

**INTERNATIONAL CONFERENCE
“COGNITIVE MODELING IN LINGUISTICS”
CML-2011, Corfu (Greece)**

EXTENDED ABSTRACT SUBMISSION FORM

1. The extended abstract should be sent electronically both to cml2011@mail.ru and cml2011@list.ru.

Name of the file:

Please, give a name to the file which is similar to the title of the paper. Files in doc (MS Word), rtf or pdf (Adobe Acrobat) format are strongly recommended.

2. The title of the report in English is:

The title should not exceed 14 words.

3. Author details:

	Title (Mr/Ms, Dr, Prof)	First Name	Family Name	e-mail	Institution (Name and Postal Address) ¹	Department (Name and Position)
1.						
2.						
3.						
4.						
5.						
6.						
7.						

4. Keywords:

Please, list less than 7 keywords, separated with comas (in English).

5. Contribution addresses the following two topics:

First Topic	Second Topic

Please, indicate two topics from the topics given below:

1. General Topics

Cognitive Models of Language Phenomena

Formal Models in Language and Cognition

Cognitively-Oriented Computer Applications and Linguistic Resources

General Problems of Cognitive Science

2. Models and Studies in:

Speech Perception and Production

Psycholinguistics and Psychosemantics

Semiotics, Semantics and Pragmatics

Language Processing, Memory and Thought

Language and Emotions

Child Speech and Language Acquisition

Language Typology

Translation and Cognition

Speech Impairments

Language Disorders

Cognitive Aspects of Theology

Cognitive Aspects of Information Technologies Development and Using

Cognitive Mechanisms for Decision Making

3. Cognitive Linguistics

Metaphor Theory

Mental Lexicon and Lexicon Ontology

Naive World Image and Verbal Form

Knowledge Conceptualisation and Verbalisation

Cognitive Mechanisms for Text Processing

Cognitive Models in Language Learning

Species-Specificity of Human Language

Thinking through Language Processing

Cognitive Slavistics

The rules for extended abstract starts on the second page:

¹ This address will be used to send CML Proceedings by surface post.

THE TITLE OF THE REPORT, TIMES NEW ROMAN, 12, BOLD, CENTERED^{2 3}

One blank line.

List of authors, Times New Roman, 12, bold, centered⁴

One blank line.

KEYWORDS

The list of keywords - Times New Roman, 10, italic.

One blank line.

The text of extended abstract – Times New Roman, 10, justified, first line – 1 cm, single-spacing, as all other sections in the paper should be. The abstract is required in English. The abstract should not exceed 1000 words (~ 2 page excluding references). Minimal size of the abstract is 500 words.

One blank line.

REFERENCES

The references are in alphabetic order, Times New Roman, 10, justified. There is an example below:

1. Barsalou, L.W., Solomon, K.O., & Wu, L.L. (1999). Perceptual simulation in conceptual tasks. In M.K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, typological, and psychological perspectives in cognitive linguistics: The proceedings of the 4th conference of the International Cognitive Linguistics Association, Vol. 3*. Amsterdam: John Benjamins.
2. Deubel, P. (2003). An Investigation of Behaviorist and Cognitive Approaches to Instructional Multimedia Design. *Journal of Educational Multimedia and Hypermedia* 12(1), 63-90. [Online]. Available: <http://dl.aace.org/12691>
3. Hassett, J., Ingram, A., Hassett, M., & Marino, E. (2003). What Do Learners Like? Ratings of Off-the-Shelf Web-Based Training Courses. *International Journal on E-Learning* 2(1), 50-60. [Online]. Available: <http://dl.aace.org/11553>
4. Levert, C., & Pierre, S. (2003). Designing Distributed Virtual Laboratories: Methodological and Telecommunications Aspects. *International Journal on E-Learning* 2(3), 18-28. [Online]. Available: <http://dl.aace.org/13504>
5. www.uvpl.com. Title of the web-page. Date.

Everything mentioned in the text sources must be given with their complete reference. It is recommended to cite the page numbers. If the source can be found on www, use: “[Online] Available:<http://xxxxxxx>”

Please, mention only First Name and Family Name of the author.

The example of abstract starts on the third page:

² © Authors' Names, 2011, Times New Roman, 10

³ References to grants support, Times New Roman, 10

⁴ Author's affiliation and contacts (University Title, Country, e-mail), Times New Roman, 10

COMPARATIVE ANALYSIS OF PHYLOGENIC ALGORITHMS ^{5 6}

Venera Bayrasheva ⁷

Renat Faskhutdinov ⁸

Valery Solovyev ⁹

KEYWORDS

Phylogeny, Phylogenic Algorithms

The phylogenetics suggests different algorithms for constructing evolutionary trees. Meanwhile the question of better algorithm is still open. The NJ algorithm has been recently used in linguistic researches. The belief in advantages of NJ algorithm is based on paper [1]. However, in [2] it was proved on the material of Sumba languages that UPGMA has better results. We compare these two algorithms as the most popular ones.

Careful analysis of the argumentation given in paper [1] shows that NJ provides better results on the trees of a certain topology (or structure). As a matter of fact the authors of the paper tested only two very specific topologies of trees. Besides, the research in [1] was initially oriented to the studies of biological evolution, but not a language one. That is why the task of systematic comparison of the UPGMA and NJ algorithms on the trees of different configuration is of vital importance as well as the constructing the realistic model of language evolution trees.

We analyze certain cases of using the algorithms mentioned (for Caucasian, Indian and other groups of languages and different types of input data), showing that UPGMA often gives better results than NJ.

The influence of the tree topology on the result is being studied. We introduce a numerical characteristic of the tree branching degree, i.e. the sum of levels of inside peaks. In this case the level of root is considered to be equal 0 and the level of an ancestor is 1 greater than the level of a direct descendant.

We made an experiment with generation of random binary trees of an arbitrary topology. The trees are generated with a given number of leaves and the length of every rib was determined as a random number on a given segment. Then a matrix of distances between leaves was constructed for every generated T tree by the lengths of its ribs. After that T-UPGMA and T-NJ trees are determined in accordance with these matrixes by UPGMA and NJ methods. The differences of the initial and constructed trees are estimated by the Robinson-Foulds method, (i.e. the number of elementary transformations, which is necessary for transforming one tree into another).

We studied the question of rib length variations in the real trees of language evolution. One of the most completely described trees is the evolution tree of the Turkic family, given in paper [3]. The lengths of all ribs in the tree (there are 77 of them) have been calculated and located in the order of increasing.

It turned out that there are several super long ribs. The longest, which is of 2130 years, corresponds to the initial separation of the Chuvash language from proto-Turkic language. The next longest ribs (1330 and 1270 years) demonstrate separating the Yakut language from the Siberian branch and the Salar language from the Oguz branch. There is one abnormally short rib of 30 years that is the rib in evolution tree of Kypchat languages. The lengths of the majority of ribs excluding the shortest and the ten longest ribs can be strictly put on the direct line. The fact that lengths of the majority of ribs except some of them can be put on the direct line means that the rib lengths can be considered as a random value with an even distribution on a segment.

The lengths vary from 90 to 650 years. Thus, the average meaning of a rib length is 370 years. The declination is ± 280 years that equals 75% average length. Similar results are obtained for other language families. This data is a basis for the algorithms of random tree generation.

We made calculations for two cases, when number of leaves is equal to 15 and 50. 1000 random trees have been generated and the results have been averaged. The branching measure for the trees with 15 leaves is from 31 to 105 and for the trees with generated random sample it was from 33 to 58. It is convenient to divide all the trees by the measure of their branching into several groups in order to analyze the data obtained. We chose four groups approximately equal by the number of trees with the following values of measure: 31-36, 37-40, 41-45, 46-105. For each group we calculated the averaged Robinson-Foulds distances, given in the Table.

Table. Averaged distances, r = 15 leaves.

⁵ © Venera Bayrasheva, Renat Faskhutdinov, Valery Solovyev, 2010

⁶ The research was supported by Russian Foundation of Basic Research (grant № 10-06-00087-a.)

⁷ Kazan State University, Russia.

⁸ Kazan State University, Russia.

⁹ Kazan State University, Russia, maki.solovyev@mail.ru

Measure of branching	UPGMA	NJ
31 - 36	4,31	5,04
37 - 40	6,41	5,72
41 - 45	8,11	6,42
46 - 105	9,04	7,43

It is clear that the efficiency of the algorithms depends on the topology of trees. For trees with a small measure of branching, which are close to a complete one, better results are provided by UPGMA algorithm. The similar result is obtained for $r = 50$.

Thus, it has been proved that NJ algorithm is not undoubtedly the best one. Both real examples and modeling by generation method of random trees shows that UPGMA is preferable in a number of cases.

The authors extend their gratitude to S. Wichmann for sensible advice. The work has been implemented with the support of RFBR.

REFERENCES

1. Saitou N., Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* V.4, N4, 1987. pp 406-425.
2. Donwey S., Halmark B., Cox M., Norquest P., Lansing S. Computational Feature-Sensitive Reconstruction of Language Relationships: Developing the ALINE Distance for Comparative Historical Linguistic Reconstruction. *Journal of Quantitative Linguistics*. V.15, N4, 2008, pp. 340-369.
3. *Sravnitel'no-istoricheskaja grammatika tjurkskih jazykov*. Red. E.R.Tenischev. Moscow: Nauka. 2002. (In Russian)