Vladimir Polyakov

NEW APPROACHES TO LANGUAGE SIMILARITY MEASURES

The Swadesh Centenary Conference, Leipzig, January 17-18, 2009

1. Introduction in the DB JM

- JM is the new tool for linguistic and cognitive researches
- It allows to carry out researches by new quantitative techniques in typology, historical and areal linguistics
- It allows to receive scientific results in the field of modeling of evolution of languages
- It allows to spend diachronic researches on the fact sheet in sphere of an origin of language and its evolution

2. Source of Data for DB JM

- Encyclopaedic issue "Jaziki Mira"(Languages of the World) – 14 volumes, printed by Institute of Linguistics of Russian Academy of Sciences from 1993 to 2006.

- Large Encyclopaedic Dictionary. Linguistics (Edited by Yarceva V.N.) – includes interpretation of all terms of model of DB.

Main work on language description in DB format was fulfilled by Yelena Yaroslavceva, DSc.

3. List of Encyclopaedic Publications "Jaziki Mira"(Languages of the World)

- Languages of the world: Uralic (1993).
- Languages of the world. Paleoasiatic languages. Moscow: Publ. "Indrick". (1996). - 231 p.
- Languages of the world: Turkic. Moscow: Publ. "Indrick". (1997). - 544 p.
- Languages of the world: Mongolic languages. Manchu-Tungus languages. Japan. Korean. (Ed.: Kibrik A.A., Rogova N.B., Romanova O.I.). Moscow: Publ. "Indrick". (1997). - 408 p.
- Languages of the world: Iranian languages. I. South-Western Iranian languages. Moscow: Publ. "Indrick". (1997). - 207 p.
- Languages of the world: Iranian languages. II. North-Western Iranian languages. Moscow: Publ. "Indrick". (1999). – 302 p.
- Languages of the world: Dardic and Nuristani languages. Moscow: Publ. "Indrick". (1998). - 143 p.
- Languages of the world: Iranian languages. III. East Iranian languages. Moscow: Publ. "Indrick". (1999). - 343 p.
- Languages of the world: Germanic languages. Celtic languages. Moscow: Publ. "Academia". (1999). - 472 p.
- Languages of the world: Caucasian languages. RAS. Institute of Linguistics. Moscow: Publ. "Academia". (2001).-480 p.
- Languages of the world: Romance languages. Moscow: Publ. "Academia". (2001). - 720 p.
- Languages of the world: Indo-Aryan languages of Ancient and Middle Period. Moscow: Publ. "Academia". (2004). - 160 p.
- Languages of the world: Slavonic languages. RAS. Institute of Linguistics. /Ed. A.M. Moldovan, S.S. Skorvid, A.A. Kibrik/ Moscow: Publ. "Academia". (2005). - 656 p.
- Languages of the world: Baltic languages. RAS. Institute of Linguistics. /Ed. V.N.Toporov, M.V.Zavyalov, A.A. Kibrik /. Moscow: Publ. "Academia". (2006), 224 p.

4. Characteristics of Data Base "Languages of the World" Content

The Data Base "Languages of the World" has the following quantitative characteristics.

- contains more than 3800 features

- the number of languages is 315 Eurasian languages

- contains the description of the following spheres of language: phonetics, morphology, syntax.

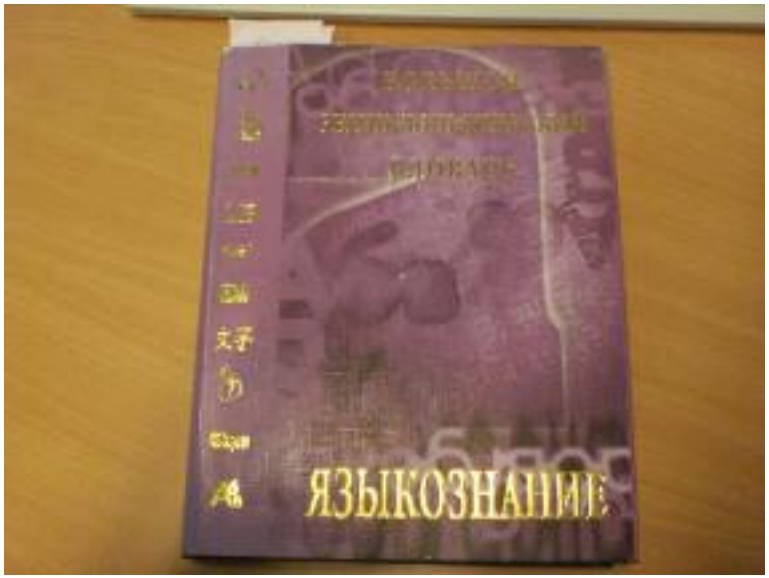
- representation of data: binary

In Data Base "Languages of the World" the following language families and unities are represented: Austroasian, Austronesian, Altaic, Afroasian, Indoeuropean, Caucasian, Paleoasian, Sinotibetic, Uralic, Hurrito-Urartean. DB contains the description of languages-isolates: Ainu, Nivch, Burushaski, Sumeran, Elamite. The unique peculiarity of Data Base "Languages of the World" is a large collection of extinct languages description, that includes 55 essays. There is no analogues of such detailed and systematic description of exinct languages.

The main principles forming of the model of language description are binarity, hierarchicity and paradigmaticity.

4.1. Areal of languages covered by JM (from Andrey Kibrik's report on CML-2009)

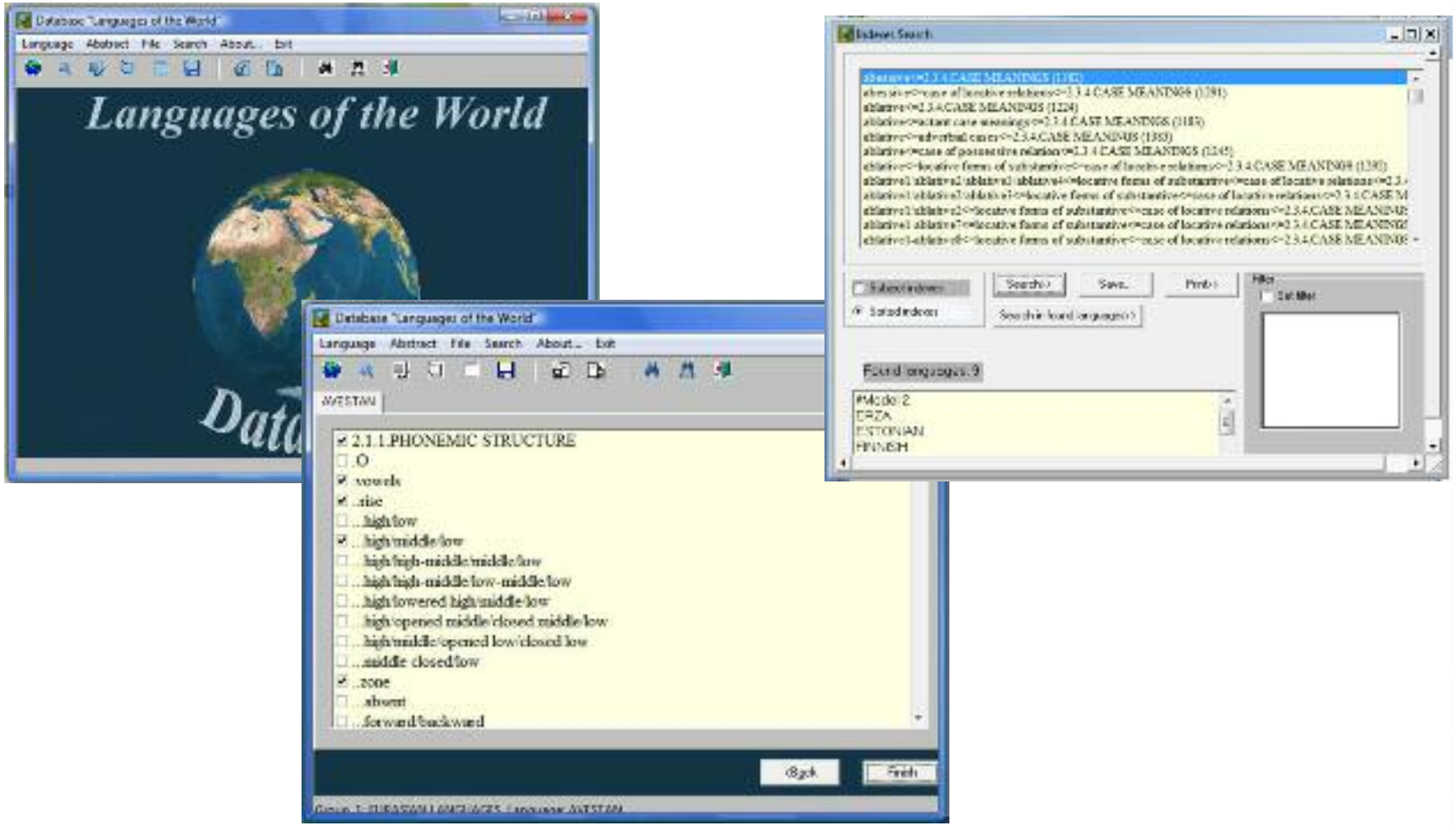
5. Dictionary and source books



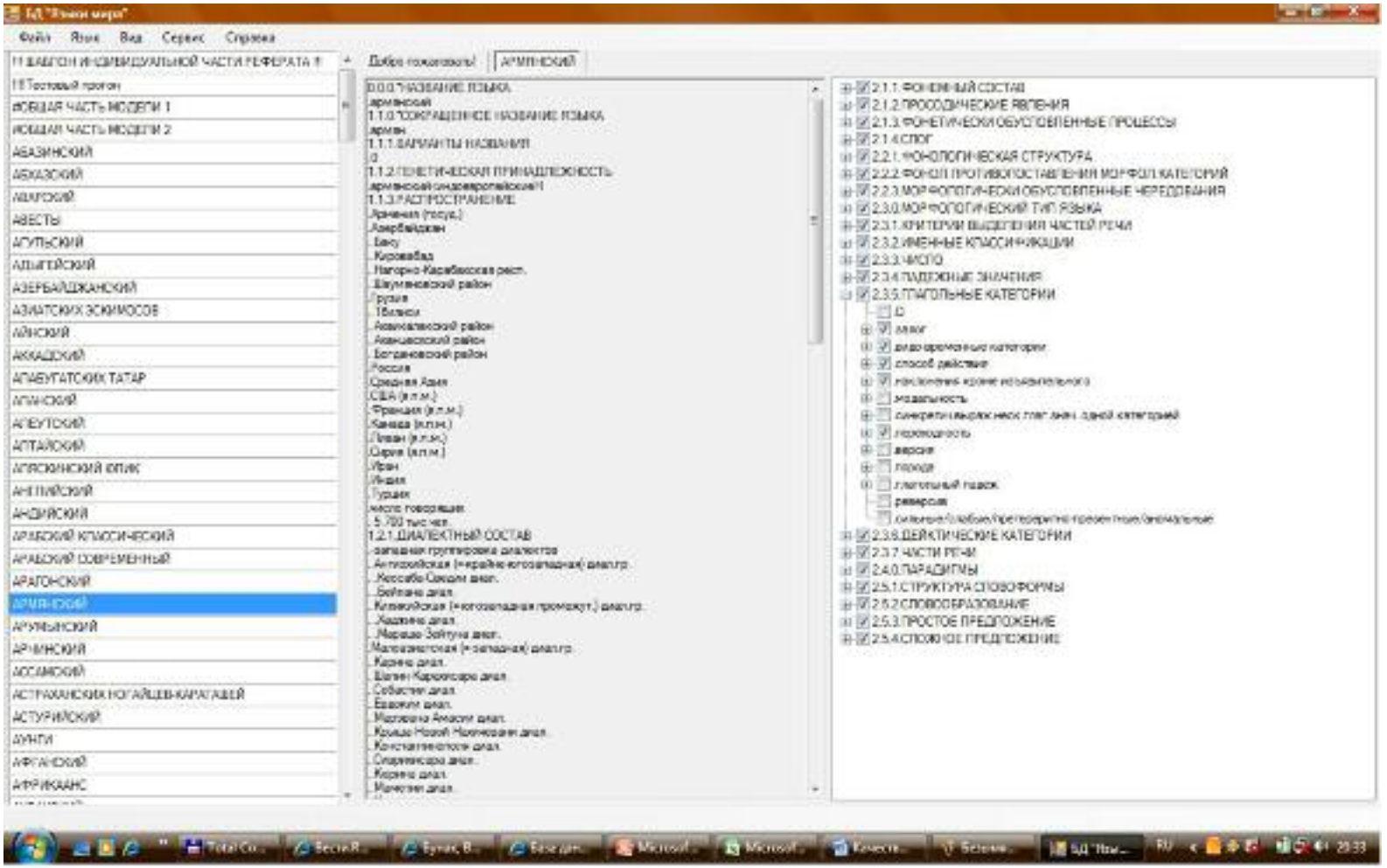
Dictionary

Two of 14 source books

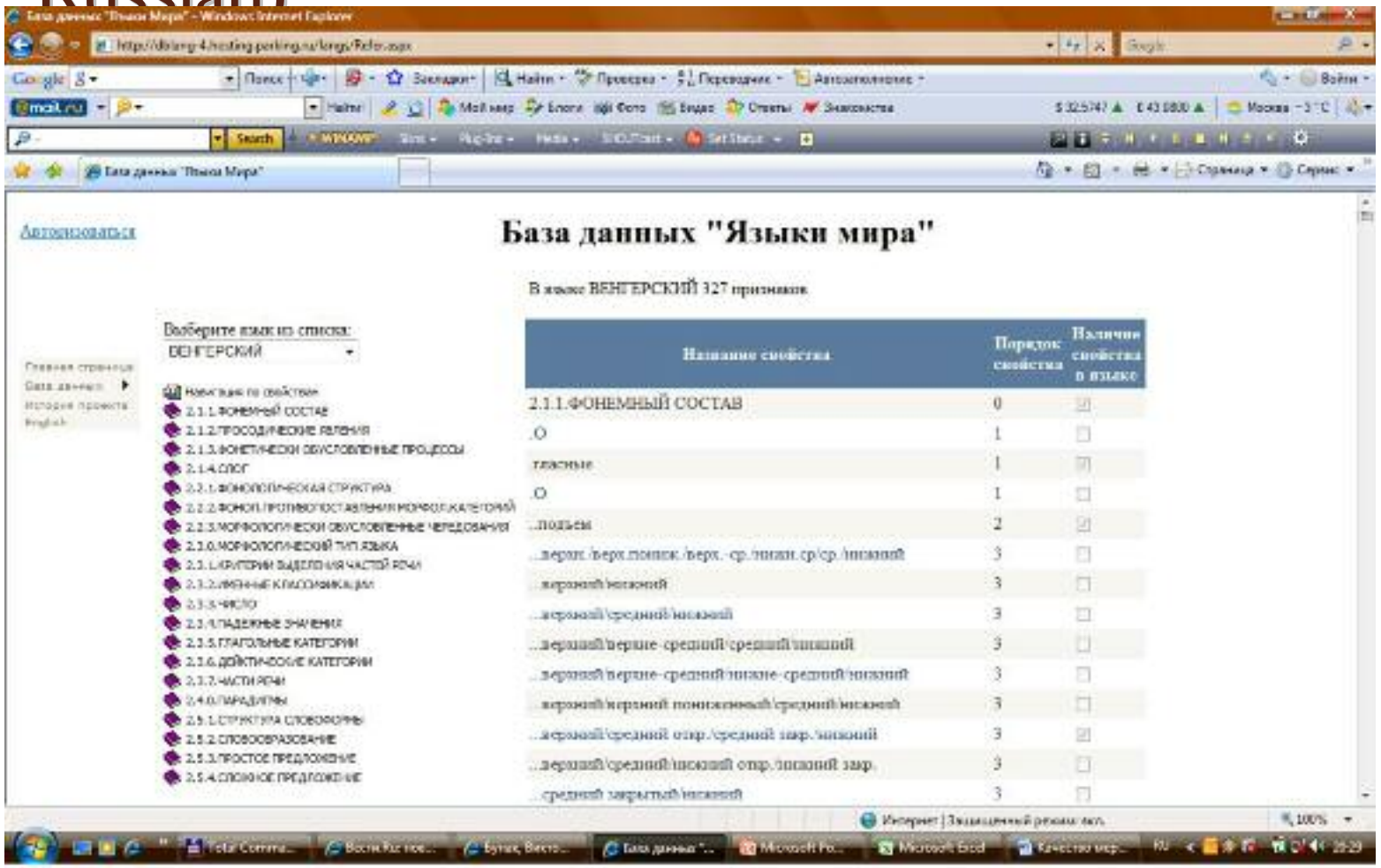
6. 1. Screenshots. Win Version (old variant)



6. 2. Screenshots. Win Version (new variant, developed by Oleg Belyaev)



6.3. Screenshots. Web Version is available on the site www,dblang.ru (while in Russian)



Also there is web-site (in English) devoted to quantitative researches on JM (www.dblang2008.narod.ru)

7. Introduction in the problem

- Similarity measure is a basis for phylogenetic calculations with the purpose of an establishment of genetic relationship between languages
- Recently (2005-2007) in works (Polyakov and Solovyev; Wichmann et al.) it has been established, that the measures constructed on typological data, reflect also genetic relationship,

BUT...

- = noise in WALS data (mainly because of absence of data) makes strong impact on results of calculations;
- = areal contacts in DB JM makes strong impact on results of calculations also.
- **Thus, in case of application of data from DB JM, the problem of a choice of a similarity measure as much as possible independent from areal contacts by the current moment is actual.**

8. Technique of an estimation of quality of a measure

Is based on the following aprioristic postulates:

- At first test set of languages is formed for which there are reliable expert data about genetic relationship.
- The technique and the formula of an estimation of the quality is offered for quantitative calculation of degree of approximation of the numerical result received by the program and an expert rating.
- In case of reception of reliable results on test set, the procedure of calculation of a measure of similarity can be transferred on the unstudied languages for check of hypotheses about their origin and genetic similarity.

9. The previous results

- The set of 48 languages (further «A.A. Kibrik's set») has been offered by group «World Languages» from Institute of Linguistics of RAS.
- The technique of estimations of quality of a similarity measure has been offered, based on ranging of languages concerning prototype language in each of eight families of the test set (Polyakov, Solovyev 2006).
- The formula of an estimation of quality of a similarity measure has been offered also.

10.1. A.A. Kibrik's set (48 languages)

N	Language		Family	Group
1	АБХАЗСКИЙ	Abkhaz	Northwest Caucasian	Northwest Caucasian
2	АГУЛЬСКИЙ	Aghul	Nakh-Daghestanian	Lezgic
3	АЗЕРБАЙДЖАНСКИЙ	Azerbaijani	Altaic	Turkic
4	АККАДСКИЙ	Akkadian	Afro-Asiatic	Semitic
5	АНГЛИЙСКИЙ	English	Indo-European	Germanic
6	АРМЯНСКИЙ	Armenian	Indo-European	Armenian
7	АССАМСКИЙ	Assamese	Indo-European	Indic
8	БАГВАЛИНСКИЙ	Bagvalal	Nakh-Daghestanian	Avar-Andic-Tsezic
9	БАШКИРСКИЙ	Bashkir	Altaic	Turkic
10	БЕЛОРУССКИЙ	Belarusan	Indo-European	Slavic
11	БЕНГАЛЬСКИЙ	Bengali	Indo-European	Indic
12	БИРМАНСКИЙ	Burmese	Sino-Tibetan	Burmese-Lolo
13	БОЛГАРСКИЙ	Bulgarian	Indo-European	Slavic
14	БУРУШАСКИ	Burushaski	Burushaski	Burushaski

10.1. A.A. Kibrik's set (48 languages)

15	БУРЯТСКИЙ	Buriat	Altaic	Mongolic
16	ВЕНГЕРСКИЙ	Hungarian	Uralic	Ugric
17	ВЕПССКИЙ	Veps	Uralic	Finnic
18	ГАЛИСИЙСКИЙ	Galician	Indo-European	Romance
19	ГРУЗИНСКИЙ	Georgian	Kartvelian	Kartvelian
20	ДАРИ	Dari	Indo-European	Iranian
21	ДАТСКИЙ	Danish	Indo-European	Germanic
22	ИСЛАНДСКИЙ	Icelandic	Indo-European	Germanic
23	ИСПАНСКИЙ	Spanish	Indo-European	Romance
24	ИТАЛЬЯНСКИЙ	Italian	Indo-European	Romance
25	ИТЕЛЬМЕНСКИЙ	Itelmen	Chukotko-Kamchatkan	Southern Chukotko-Kamchatkan
26	КАЛМЫЦКИЙ	Kalmyk_Oirat	Altaic	Mongolic
27	КОРЯКСКИЙ	Koryak	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan
28	ЛЕЗГИНСКИЙ	Lezgi	Nakh-Daghestanian	Lezgic

29	МАКЕДОНСКИЙ	Macedonian	Indo-European	Slavic
30	МОГОЛЬСКИЙ	Mogholi	Altaic	Mongolic
31	МОНГОРСКИЙ	Tu	Altaic	Mongolic
32	НЕМЕЦКИЙ	German	Indo-European	Germanic
33	НИВХСКИЙ	Gilyak	Nivkh	Nivkh
34	НОРВЕЖСКИЙ	Norwegian, Bokmål & Nynorsk	Indo-European	Germanic
35	ПЕРСИДСКИЙ	Western Farsi	Indo-European	Iranian
36	ПОЛЬСКИЙ	Polish	Indo-European	Slavic
37	ПОРТУГАЛЬСКИЙ	Portuguese	Indo-European	Romance
38	РУМЫНСКИЙ	Romanian	Indo-European	Romance
39	РУССКИЙ	Russian	Indo-European	Slavic
40	ТАДЖИКСКИЙ	Tajik	Indo-European	Iranian
41	ТАТАРСКИЙ	Tatar	Altaic	Turkic
42	ТУРЕЦКИЙ	Turkish	Altaic	Turkic
43	ТУРКМЕНСКИЙ	Turkmen	Altaic	Turkic
44	ФИНСКИЙ	Finnish	Uralic	Finnic
45	ХАНТЫЙСКИЙ	Khanty	Uralic	Ugric
46	ЧУКОТСКИЙ	Chukot	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan
47	ШУГНАНСКИЙ	Shughni	Indo-European	Iranian
48	ЭСТОНСКИЙ	Estonian	Uralic	Finnic

10.2. The formula of an estimation of quality of a similarity measure

	Group	Languages	Language-prototype	Ng	Ki
1	Uralic	Hungarian, Veps, Finnish, Khanty, Estonian	Finnish	5	K_1
2	Turkic	Azerbaijani, Bashkir, Tatar, Turkish, Turkmen	Turkish	5	
3	Mongolian	Buriat, Kalmyk_Oirat, Mogholi, Tu	Kalmyk_Oirat	4	K_3
4	Slavic	Belarusan, Bulgarian, Macedonian, Polish, Russian	Belarusan	5	K_4
5	Iranian	Dari, Western Farsi, Tajik, Shughni	Western Farsi	4	K_5
6	Germanian	English, Danish, Icelandic, German, (Norwegian, Bokmål & Nynorsk)	German	5	K_6
7	Romance	Galician, Spanish, Italian, Portuguese, Romanian	Spanish	5	K_7
8	Caucasian-1 (Nakh-Daghestanian)	Aghul, Bagvalal, Lezgi	Lezgi	3	K_8
9	Caucasian-2	Abkhaz, Georgian	-	-	-
10	Paleoasian	Burushaski, Itelmen, Koryak, Gilyak (Nivkh), Chukot	-	-	-
11	Others	Akkadian, Burmese, Armenian, Assamese, Bengali	-	-	-

All languages from A.A.Kibrik's set were divided on 11 groups according to genetic relationship

10.2. The formula of an estimation of quality of a similarity measure

- After calculation of a measure all languages are sorted eight times relatively to prototype languages in each group.
- Quality of measure K:

$K = (K_1+K_2+K_3+K_4+K_5+K_6+K_7+K_8)/8$

$Ki = Np/Ng$

Np – a number of related languages placed after prototype language.

Ng – a number of related languages in each group.

- Example

See tables with other measures at www.dblang2008.narod.ru

10.3. Results of calculations (Polyakov, Solovyev 2006)

- During DB testing great volume of works has been spent for choice the best variant of a similarity measure and to research of influence of different factors on quality of a measure. Among these factors there are types of features, their frequency, hierarchy in abstract structure, the contribution of various sections of the language description.
- Calculation of one variant of a measure on the set of 48 languages occupies about 20 minutes on the computer with processor Intel Pentium of 1,6 GHz. Calculation on one section of the language description lasts about 5 minutes. Full calculation on all data base (315 languages) is carried out over 10 hours.
- It has actually been established, that the best values of the measure quality reaches at simple additive sum of all conterminous features without restrictions on their frequency, hierarchy or an accessory to section of description (see table low). In this case on two groups (Ural, Turkic) it is reached full coincidence to traditional genetic representation and factor of quality K is equal 0,667. All other combinations of features yielded the worst result.
- The separate measure for each of sections of the description of language in DB is less than total measure under all model.

10.4. Results of calculations (Polyakov, Solovyev 2006)

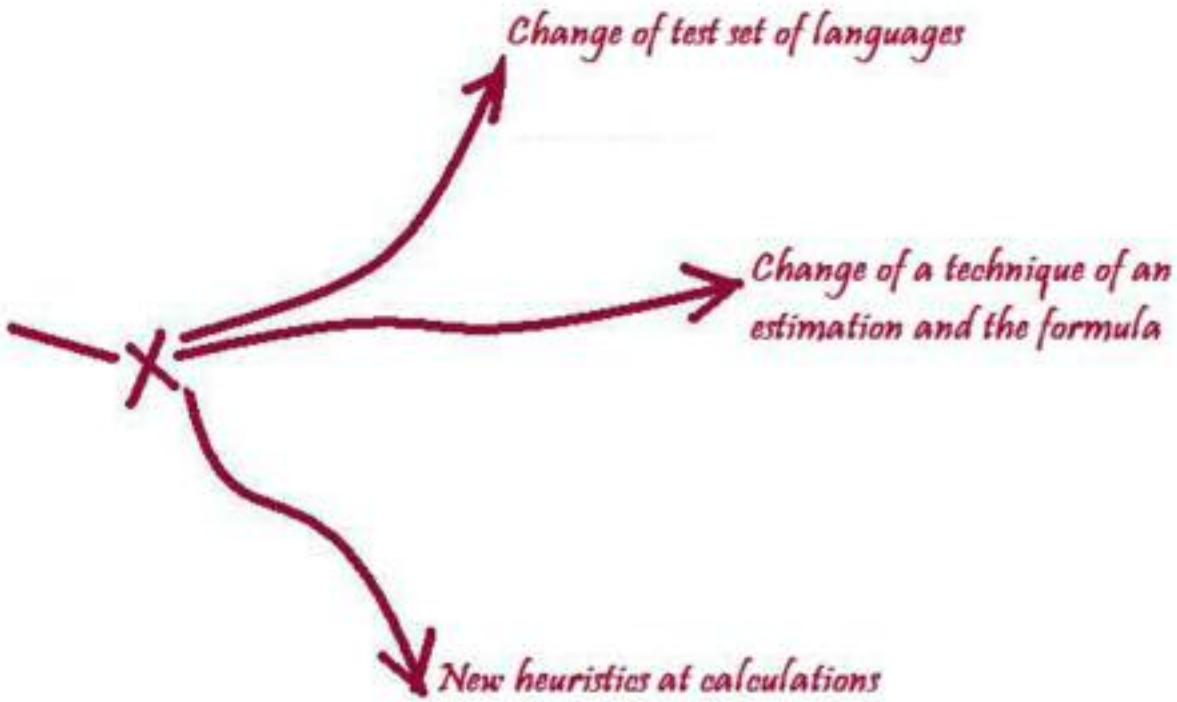
	All features	Only present in the languages	Only absent in the languages	Only classifying	Only factographic	Only present classifying	Only present factographic
K_1-Uralic (5 lang.)	1	0,5	0,75	0,5	1	0,25	0,5
K_2-Turkic (5 lang.)	1	0,75	0,5	0,75	0,75	0	0,75
K_3-Mongolian (4 lang.)	0,67	0,33	1	0,33	0,67	0,33	0,33
K_4-Slavic (5 lang.)	0,5	0,5	0	0,25	0,5	0,25	0,75
K_5-Iranian (4 lang.)	0,67	0,33	0	0,33	0,67	0	0,33
K_6-Germanian (5 lang.)	0,5	0,75	0,25	0,75	0,5	0,75	0,5
K_7-Romance (5 lang.)	0,5	0,5	0,5	0,5	0,5	0,25	0,75
K_8-Caucasian-1 (3 lang.)	0,5	0	0,5	0,5	0,5	0	0
K-Total	**0,67**	**0,46**	**0,44**	**0,49**	**0,64**	**0,23**	**0,49**

Section of essay	All features	Only present in the languages	Only classifying	Only factographic	Only present classifying	Only present factographic
2.1.1 Phonological structure	0,30	0,35	0,32	0,30	0,23	**0,36**
2.1.2 Prosody	0,29	**0,34**	0,24	0,26	0,19	0,33
2.1.3. Phonetics	0,09	0,10	**0,17**	0,06	0,06	0,14
2.1.4. The syllable	0,20	0,16	***0,50***	0,09	0,29	0,14
2.2.1. Phonotactics	0,13	**0,22**	0,13	0,16	0,19	0,16
2.2.2. Phonological opposition between morphological categories	0,16	**0,19**	**0,19**	0,13	0,13	0,16
2.2.3. Morphologically motivated alternations	**0,33**	0,17	0,18	0,17	0,09	0,10
2.3.0 Morphological type	0,34	**0,46**	0,33	0,26	0,21	0,23
2.3.1. Criteria for parts of speech assignment	0,07	0,07	0,07	0,07	0,07	0,07
2.3.2. Nouns	**0,23**	0,09	0,15	**0,23**	0,03	0,20
2.3.3. Number	0,16	0,13	**0,21**	0,20	0,17	0,20
2.3.4. Case	0,46	0,44	0,30	**0,53**	0,26	0,51
2.3.5. Verbal categories	**0,32**	0,29	0,20	0,29	0,14	0,20
2.3.6. Deictic categories	***0,47***	0,29	0,41	0,39	***0,36***	0,29
2.3.7. Parts of speech	0,44	**0,56**	0,20	***0,53***	0,07	***0,53***
2.4.0. Structure of morphological paradigms	**0,42**	0,40	0,30	0,32	0,22	0,27
2.5.1. Word structure	0,17	**0,30**	0,18	0,13	0,17	0,27
2.5.2. Word formation	0,18	**0,20**	0,17	0,18	0,10	0,17
2.5.3. The simple sentence	**0,41**	0,34	0,36	0,33	0,06	0,34
2.5.4. The complex sentence	0,10	0,17	0,18	0,10	0,09	**0,20**
Total sum in column:	**5,26**	**5,18**	**4,77**	**4,73**	**3,14**	**4,86**

11. Preliminary conclusions

- The measure reflects genetic similarity
- The contribution of structure of the description of language is insignificant
- The contribution of sections is rather essential

12. Directions of the further researches



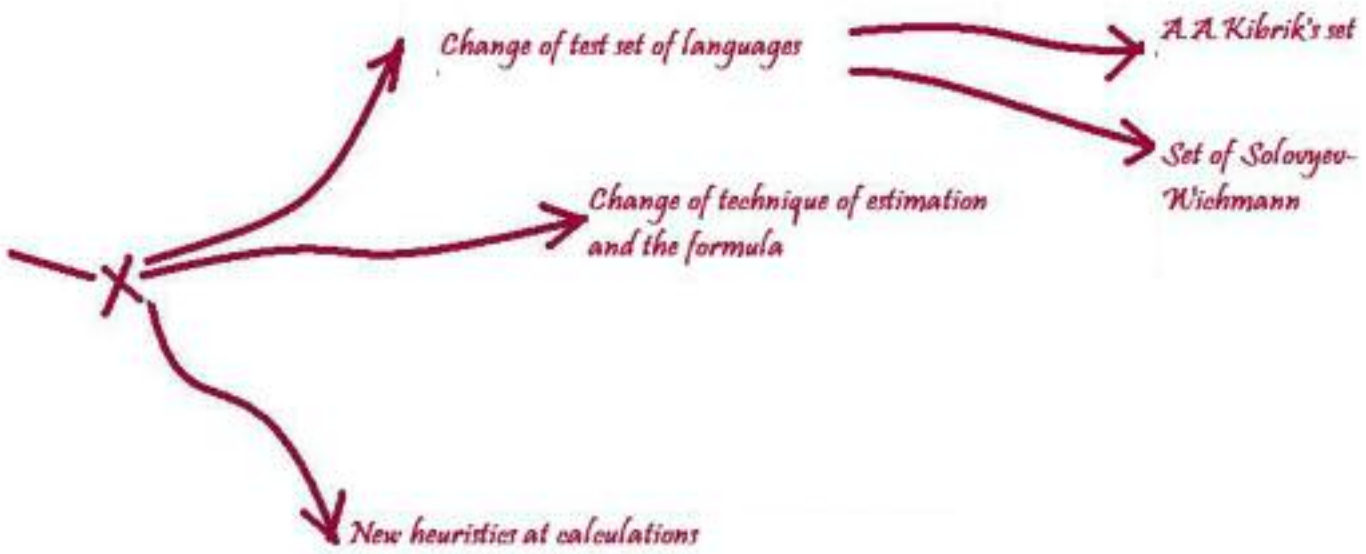
13. Aims of the investigation

- To choose new set of languages (comparable to content WALS, project ASJP and DB JM)
- To develop new, more thin technique of quality estimation
- **To find new heuristics, allowing to improve quality of a similarity measure. To establish a new benchmark in this field.**

14.1. The new set of languages comparable to content of WALS, project ASJP and DB JM

- The set is offered by Valery Solovyev and specified by Søren Wichmann in 2007
- The set includes the list from 39 (then reduced to 37) languages presented in WALS, JM and ASJP
- ***Thus there is a possibility not only to estimate quality of a similarity measure calculated on DB JM, but also to compare the genetic trees received from three linguistic sources.***
- ***Also there is a possibility of quantitative comparison of three projects on degree of coincidence of trees with the etalon.***

14.2. Alternatives on sets of languages



14.3. Set of Solovyev-Wichmann (39 languages)

	Language	Family	Genus
1	Modern Hebrew	Afro-Asiatic	Semitic
2	Chuvash	Altaic	Turkic
3	Yakut	Altaic	Turkic
4	Uzbek	Altaic	Turkic
5	Bashkir	Altaic	Turkic
6	Tatar	Altaic	Turkic
7	Azerbaijani	Altaic	Turkic
8	Kirghiz	Altaic	Turkic
9	Burushaski	Burushaski	Burushaski
10	Chukchi	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan
11	Itelmen	Chukotko-Kamchatkan	Southern Chukotko-Kamchatkan
12	Breton	Indo-European	Celtic
13	Dutch	Indo-European	Germanic
14	Swedish	Indo-European	Germanic
15	Icelandic	Indo-European	Germanic
16	Danish	Indo-European	Germanic
17	Bengali	Indo-European	Indic

	Language	Family	Genus
18	Persian	Indo-European	Iranian
19	French	Indo-European	Romance
20	Italian	Indo-European	Romance
21	Portugese	Indo-European	Romance
22	Catalan	Indo-European	Romance
23	Russian	Indo-European	Slavic
24	Polish	Indo-European	Slavic
25	Bulgarian	Indo-European	Slavic
26	Czech	Indo-European	Slavic
27	Ukrainian	Indo-European	Slavic
28	Georgian	Kartvelian	Kartvelian
29	Lezgian	Nakh-Daghestanian	Lezgic
30	Chechen	Nakh-Daghestanian	Nakh
31	Abkhaz	Northwest Caucasian	Northwest Caucasian
32	Kabardian	Northwest Caucasian	Northwest Caucasian
33	Finnish	Uralic	Finnic
34	KomiZyrian	Uralic	Finnic
35	Nenets	Uralic	Samoyedic
36	Selkup	Uralic	Samoyedic
37	Hungarian	Uralic	Ugric
38	Khanty=Yakut	Uralic	Ugric
39	Ket	Yeniseian	Yeniseian

14.4. Set of Solovyev-Wichmann (39 languages)

- Examples of trees, built on different data.
- Tree from JM data
- Tree from WALS data
- Tree from ASJP data

ASJP tree is the most reliable in its quality to describe genealogic relationship. JM tree is placed at the second place and WALS tree is at the third place.

15.1. New more thin techniques of an estimation of quality of similarity measures

- After calculation of a measure all languages are sorted 39 times relatively to each languages.
- Quality of measure K:

$K = \sum(K_i)/39,\ i = i\ldots39$

$Ki = Np/Ng,\ i = i\ldots39$

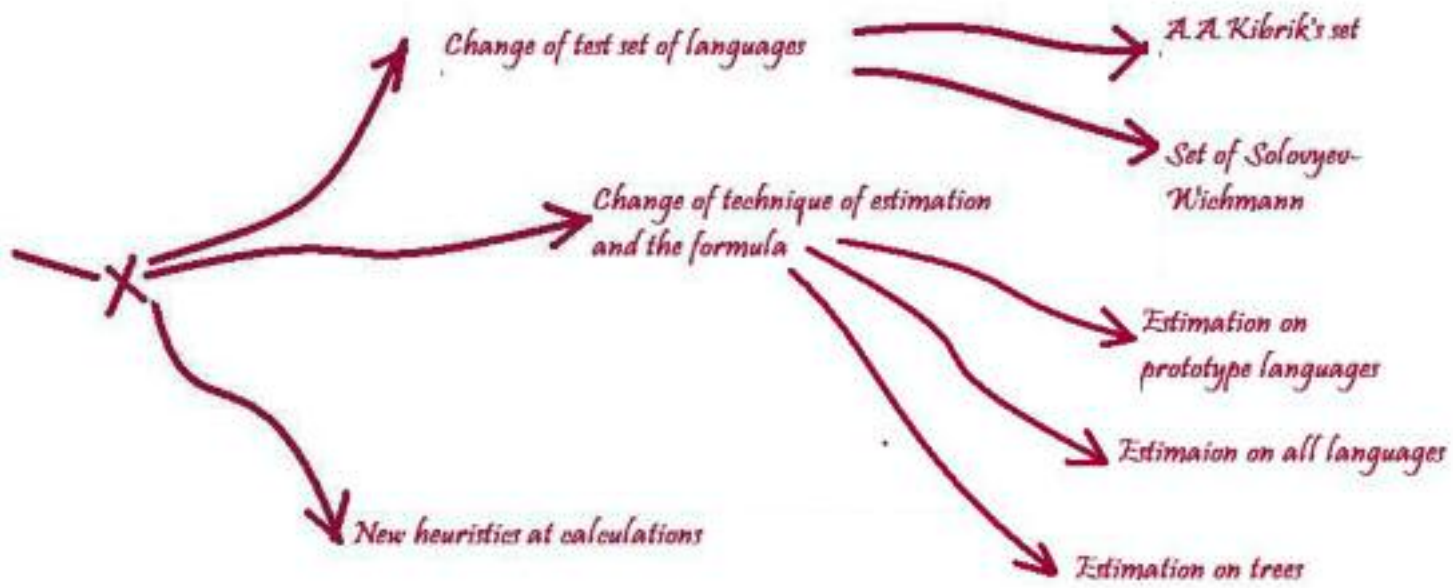
Np – a number of related languages placed after each language.

Ng – a number of related languages in each group.

15.2. New more thin techniques of an estimation of quality of similarity measures

Also different techniques exist that allow to compare trees immediately. In this case a quality measure is calculated as editorial distance (for ex. Robinson and Foulds topological distance) ***but in this case reference tree is needed.***

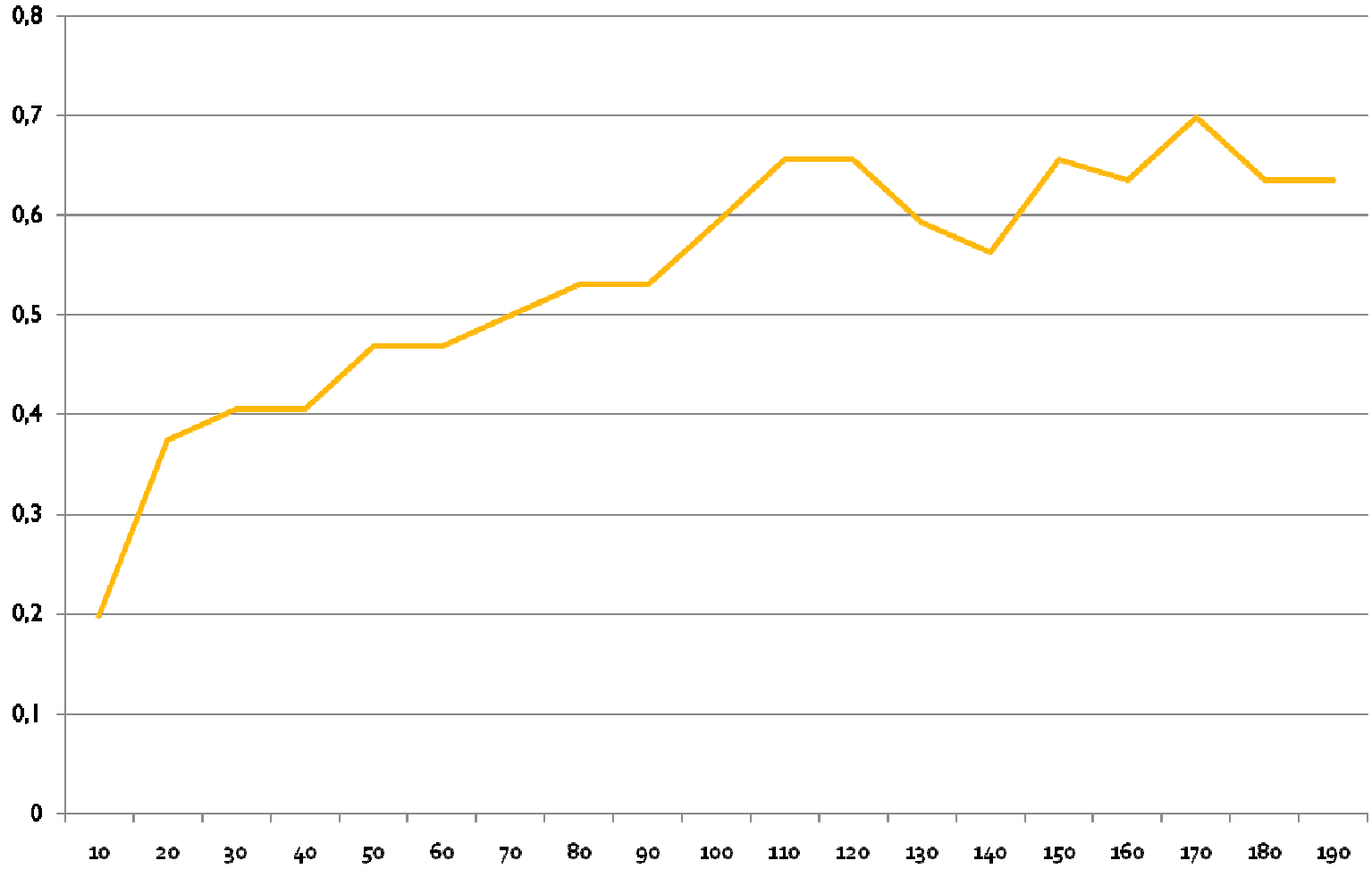
15.3. Alternatives on techniques of estimation of measure quality



16.1. New heuristics, allowing to improve quality of similarity measure (on A.A.Kibrik's set)

- Restriction on frequency of features (T N=170 lang.) gives increase in a measure to 0,697
- Restriction on description sections gives increase in a measure to 0,760
- Restriction by filter of genealogic markers (K = 2) gives a measure = 0,531

16.2 Dependency of the quality of measure from the frequency restriction (N, lang)



16.3. New heuristics, allowing to improve quality of similarity measure

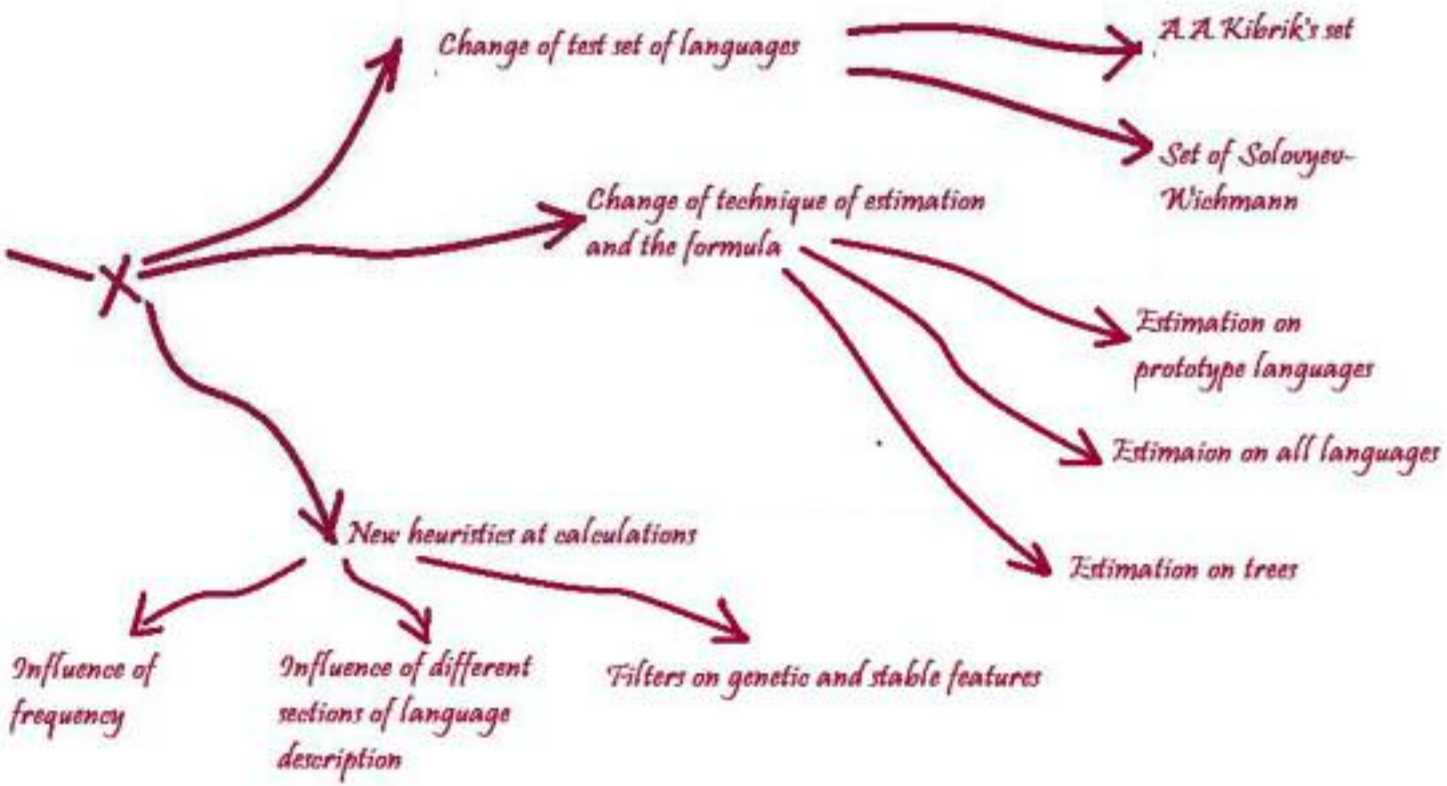
- **The sections of essay were chosen that has a quality value more than 0,25. The list of these sections includes numbers {1,2,7,8,12,13,14,15,16,19}.**

- **See table at slide 13.**

16.4 Sections of essay used to improve result of calculation

- **2.1.1 Phonological structure**
- **2.1.2 Prosody**
- 2.1.3. Phonetics
- 2.1.4. The syllable
- 2.2.1. Phonotactics
- 2.2.2. Phonological opposition between morphological categories
- **2.2.3. Morphologically motivated alternations**
- **2.3.0 Morphological type**
- 2.3.1. Criteria for parts of speech assignment
- 2.3.2. Nouns
- 2.3.3. Number
- **2.3.4. Case**
- **2.3.5. Verbal categories**
- **2.3.6. Deictic categories**
- **2.3.7. Parts of speech**
- **2.4.0. Structure of morphological paradigms**
- 2.5.1. Word structure
- 2.5.2. Word formation
- **2.5.3. The simple sentence**
- 2.5.4. The complex sentence

16.5. Alternatives on heuristics



17.1 Table of classification of features and genealogical markers

Classification of features

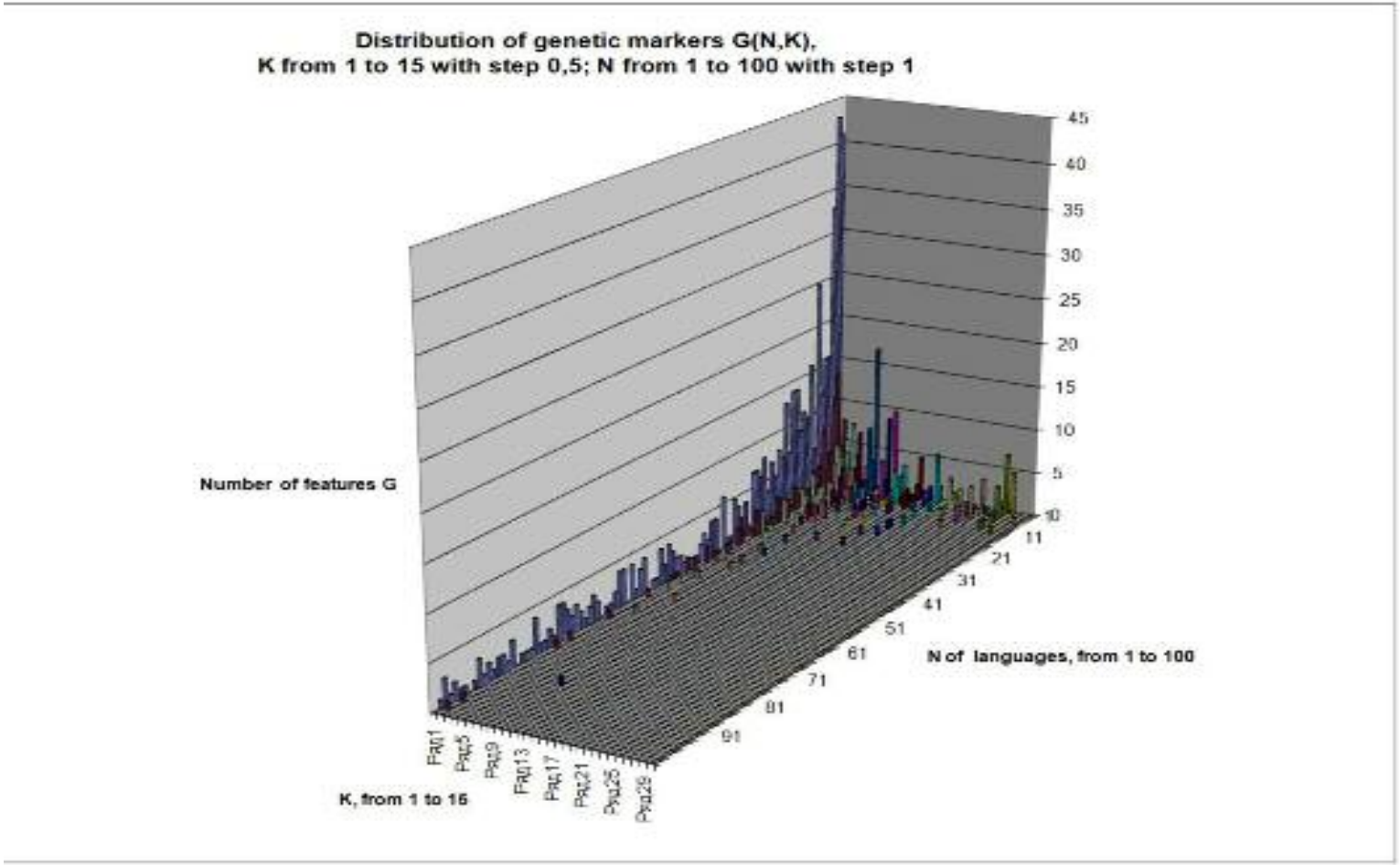
Group of feature		General number of language having a feature in JM (DB)	Frequency of a feature in a family (Fam)	Frequency of a feature in a genera / group / subgroup (Gen)	Frequency of a feature in other families genera / groups / subgroups (Oth)
The I-st group	Absolute universals	F_{DB} = 100 % from all number of languages	-	-	-
The II-nd group[1]	Statistic universals	-	$F_{Fam} >= 0,5$	-	$F_{Oth} >= 0,5$
The III-rd group	Genealogically stable features for family	$N_{DB} <= 2*N_{Fam}$	$F_{Fam} >= 0,5$	-	$F_{Oth} << 0,5$
The IV-th group	Genealogically stable features for genera / group / subgroup	$N_{DB} <= 2*N_{Gen}$	$F_{Fam} <= 0,5$	$F_{Gen} >= 0,5$	$F_{Oth} << 0,5$
The V-th group	Genealogically stable features for part of genera / group / subgroup	$N_{DB} <= 2*N_{Gen}$	$F_{Fam} << 0,5$	$F_{Gen} << 0,5$	$F_{Oth} << F_{Gen}$
The VI-th group	Genealogically unstable features	-	$F_{Fam} << 0,5$	$F_{Gen} << 0,5$	$F_{Oth} \sim F_{Gen}$
The-VII-th group	Unique features	N_{DB} = 1 (in absolute values)	-	-	-

F – frequency of feature
N – number of language

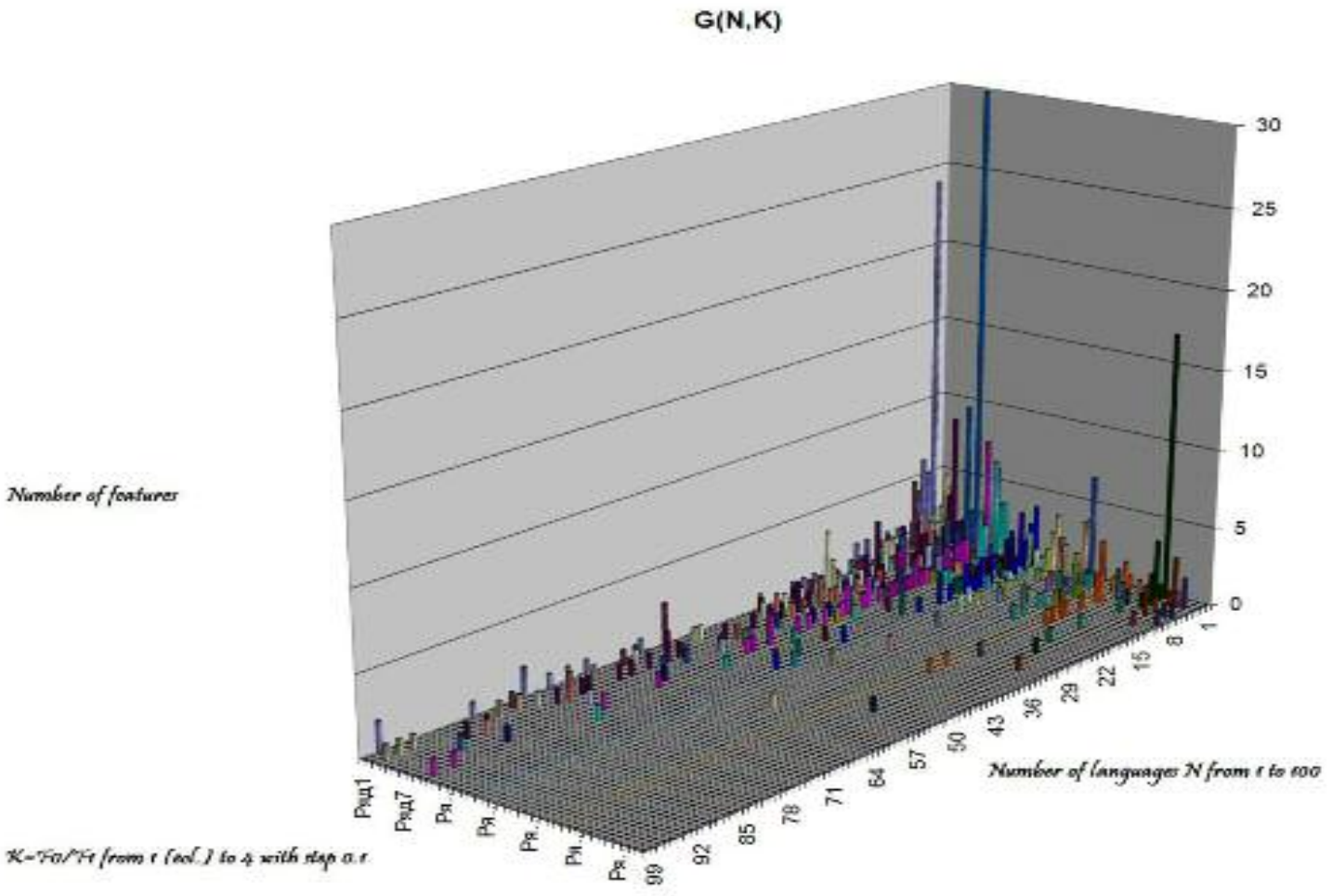
17.2 Extended list of genealogical markers includes:

- Positive markers that are dominant only in one family / genera / group / subgroup
- Negative markers that are absent (or most absent) only in one family / genera / group / subgroup
- Double positive markers that are dominant only in two family / genera / group / subgroup
- Double negative markers that are absent (or most absent) only in two family / genera / group / subgroup (very rare cases)

17.3. Distribution of genealogical markers in JM-1



17.4. Distribution of genealogical markers in JM-2



18. Parts of data used in different heuristics

Heuristics	Quality of measure	Part of data used
No restrictions	0,667	100 %
Restriction in frequency (N <= 170	0,698	52,1 %
Restriction in parts of model (ten the best parts used)	0,760	47,7 %
Using of positive and negative genealogical markers (K=2)	0,531	38,8 %

19. Conclusions and the future researches

- New heuristics (frequency and the filter on sections) allow to improve quality of a measure
- In the future:

  - It is planned to use such factors as stability (Wichmann and Holman, 2008; Belyaev, 2008), full list of genealogical markers, weights from linear regression decision;

(***It is necessary to notice, that use of similar techniques moves the problem from the area of clusterization in the classification area.***)

- It is supposed to apply more thin measures of an estimation of quality;

- It is more preferable to use the set comparable to other linguistic resources (WALS, ASJP, etc.)

NEW EVENT INFORMATION

TUTUTORIAL IN COMPUTATIONAL LANGUAGE TYPOLOGY AND QUANTITATIVE COMPARATIVISTICS

Joined with CML Conferences

Took places in Sofia (Bulgaria, 2007) and Bechichi (Montenegro, 2008)

The next tutorial is planned in Constantsa

(Romania, in September 2009)

YOU ARE WELCOMED!

Additional information will be soon at

cml.msisa.ru

Contacts:

Vladimir Polyakov

Institute of Linguistics of RAS

pvn-65@mail.ru

www.dblang.ru

www.dblang2008.narod.ru

www.cml.msisa.ru

The research is supported by RFBR grant (www.rfbr.ru), № 07-06-00229a

Thanks!