ОБРАБОТКА ТЕКСТА И КОГНИТИВНЫЕ ТЕХНОЛОГИИ

№ 19

TEXT PROCESSING AND COGNITIVE TECHNOLOGIES

XII Международная конференция

«Когнитивное моделирование в лингвистике»

Труды

*

The XII-th International Conference

«Cognitive Modeling in Linguistics»

Proceedings

September 7 — 14

2010

Croatia

Dubrovnik

УДК 004.9:81'322
ББК 81.1
О23

Ответственные научные редакторы:
докт. физ.-мат. наук, проф. *В.Д. Соловьёв;*
канд. технич. наук, доц. *В.Н. Поляков*

О23 Обработка текста и когнитивные технологии: сборник статей. Когнитивное моделирование в лингвистике: [Текст] : Труды XII Междунар. науч. конф. (Дубровник, 7–14 сентября 2010 г.) / отв. науч. ред. В.Д. Соловьёв, В.Н. Поляков. – Казань : Казан. ун-т, 2010. – Вып. 19. – 404 с.

Данный том содержит труды XII Международной конференции "Когнитивное моделирование в лингвистике" (Дубровник, 2010), посвященной перспективным направлениям в когнитивистике. Целью конференции является объединение усилий специалистов по когнитивной психологии, когнитивной инженерии, обучающим технологиям, нейронным сетям, когнитивной лингвистике, инженерии знаний, семиотике, психолингвистикé и другим направлениям, связанным с исследованиями когнитивных процессов обработки языковой информации. Адрес сайта конференции – cml.msisa.ru.

УДК 004.9:81'322
ББК 81.1

Text Processing and Cognitive Technologies. Collection of Papers. Issue 19. Cognitive Modeling in Lingvistics: Proceeding of the XII Internatonal Conference / Ed. by V. Solovyev, V. Polyakov. Kazan: KSU, 2010, 404 p.

The present volume has been comprised by the papers which represent proceedings of the XII-th International Conference "Cognitive Modeling in Linguistics" (Dubrovnik, 2010) was devoted to several perspective branches of Cognitive Science. The objective of the conference integration of efforts of specialists in cognitive psychology, cognitive engineering, learning technologies, neural networks, cognitive linguistics, knowledge engineering, semiotics, psycholinguistics etc. related to investigation of cognitive processing of verbal information. Out conference site is cml.msisa.ru is devoted to the conference.

THE PHYLOGENY OF MALAGASY DIALECTS[145]

Maurizio Serva[146]

Philippe Blanchard[147]

Eric W. Holman[148]

Filippo Petroni[149]

Dima Volchenkov[150]

Søren Wichmann[151]

KEYWORDS

Phylogeny, Malagasy Dialects

All dialects spoken in Madagascar belong to the Indonesian subgroup of the Austronesian family and they are sufficiently homogeneous that nearby populations can easily reciprocally understand but not faraway ones.

The Indonesian language most closely related to Malagasy is probably Maanyan of South-east Kalimantan with a 45 percent of shared basic vocabulary, but close languages can be found in Sulawesi, Malaysia and Philippine Islands. For this reason, the history of Madagascar peopling and settlement is subject to dispute and alternative interpretations among scholars. It seems that Indonesian sailors reached Madagascar by a maritime trek at a time which is disputed and can be between one and two thousand years ago,

---

146 Dipartimento di Matematica, Università dell'Aquila, Italy

147 Universität Bielefeld, Fakultät für Physik, Germany

148 Psychology Department, University of California, Los Angeles, USA

149 Facoltà di Economia, Università la Sapienza, Roma, Italy.

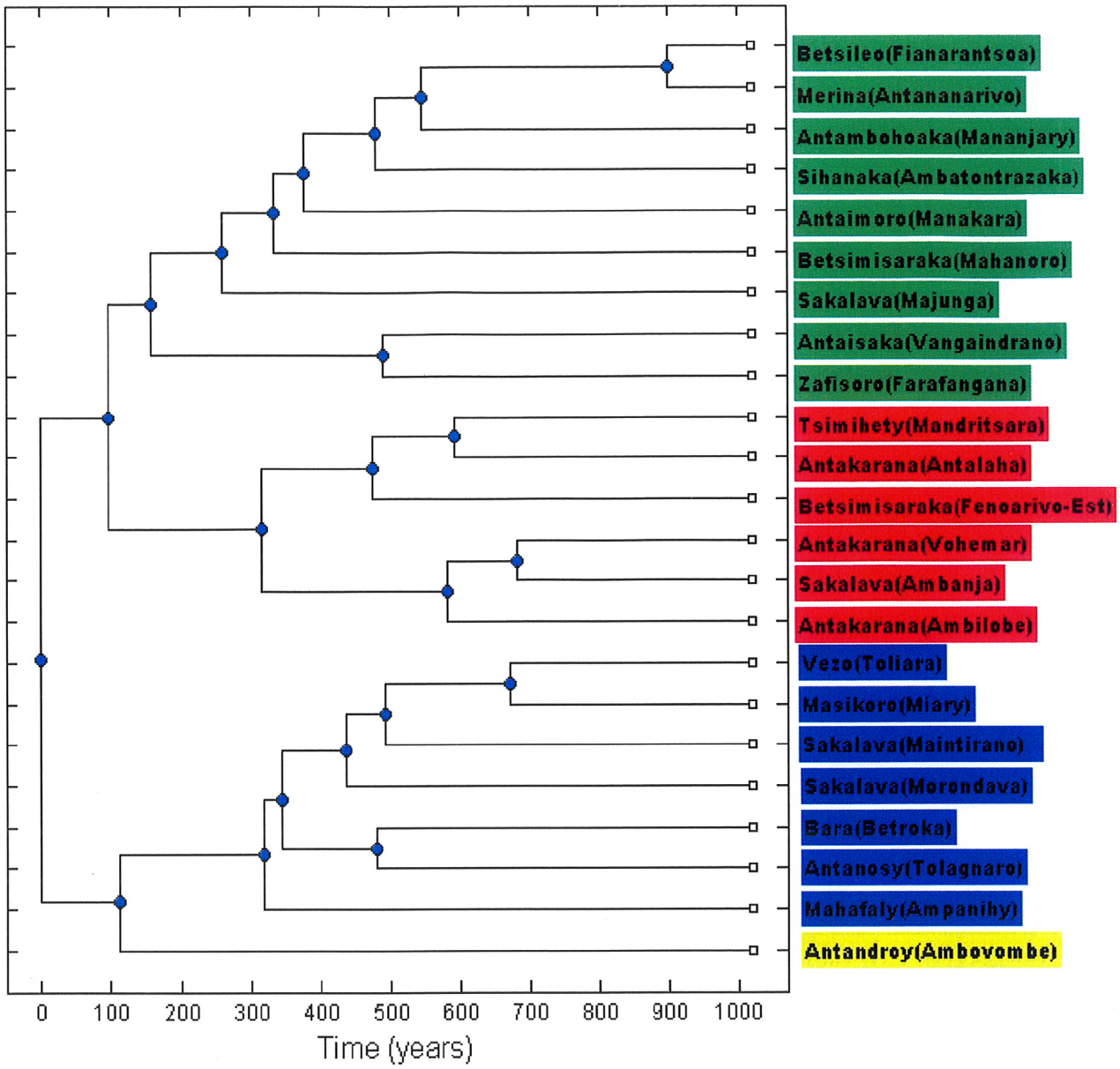
150 Center of Excellence Cognitive Interaction Technology, Bielefeld, Germany

151 Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

furthermore it is not clear if there were multiple settlements during centuries or a single one. More mystery is added by the fact that the Maanyan, which seems to be the closest language to Malagasy, is spoken by a population which lives along the rivers of Kalimantan and which does not possess the necessary skill for long maritime navigation. A possible explanation is that they arrived as slaves of Malay sailors and, in this case, the dialects should show both a Malay and a Maanyan contribution.

In our research we would like to try to answer to these problems by a glottochronological analysis based on the automated method recently proposed by the authors (Serva and Petroni 2008, Holman *et al.* 2008, Petroni and Serva 2008, Bakker *et al.* 2009).

The data were collected by one of us (M.S.) at the beginning of 2010 with the invaluable help of Joselinà Soafara Néré and consist in a vocabulary of 200 words for 23 dialects covering all areas of the Island, probably the largest collection of comparative Swadesh lists for what concerns Madagascar.



The automated method uses Levenshtein normalized distance to compute distance between words with same meaning and it averages on all the words contained in the list to obtain distance between pairs of dialects or languages. Then, these lexical distances are transformed, by a simple rule, in separation times (genealogical distances).

We are able to compare pairs of Malagasy dialects or a Malagasy dialect with an external language (like Maanyan). We obtain results concerning the internal relationships among the 23 Malagasy dialects considered, which can give both information about dates and modalities of the settlement, as well we obtain results concerning external relationships, which can give information concerning the Indonesian origin of Malagasy ancestors.

We only show in this abstract the tree that we obtain for the 23 Malagasy dialects. It can be seen that they are divided into two main groups: south-west and east-center-north. This differs from previous results (Vérin *et al.* 1969) which found a main division north/others. Our results are confirmed by a network analysis based on a method that we recently proposed (Blanchard *et al.* in press).

REFERENCES

M. Serva and F. Petroni, *Indo-European languages tree by Levenshtein distance.* EuroPhysics Letters 81, (2008), 68005.

E. W. Holman, S. Wichmann, C. H. Brown, V. Velupillai, A. Müller and D. Bakker, *Explorations in automated lexicostatistics.* Folia Linguistica 42.2, (2008), 331-354.

F. Petroni and M. Serva, *Languages distance and tree reconstruction.* Journal of Statistical Mechanics: theory and experiment, (2008), P08012.

D. Bakker, C. H. Brown, P. Brown, D. Egorov, A. Grant, E. W. Holman, R. Mailhammer, A. Muller, V. Velupillai and S. Wichmann, *Adding typology to lexicostatistics: a combined approach to language classification,* Linguistic Typology 13, (2009), 167-179.

P. Vérin, C. P. Kottak and P. Gorlin, *The glottochronology of Malagasy speech communities,* Oceanic Linguistics 8, (1969), 26-83

Ph. Blanchard, F. Petroni, M. Serva and D. Volchenkov. *Geometric Representations of Language Taxonomies.* Computer Speech and Language, (in press).