PROCEEDINGS
OF THE INTERNATIONAL CONFERENCE
"TURKIC LANGUAGES PROCESSING"

TurkLang-2015

September 17–19, 2015, Kazan, Tatarstan, Russia

Kazan
2015

UDC 002.001.4
BBC 32.973.8122
P 93

Tatarstan Academy of Sciences
Research Institute of Applied Semiotics
L.N. Gumilyov Eurasian National University
Ministry of Education and Science of the Republic of Kazakhstan
Scientific Research Institute of Artificial Intelligence
Russian Foundation of Basic Research
Russian Association of Artificial Intelligence
Kazan Federal University
Institute of Philology and Intercultural Communication
Institute of Computational Mathematics and InformationTechnologies
The Higher Institute for Information Technology and Information Systems
"Selet" Tatarstan Republic Youth Social Fund

The publication was supported by RFBR, research project No. 15-46-07007.

Printed by decision of the Editorial Board of the Tatarstan Academy of Sciences

P 93 **Proceedings of the International Conference "Turkic Languages Processing: TurkLang-2015".** – Kazan: Academy of Sciences of the Republic of Tatarstan Press, 2015. – 488 c.
ISBN 978-5-9690-0262-3

These proceedings include papers presented at the International Conference on Turkic languages processing "Turklang-2015" (Kazan, Tatarstan, Russia, 17–19 September 2015). The Conference is focused on the relevant problems of computational linguistics in Turkic languages. The participants discussed issues related to the development of formal linguistic models, corpora projects, machine translation tasks, applied systems and technologies of computer and cognitive linguistics. These proceedings were designed for researchers, teachers and students specializing in the field of computer and cognitive linguistics and its applications.

UDC 002.001.4
BBC 32.973.8122

ISBN 978-5-9690-0262-3

© Academy of Sciences
of the Republic of Tatarstan Press, 2015

СОДЕРЖАНИЕ

SECTION 1. **MACHINE TRANSLATION TECHNOLOGIES**

STUDY OF THE PROBLEM OF CREATING STRUCTURAL TRANSFER RULES AND LEXICAL SELECTION FOR THE KAZAKH-RUSSIAN MACHINE TRANSLATION SYSTEM ON APERTIUM PLATFORM *Abduali Balzhan, Akhmadieva Zhadyra, Zholdybekova Saule, Tukeyev Ualsher, Rakhimova Diana* 5

CHOOSING THE MODEL FOR SOLVING THE PROBLEM OF LEXICAL SELECTION FOR ENGLISH-KAZAKH LANGUAGE PAIR IN THE FREE/OPEN-SOURCE PLATFORM APERTIUM *Dina Amirova* 10

THE ONTOLOGICAL MODEL OF NOUN FOR KAZAKH-TURKISH MACHINE TRANSLATION SYSTEM *Lena Zhetkenbay, Altynbek Sharipbay, Gulmira Bekmanova, Unzila Kamanur* 15

THE ALGORITHM OF MACHINE TRANSLATION FROM UZBEK TO KARAKALPAK *Azizbek Kadirov* 24

LEXICAL SELECTION RULES FOR KAZAKH-TO-ENGLISH MACHINE TRANSLATION IN THE FREE/OPEN-SOURCE PLATFORM APERTIUM *Aidana Karibayeva* 28

REALISATION OF STATISTICAL MACHINE TRANSLATION BASED ON A PARALLEL TATAR-RUSSIAN CORPUS OF LEGAL TEXTS *Aliya Mirzagitova* .. 39

THE HISTORY OF TRANSLATION IN YAKUTIA : ACHIEVEMENTS AND PROBLEMS *Dr. Alina Nakhodkina* 50

RESEARCH OF PROBLEM OF THE SEMANTIC ANALYSIS AND SYNTHESIS OF PRETEXTS IN THE RUSSIAN-KAZAKH MACHINE TRANSLATION *Diana Rakhimova* 59

EXPERIENCE OF CREATION OF TATAR-RUSSIAN STATISTICAL MACHINE TRANSLATION IN YANDEX *Andrey Sokolov, Andrey Egorov, Sergey Gubanov, Dmitriy Khristich, Mariya Schmatova, Irina Galinskaya, Alexey Baytin* 67

A FREE/OPEN-SOURCE MACHINE TRANSLATION SYSTEM FOR ENGLISH TO KAZAKH *Aida Sundetova, Mikel Forcada, Francis Tyers*..... 78

AUTOMATON MODELS OF THE MORPHOLOGY ANALYSIS AND THE COMPLETENESS OF THE ENDINGS OF THE KAZAKH LANGUAGE *Ualsher Tukeyev* 91

SECTION 2 . **LINGUISTIC SOFTWARE**

STUDY ON FREQUENCY STATISTIC OF KAZAKH COMMON-USE WORD *Gulila Altenbek* 101

SECTION 1 | **MACHINE TRANSLATION TECHNOLOGIES**

STUDY OF THE PROBLEM OF CREATING STRUCTURAL TRANSFER RULES AND LEXICAL SELECTION FOR THE KAZAKH-RUSSIAN MACHINE TRANSLATION SYSTEM ON APERTIUM PLATFORM

Abduali Balzhan[1], Akhmadieva Zhadyra[2], Zholdybekova Saule[3], Tukeyev Ualsher[4], Rakhimova Diana[5]

[1] KazNU named after Al-Farabi, Almaty, Kazakhstan
balzhan_5696@mail.ru
[2] KazNU named after Al-Farabi, Almaty, Kazakhstan
[3] KazNU named after Al-Farabi, Almaty, Kazakhstan
[4] KazNU named after Al-Farabi, Almaty, Kazakhstan
[5] KazNU named after Al-Farabi, Almaty, Kazakhstan

Active integration of Kazakhstan into the world community and the increasing volume of information flow between our country and its foreign partners, and a real need of different segments of population for operational machine translation while using the Internet, determine the relevance of machine translation between the Kazakh language and various major world languages, like English, Russian, French, German, and recently, Chinese languages, as well as in the vice versa machine translation. The priorities of information interaction for the population of Kazakhstan with foreign partners and internally are mainly defined by interaction in three languages: Kazakh, English and Russian. In this regard, it is highly relevant to have highly efficient instrumental support machine translation for the trilingual language interaction. So are actual research and development industrial quality machine translation systems from Russian language to Kazakh language, and vice versa. Analysis of the state of research in the field of machine translation from Russian into