



Л. Н. ГУМИЛЕВ АТЫНДАҒЫ
ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л. Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
ЕВРАЗИЙСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ ИМ. Л. Н. ГУМИЛЕВА

THE MINISTRY OF EDUCATION AND SCIENCES OF REPUBLIC KAZAKHSTAN
L.N.GUMILYOV EURASIAN NATIONAL UNIVERSITY

**«Түркі тілдерін компьютерлік өңдеу»
атты I халықаралық конференция
ЕҢБЕКТЕРІ**

**ТРУДЫ
I Международной конференции
"Компьютерная обработка тюркских языков"**

**PROCEEDINGS
Of the I International Conference
on Computer processing of Turkic Languages
(TurkLang-2013)**

АСТАНА, 2013

УДК 81'322
ББК 81.1
Т 90

Т 90 ТҮРКІ ТІЛДЕРІН КОМПЬЮТЕРЛІК ӨНДЕУ. Бірінші халықаралық конференция:
Еңбектері/ Астана: Л.Н.Гумилев атындағы ЕҰУ баспасы, 2013- 328 бет
КОМПЬЮТЕРНАЯ ОБРАБОТКА ТЮРКСКИХ ЯЗЫКОВ. Первая международная конференция:
Труды. – Астана: ЕНУ им. Л.Н. Гумилева, 2013. – 328 с.

ISBN 978-601-7454-85-2

Жинақта «Түркі тілдерін компьютерлік өңдеу» атты I халықаралық конференция қатысушыларының баяндамалары енген.

Компьютерлік лингвистика бағыты бойынша оқитын студенттерге, магистранттарға, докторанттарға және мамандарға арналған.

В сборнике представлены доклады участников I международной конференции «Компьютерная обработка тюркских языков».

Предназначен для студентов, магистрантов, докторантов и специалистов специализирующихся в области компьютерной лингвистика.

Техникалық редакция: Бурибаева А.К.
Муканова А. С.
Ергеш Б.Ж.
Елибаева Г.З.

УДК 81'322
ББК 81.1

ISBN 978-601-7454-85-2

© Л.Н.Гумилев атындағы ЕҰУ, 2013

**МӘТІНДЕРДІ СЕМАНТИКАЛЫҚ ӨНДЕУ ЖҮЙЕЛЕРІ
СИСТЕМЫ СЕМАНТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ
SYSTEMS OF SEMANTIC TEXT PROCESSING**

- 1 **Бекманова Г.Т., Жеткенбай Л.**
*Л.Н. Гумилев атындағы Еуразия ұлттық университеті, «Жасанды зерде» ҒЗИ,
Астана* 237
ҚАЗАҚ ТІЛІНІҢ КҮРДЕЛІ СӨЗДЕРІН ФОРМАЛДАУ НЕГІЗІНДЕ ЖАСАУ
- 2 **Бекманова Г.Т., Жеткенбай Л.,**
*Л.Н. Гумилев атындағы Еуразия ұлттық университеті, «Жасанды зерде» ҒЗИ,
Астана* 243
ҚАЗАҚ КҮРДЕЛІ СӨЗДЕРІН ТҮРЛЕНДІРУДІҢ СЕМАНТИКАЛЫҚ МОДЕЛІ
- 3 **Ергеш М.**
Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана 248
ҚҰЖАТТАРДАҒЫ КІЛТТІК СӨЗДЕРДІ ВЕКТОРЛЫҚ МОДЕЛЬ АРҚЫЛЫ ІЗДЕУ
- 4 **Хакимов М.Х., Арипов М.М.**
*Национальный Университет Узбекистана им. Мирзо Улугбека (г. Ташкент,
Республика Узбекистан)* 250
СЕМАНТИЧЕСКИЕ БАЗЫ РУССКОГО ЯЗЫКА

**МАШИНАЛЫҚ АУДАРУ ЖҮЙЕЛЕРІ
СИСТЕМЫ МАШИННОГО ПЕРЕВОДА
MACHINE TRANSLATION SYSTEMS**

- 1 **Сулейманов Д.Ш., Гатиатуллин А.Р., Гильмуллин Р.А., Аюпов М.М.**
НИИ “Прикладная семиотика” академии наук Республики Татарстан 256
К РАЗРАБОТКЕ ТАТАРСКО-ТУРЕЦКОГО МАШИННОГО ПЕРЕВОДЧИКА
- 2 **Хакимов М.Х.**
*Национальный Университет Узбекистана им. Мирзо Улугбека (г. Ташкент,
Республика Узбекистан)* 262
МОДЕЛИРУЕМАЯ ТЕХНОЛОГИЯ МАШИННОГО ПЕРЕВОДА
- 3 **Тукеев У.А., Сапақова С.З., Маратқызы А., Өтепова Қ.**
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан 269
ҚАЗАҚША-ОРЫСША МАШИНАЛЫҚ АУДАРМАСЫНЫҢ МӘЛІМЕТТЕР БАЗАСЫ ЖӘНЕ ОНЫҢ ҚҰРЫЛЫМЫ
- 4 **Төкеев У.А., Сапақова С.З.**
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан 276
ҚАЗАҚ ТІЛІНЕН ОРЫС ТІЛІНЕ МАШИНАЛЫҚ АУДАРМА
- 5 **Abdurakhmonova N.Z.**
National University of Uzbekistan named after Mirzo, Tashkent, Uzbekistan 283
GRAMMATICAL ANALYZE IN MACHINE TRANSLATION BETWEEN ENGLISH AND UZBEK

ҚАЗАҚ ТІЛІНЕН ОРЫС ТІЛІНЕ МАШИНАЛЫҚ АУДАРМА

1. Қазақ тілді машиналық аудармашыларға қысқаша шолу

Қазіргі таңда қазақ тілінен өзге тілдерге аударатын программалар, онлайн-аудармалар баршылық, бірақ олардың жұмыс нәтижесі мардымды емес. Оның ең негізгі себебі қазақ тілінің грамматикасының басқа тілдің грамматикасына қарағанда анағұрлым күрделілігі, өзге тілдің грамматикасына ұқсамайтындығында. Еліміздегі қазіргі кезде кеңінен қолданылып жүрген *sozdik.kz*, *soylem.kz*, *sanasoft.kz* секілді онлайн аудармашылармен қатар «Ізет-тілмәш» қолданбалы программасында қазақ тілінен орыс тіліне аудару мүмкіндігі бар. Бірақ бұл аудармашы программалар енгізілген сөздерді аударғанымен сөйлемнің құрылымына, сөз мағынасына аса мән бермейтінін олардан алынған нәтижелерден көре аламыз. Айта кететін жайт осы айтылған машиналық аудармаларда сөйлемдер енгізіп, оларды аударатын болсақ ол сөздердің көп жағдайда орнын өзгертпей, басқа мағыналарын қарастырмайтынын көреміз, яғни түсініксіз аудармаларға грамматика толық қарастырылмағандықтан тап боламыз, оны дәлелдемесекте болады. Сонымен қатар, бұл бағытта атқарылып жатқан іс-шараларды да атап кететін болсақ, «ағылшын- қазақ» машиналық аударма бағытында *Apertium* програмасы Микель L. Forcada (Испания) басшылығымен және әл-Фараби атындағы Қазақ Ұлттық Университетінің қолдауымен, *free/* машиналық аударманың ашық кодты платформасы пайдаланылуда. *Apertium* – бұл машиналық аударма жүйесінің ашық кодасын құруға арналған құралдардың жиыны, әсіресе өзара байланысқан тілдер жұбы үшін ыңғайлы, оның құрамына ашық лингвистикалық мәліметтерге арналған бірнеше сөздіктер, техникалық қызмет көрсету т.с.с. енетінін білеміз. Осы бағдарлама негізінде «қазақ- татар» тілдер бағытындағы ашық кодалы жүйе құрып, онымен қарқынды айналысып жатқан ғалымдарды: Ильнар Салимзянов, Джонатан Вашингтон және Фрэнсис Туерс атап кетуге болады.

Ұсынылып отырған жұмыс қазақ-орыс бағытында құрылған машиналық аударма жүйесінің негізгі жұмыс істеу принциптеріне, қазіргі таңда туындаған мәселерге тікелей байланысты. Жұмыс нәтижесінде шағын «**kaz-rus translator**» қолданбалы программасы жасалынды және одан әрі дамыту үстіндеміз, бұл программа *Visual Studio 2010* және *SQLite* орталарында орындалды.

2. Қазақ тілінен орыс тіліне аудару барысындағы морфологиялық талдау сұлбасы

Қазіргі уақытта машиналық аударма барысында бірнеше бөліктерден тұратын күрделі жүйе қолданылады, мысалы:

- Морфологиялық талдау – мәтіндегі сөздерді талдау
- Синтаксистік талдау – сөйлемдерді, грамматиканы және сөздер арасындағы байланыстарды талдау;
- Семантикалық талдау – белгілі бір пәндік аймаққа бағытталған деректер қоры негізінде әр сөйлемнің мағынасын талдау.
- Прагматикалық талдау- өзіндік мәліметтер қоры негізінде белгілі бір контекстің ауқымында сөйлемнің мағынасын талдау.