაველის პრ.ს. 52
ტელ.: 99 88 91; ფაქსი: 99 88 23
ელ. ფოსტა: academy@science.org.ge