

TurkLang-2018

**VI International Conference on Computer Processing
of Turkic Languages**

**Tashkent, Uzbekistan
18-20 October, 2018**

ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ ИМЕНИ АЛИШЕРА НАВОИ
АКАДЕМИЯ НАУК РЕСПУБЛИКИ ТАТАРСТАН
ИНСТИТУТ ПРИКЛАДНОЙ СЕМИОТИКИ
ЕВРАЗИЙСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ Л. Н. ГУМИЛЁВА
МИНИСТЕРСТВА ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН
НИИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

VI МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ
ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ
ТЮРКСКИХ ЯЗЫКОВ

«TURKLANG-2018»

(труды конференции)



ТАШКЕНТ
ИЗДАТЕЛЬСКО-ПОЛИГРАФИЧЕСКИЙ
ДОМ «NAVOIY UNIVERSITETI»
2018

С 23

УДК 811.512.1 (063)

ББК 81.2 ТЮРК (я43)

Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) –Ташкент: Издательско-полиграфический дом «NAVOIY UNIVERSITETI», 2018. – 298 с.

Научные редакторы:

PhD Н.З. Абдурахмонова;

к.т.н. А.Р. Гатиатуллин

Сборник содержит материалы Шестой Международной конференции по компьютерной обработке тюркских языков «TurkLang-2018» (Ташкент, Узбекистан, 18–20 октября 2018 г.)

Данная публикация предназначена для научных работников, преподавателей, аспирантов и студентов, специализирующихся в области компьютерной лингвистики и ее приложений.

(Издается в авторской редакции)

ISBN 978-9943-5635-1-3

Издание рекомендовано к публикации Постановлением №8 Ученого совета Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 28 декабря 2018 года.

© Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018», 2018
© Издательско-полиграфический дом «NAVOIY UNIVERSITETI», 2018

ПРЕДИСЛОВИЕ

В этом году на базе Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои прошла уже шестая международная конференция по компьютерной обработке тюркских языков TurkLang-2018. Предыдущие конференции прошли в Астане (2013), Стамбуле (2014), Казани (2015, 2017), Бишкеке (2016). География проведения, представленные труды и состав участников конференции подтверждают, что в настоящее время тематика конференции продолжает оставаться весьма актуальной.

Целью серии международных конференций TurkLang является создание пространства совместных компьютерных лингвистических исследований для тюркских языков. На конференции представляются качественно новые результаты, связанные с разработкой компьютерных лингвистических приложений для тюркских языков. Как известно, тюркские языки обладают сложной и самобытной грамматической и семантической системами, поэтому простой перенос решений, полученных на материале других языков (в том числе английского и русского) порой практически не возможен.

Мы надеемся, что проведение этой конференции на базе Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои послужит очередным толчком для развития в стенах нашего университета направления компьютерной лингвистики.

В сборник трудов включены статьи участников VI Международной конференции по компьютерной обработке тюркских языков «TurkLang-2018» (Ташкент, Узбекистан, 18–20 октября 2018 г.). Участниками конференции, учеными и специалистами из Узбекистана, Казахстана, Кыргызстана, Турции, Азербайджана, России (Татарстан, Башкортостан, Москва, Саха (Якутия), Чувашия, Тува, Крым и др.), Португалии, были представлены доклады, посвященные актуальным проблемам компьютерной и когнитивной лингвистики в плане разрешения их в контексте тюркских языков. В ходе конференции активно и плодотворно обсуждались вопросы разработки формальных лингвистических моделей, электронных корпусов, систем машинного перевода, речевых технологий, а также проблемы, связанные с функционированием национальных языков в Интернет-технологиях. Участники отметили конструктивность обсуждения на секциях и круглых столах проблем разработки общей терминологии, общей системы обозначений лексико-грамматических категорий, использования для реализации своих национальных проектов аналогичных подходов, методов и технологий, особенно с учетом близости тюркских языков практически во всех компонентах, включая лексику, морфологию, синтаксис и семантику.

Тематика конференции находится в постоянном развитии. В список новых обсуждаемых тем в 2018 году включен совместный проект по созданию компьютерной онтологии тюркской грамматики. Участниками проекта являются ученые Евразийского национального университета имени Л.Н.Гумилева, Кыргызского государственного технического университета имени И.Раззакова, Стамбульского технического университета, Академии наук Республики Татарстан, а также нашего университета. Как показывает обсуждение реализации проекта и задач по этой проблематике и путей их решения, унификация систем понятий и терминов не является тривиальной практической задачей и требует теоретического пересмотра многих традиционных грамматических описаний.

**Ректор Ташкентского государственного университета
узбекского языка и литературы имени Алишера Навои,
профессор Ш. С. Сирожиддинов**

ПРОГРАММНЫЙ КОМИТЕТ

1. Сирожиддинов Шухрат Самариддинович (Ташкент, Узбекистан) – председатель
2. Сулейманов Джавдет Шевкетович (Казань, Татарстан, Россия) – сопредседатель
3. Шарипбаев Алтынбек Амирович (Астана, Казахстан) – сопредседатель
4. Ешреф Адалы (Стамбул, Турция)
5. Алтынбек Гулила (Урумчи, Китай)
6. Гатиатуллин Айрат Рафизович (Казань, Татарстан, Россия)
7. Дыбо Анна Владимировна (Москва, Россия)
8. Желтов Валериан Павлович (Чебоксары, Чувашия, Россия)
9. Исраилова Нелла Амантаевна (Бишкек, Кыргызстан)
10. Кубединова Ленара Шакировна (Симферополь, Крым, Россия)
11. Мамедова Масума Гусейновна (Баку, Азербайджан)
12. Офлазер Кемаль (Доха, Катар)
13. Садыков Ташполот (Бишкек, Кыргызстан)
14. Салчак Аэлига Яковлевна (Кызыл, Тыва, Россия)
15. Сиразитдинов Зиннур Амирович (Уфа, Башкортостан, Россия)
16. Татевосов Сергей Георгиевич (Москва, Россия)
17. Торотоев Гаврил Григорьевич (Якутск, Саха, Россия)
18. Арипов Мирсаид Мирсиддинович (Ташкент, Узбекистан)

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ

1. Сирожиддинов Шухрат
2. Мухамедова Саодат
3. Абдурахмонова Нилуфар
4. Гатиатуллин Айрат
5. Хакимов Муфтах
6. Аббасова Татьяна
7. Участники лаборатории «Компьютерная лингвистика» в ТашГУУЛ

Рецензенты:

д.ф.н., профессор **Б. Менглиев (Узбекистан)**

PhD, профессор **Карлос Гомез (Испания)**

The conclusion is if not to require too much from the MT, then perhaps the dreams of qualitative machine translation will be soon fulfilled. Afterwards, we must remember that a good translation of the text is not only creative, but also quite labor-intensive work. Even a very good translation, as a rule, needs an editorial amendment. As for the creative part, in the nearest future computer-human competition will be probably won by the artificial intelligence created by a man.

REFERENCES:

1. Blatz, J., Fitzgerald, E., Foster, G., Gandrabur, S., Goutte, C., Kulesza, A., Sanchis, A., and Ueffing, M. Confidence Estimation for Machine Translation.- In Proceedings of COLING: Geneva, 2004.- P. 315-321
2. Byrne Jody Technical Translation: Usability Strategies for Translating Technical Documentation. -Dordrecht: Springer, 2006
3. Hutchins W. J. Current commercial machine translation systems and computer-based translation tools: system types and their uses // International J. of Translation, 2005. Vol. 17. № 1-2.- P. 5-38.
4. Turian J.P., Shen, L., and Melamed, I.D. Evaluation of Machine Translation and its Evaluation.- In Proceedings of MT Summit IX: New Orleans, USA, 2003.- P. 23-28
5. Ulitkin, I. Computer-assisted Translation Tools: A Brief Review.// Translation Journal, Vol. 15, No. 1, January 2011.
6. Марчук Ю. Н. Компьютерная лингвистика: Учеб. Пособие: М., 2007. — 317 с.



KOMPYUTER ANALIZI VA INGLIZ TILIDAGI GAPLARNI O'ZBEK TILIGA TARJIMA QILISH ALGORITMI

S. Muhamedova, Alisher Navoiy nomidagi Toshkent davlat universiteti, o'zbek tili va adabiyoti universiteti, Toshkent, prof.s.muhamedova@gmail.com

Annotatsiya: mazkur maqolada kompyuter analizi va ingliz tilidagi gaplarni o'zbek tiliga tarjima qilish algoritmlari haqida so'z boradi. Unda ingliz tilidan o'zbek tiliga kompyuterda tarjima qilish dasturi vositachi tilsiz amalga oshiriladi. Shuningdek, maqolada tarjima gaplarning sintaktik analiz qilish algoritmlari va formal modellari bazasi asosida ketma-ket amalga oshirilishi ko'rsatib berilgan.

Tayanch so'zlar: kompyuter analizi, tarjima qilish algoritmi, vositachi til, sintaktik analiz, formal modellar bazasi.

КОМПЬЮТЕРНЫЙ АНАЛИЗ И АЛГОРИТМ ПЕРЕВОДА ПРЕДЛОЖЕНИЙ С АНГЛИЙСКОГО НА УЗБЕКСКИЙ

С. Мухамедова, Ташкентский государственный университет узбекского языка и литературы им. Алишера Навои, Ташкент, Узбекистан, prof.s.muhamedova@gmail.com

Аннотация: в статье рассматривается алгоритм компьютерного анализа перевода предложений с английского на узбекский язык. В данной программе перевода предложений с английского на узбекский язык, перевод осуществляется без языка-посредника. Статья описывает базовые формальные модели и алгоритмы синтаксического анализа предложений.

Ключевые слова: компьютерный анализ, алгоритм перевода, язык-посредник, синтаксический анализ, база формальных моделей.

COMPUTATIONAL ANALYSIS AND ALGORITHM FOR TRANSLATING SENTENCES FROM ENGLISH INTO UZBEK

*S. Muhamedova, Alisher Navo'i Tashkent State University of the Uzbek
Language and Literature, Tashkent, Uzbekistan, prof.s.muhamedova@mail.ru*

Annotation: this article contains algorithms for computer analysis of the translation of sentences from English into Uzbek. In the program of translating sentences from English into Uzbek, the translation is carried out without an intermediary language. The article also indicates on the basis of formal models a sequence of algorithms for the syntactic analysis of sentences

Key words: computer analysis, translation algorithm, intermediate language, syntactic analysis, base of formal models.

Ingliz tilidagi matnlarni o'zbek tiliga va aksincha, o'zbek tilidagi matnlarni ingliz tiliga o'zgarishning ommaviy kompyuter metodlarini qo'llash, tillarni kompyuter yordamida o'qitish, matnlarni baholash, matnlarni tahrirlash eng dolzarb muammolar hisoblanadi.

Inglizcha-o'zbekcha va o'zbekcha-inglizcha kompyuter tarjimasini dasturlari ham juda muhim ahamiyatga ega. Ma'lumki, ingliz va o'zbek tillari leksik-grammatik xususiyatlariga ko'ra bir-biridan tubdan farq qiladi. Shuning uchun ingliz tilidan o'zbekchaga va o'zbekchadan inglizchaga kompyuter tarjimasini yaratish o'ziga xos qiyinchiliklarni keltirib chiqaradi. Hozirgi kunda rus tili vositasida ingliz tilidan o'zbek tiliga avtomatik tarjima qiluvchi dasturlarning versiyalari e'lon qilingan.

Ammo biz taklif qilayotgan ingliz tilidan o'zbek tiliga kompyuterda tarjima qilish usuli vositachi tilsiz amalga oshiriladi. Ta'kidlash lozimki, mazkur ish gaplarning sintaktik analiz qilish algoritmlari va formal modellari bazasi asosida amalga oshiriladi.

INGLIZ TILIDAN O'ZBEK TILIGA KOMPYUTERDA TARJIMA QILISH ALGORITMI

Algoritm quyidagi vazifalarni hal qilish uchun mo'ljallangan:

I. Analiz-bunda ingliz tilidagi gap quyidagi soddalashtirilgan model ramkasida tahlil qilinadi:

Ushbu model faqat sodda gaplarni tahlil qiladi.

Gapning har bir bo'lagi bitta so'zdan iborat bo'ladi.

Gaplarda aniqlovchilar bo'lmaydi.

Gaplarning standart tiplari ko'rib chiqiladi (darak gap (ega + kesim + harakatchi + hol), so'roq, inkor va so'roqchilik gaplar).

5. Fe'lning quyidagi tuslanishli shakllarni qamrab oluvchi kesimli gaplar ko'rib chiqiladi:

a) shaxs (I, II, III shaxs);

b) son (birlik va ko'plik);

d) zamon (Past, Present, Future);

e) harakat tipiga ko'ra (Simple, Continuous, Perfect, Perfect-Continuous)

f) maylga ko'ra ((Indicative mood)

g) nisbatga ko'ra (Active i Passive voices).

II. Tarjimada gaplar ingliz tilidan o'zbek tiliga o'giriladi. Algoritm quyidagi gaplardan tashkil topadi:

Gap kiritiladi;

Gapning har bir so'zi I massivining elementlariga qo'shiladi;

I massivining elementlari yordamida gap elementlari bilan taqqoslanadi, bu taqqoslashda

taqqoslashda olmosh, ko'makchilar, ko'makchi

va modal fe'llar, artikllar va noto'g'ri fe'llar ro'yxati mavjud bo'ladi;

4) agar so'zlar yordamchi lug'atda

topilmasa, unda taqqoslash maxsus lug'at

yordamida davom ettiriladi;