ЎзМУ ХАБАРЛАРИ

ВЕСТНИК НУУз

ACTA NUUz

МИРЗО УЛУҒБЕК НОМИДАГИ ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ ИЛМИЙ ЖУРНАЛИ

ЖУРНАЛ 1997 ЙИЛДАН ЧИҚА БОШЛАГАН

2019
1/3

Ижтимоий-гуманитар фанлар туркуми

ТОШКЕНТ – 2019

МУНДАРИЖА

ТАРИХ

ФАЛСАФА. ПЕДАГОГИКА. МЕТОДИКА. СОЦИОЛОГИЯ

ФИЛОЛОГИЯ

O'ZBEKISTON MILLIY UNIVERSITETI XABARLARI, 2019, [1/3]
ISSN 2181-7324

Nilufar ABDURAKHMONOVA,
PhD Associate professor of the department of Information Technology, TSUULL
E-mail:abdurahmonova.1987@mail.ru
Anvar TURDIALIYEV,
2-course master student, TSUULL
anvarbekturdaliyev@list.ru

THE USE OF MORPHOLOGICAL ONTOLOGICAL MODELS FOR TURKISH-UZBEK MACHINE TRANSLATION

Abstract

The article deals with ontological models of morphological categories of Uzbek language comparing Turkish. In this paper analyzed grammatical relations and models in order to create meta language for NLP and other linguistic technology in the frame of Turkic languages under the project of «AP05132249 Processing of electron thesaurus of Turkic languages for creation multilingual information retrieval system and extracting knowledge» on agreement №132 on «12 » March 2018.

Key words: protégé, NLP, Machine translation, Uzbek, Turkish, ontological models.

ИСПОЛЬЗОВАНИЕ МОРФОЛОГИЧЕСКИХ ОНТОЛОГИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ТУРЕЦКО-УЗБЕКСКОГО ПЕРЕВОДА МАШИНЫ

Аннотация

В статье рассматриваются онтологические модели морфологических категорий узбекского языка в сравнении с турецким. В данной статье проанализированы грамматические отношения и модели с целью создания метаязыка для НЛП и других лингвистических технологий в рамках тюркских языков в рамках проекта «AP05132249 Обработка электронного тезауруса тюркских языков для создания многоязычной информационно-поисковой системы и извлечения знаний» на Соглашение №132 от «12» марта 2018 года.

Ключевые слова: protégé, NLP, машинный перевод, узбекский, турецкий, онтологические модели.

TURKCHA-O'ZBEKCHA MASHINA TARJIMASI UCHUN MORFOLOGIK ONTOLOGIK MODELLARNING QO'LLANILISHI

Annotatsiya

Maqolada turk va o'zbek tillarning qiyosiy jihatdan morfologik kategoriyalarining ontologik modellari analiz qilingan. Shuningdek, «AP05132249 Обработка электронного тезауруса тюркских языков для создания многоязычной информационно-поисковой системы и извлечения знаний» nomli loyiha doirasida amalga oshirilayotgan NLP va boshqa lingvistik texnologiyalar uchun metatilni yaratish uchun grammatik munosabatlar tahlili bayon etilgan.

Kalit so'zlar: protégé, NLP, mashina tarjimasi, o'zbek tili, turk tili, ontologik modellar.

Introduction. Due to globalization of the world is making disappear obstacles between communications among the people. Machine translation (MT) as the means of connecting people with different types of information (synchronic and written) has already become easy and flexible tool for human being to save money and time.

Recent tackling many problems regarding to development computer technologies and linguistic resources in the sphere of computational linguistics and NLP (Natural language processing) though decade ago there were some of challenges natural language in the scope of MT. Hence, machine translation is considered high consolidation language technology as an artificial intelligence. Now as we can see different type of MT systems based on diverse approaches in order to solve linguistic issues like homonymy and polysemy.

There are different approaches to implement MT: Rule based (interlingual, transfer, dictionary based), Statistical (example based or corpus based MT), Knowledge based and Hybrid MT. Each technologies has pros and cons regarding to what resources are available.

Our work has been implemented based on rule based machine translation (RBMT) system through Turkish and Uzbek relied on building ontological models of languages. MT between closely related languages is easier than between not related languages. RBMT is suitable one that grammatical rules in common, henceforth it relies on built-in linguistic rules and a number dictionaries for each language pair. Owing to the sub approaches in rule-based Machine translation (direct, transfer-based, interlingua) there will be opportunity to construct ontological models in the frame of close related languages like Uzbek and Turkish. It is obvious that both languages comprised in one group language family. Henceforth, some grammatical rules might be the same for morphological and syntactic parsing in machine translation technology for related languages. Due to being multiple morphemes of Turkic languages, it makes

to complex the process of translation if there are not related.

Related work

Some implementation has been done in the scope of Turkic languages. For example, between Turkmen and Turkish machine translation system relies on ambiguous lexical and morphological transfer augmented with target side rule-based repairs and rescoring with statistical language models [A. Guneyd Tantug, 2007].

Studies on machine translation between close languages are generally concentrated around certain **Slavic languages (e.g., Czech→Slovak, Czech→Polish, Czech→Lithuanian (Hajic et** al., 2003)) and languages **spoken in the Iberian Peninsula (e.g., Spanish↔Catalan (Canals et al., 2000), Spanish↔Galician (Corbi-**Bellot et **al., 2003) and Spanish↔Portugese (Garrido-**Alenda et **al., 2003); Turkic languages Turkman↔Turkish (A.** Guneyd Tantug, Eshref Adali, Kemal Oflazer), Turkish **↔Crimean Tatar (Kemal Altintas, Ilyas Cicekli).**

In this case, the algorithm of machine translation authors [Guneyd Tantug, 2007] indicates the stages for analyzing:

1. Source Language (SL) Morphological Analysis
2. SL Morphological Disambiguation
3. Multi-Word Unit (MWU) Recognizer
4. Morphological Transfer
5. Root Word Transfer
6. Statistical Disambiguation and Rescoring (SLM)
7. Sentence Level Rules (SLR)
8. Target Language (TL) Morphological Generator

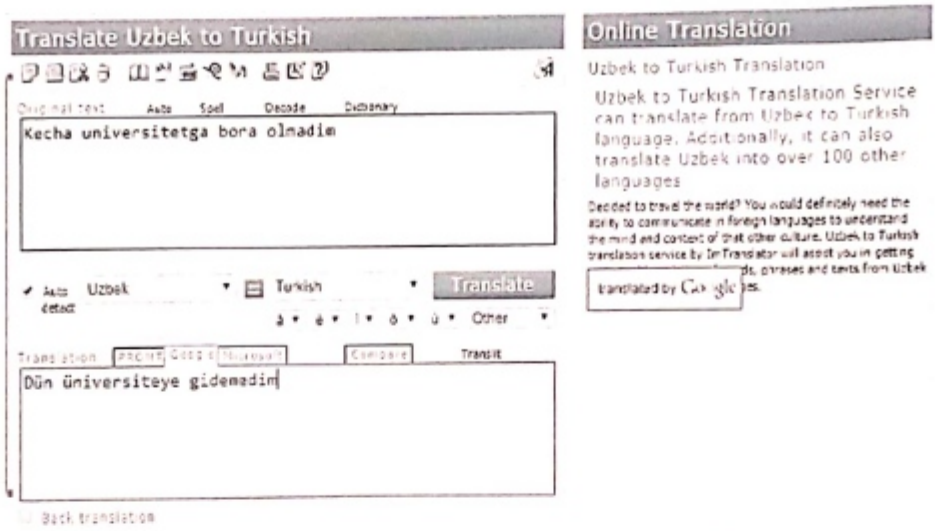
Between Crimean Tatar and Turkish authors [Kemal Altintas, 2001: 41] figure out following stages of algorithm:

- Morphological analysis of Turkish text
- Morphological disambiguation
- Application of context dependent and grammatical translation rules
- General one-to-one translation of words.
- Morphological generation of Crimean Tatar text

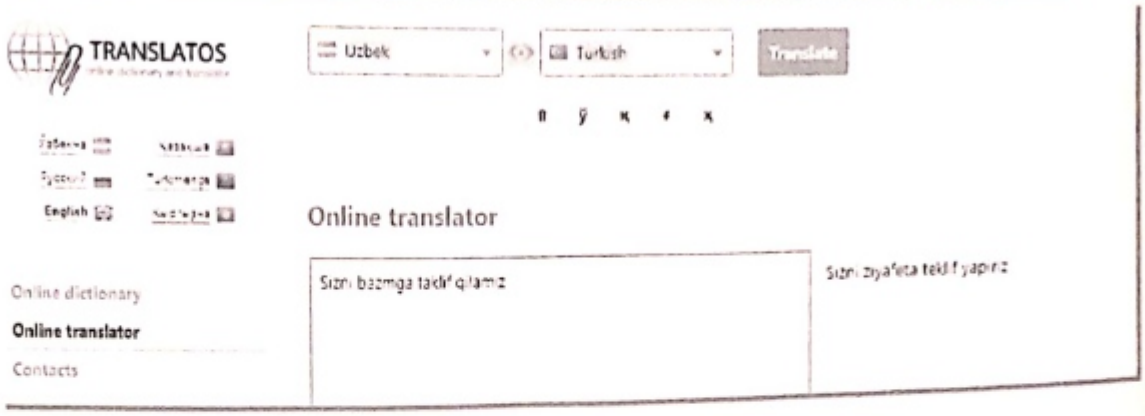
We consider above pointed stages are enough for Turkic languages analysis, because of similarities of grammatical categories and rules.

There are some online services which implement for Turkish-Uzbek machine translation.

For example: http://imtranslator.net/translation/uzbek/to-turkish/translation/



Moreover http://www.translatos.com/en/translate/uzbek-turkish/ is also open available online translator for the languages.



https://beta.apertium.org/index.rus.html this platform is lack of information Uzbek language resources. Comparing and analyzing aforementioned both applications are not considered perfect translation one. Because there are some problems with analytical word forms and combination of morphemes from Uzbek into Turkish.

ONGOING WORK IN THE FIELD

It is known that machine translation is a huge problem for any language if there is lack of resources. However, it can be considered as a very large problem for Uzbek language than others. Because as other Turkic languages Uzbek is very non structured language and applying some strike method to it is very difficult. Some of its difficulties has been mentioned above. According to these issues, it can be useful that if we will create a method or program for this language which analyze its parts. That, it should identify type and meanings of words in sentences. For this, we should analyze only words very first. It is called **morphoanalyzer**. Using this analyzer we can make a decision about words and their meanings, morphological or other changings in it as well.

So, creating this analyzer also can be divided several steps:

- Identifying a base of words;
- Identifying a type of word base;
- Parsing all affixes added to the word according to base;
- Identifying types of all parsed affixes and noticing them.

In the fact that Uzbek and Turkish word order pattern of syntactic structures is more flexible. For MT not only morphological analyzing, but also syntactic parsing is core issue for the next step. We would like to discuss this issues in the further our research how to implement in this case. Moreover, for now morphological analyzing is the key for syntax also according to syntactic markers given by morphological features somehow.

ONTOLOGICAL MODELS OF MORPHOLOGICAL FEATURES

Ontology is used for formal and specialized concept and relations that belong to exact domain. Having advantage of ontology in NPL to create metalanguage in the sphere of machine translation (mainly, rule-based machine translation) or other purposes (information retrieval system, text analysis, annotation of text). Thanks to ontology, creating structure of information based on systematical and hierarchical data it aids to ease computational processing of the natural language. Effective way to create of ontology is representing OWL. "There are several types of ontologies. The word "ontology" can designate different computer science objects depending on the context. For example, an ontology can be:

- a thesaurus in the field of information retrieval or
- a model represented in OWL in the field of linked-data or
- a XML schema in the context of databases
- etc " [Catherine Roussey, (2011)].

Once analyzed on concrete artifact, a model can support relation and conclusion about important aspects of the underlying sphere. For the reason that any kind of model is a construct of understandings according to a certain conceptualization. Furthermore, ontological model structure defines the set of grammatically construction in terms of the order parts of speech.

Tagging considered the core to understand each element of Noun. Therefore, we have such general results of languages morphological categories of Noun. In fact, there is not syngormonism in Uzbek; we can see more this position in those cases in Kazakh language. However there is unique similarities of grammatical point of view {SIMP, CMPL, FUSW, PAIR, CMPN, ABBR, UNDR, DRVT, COMP, ANIM, INAM, CMMN, PRPR, CNCR, ABST, Num., SG, COL} is common both of languages. Variables of case of Turkish differ from Uzbek:

ases			Uzbek_affixes	Turkish_affixes				

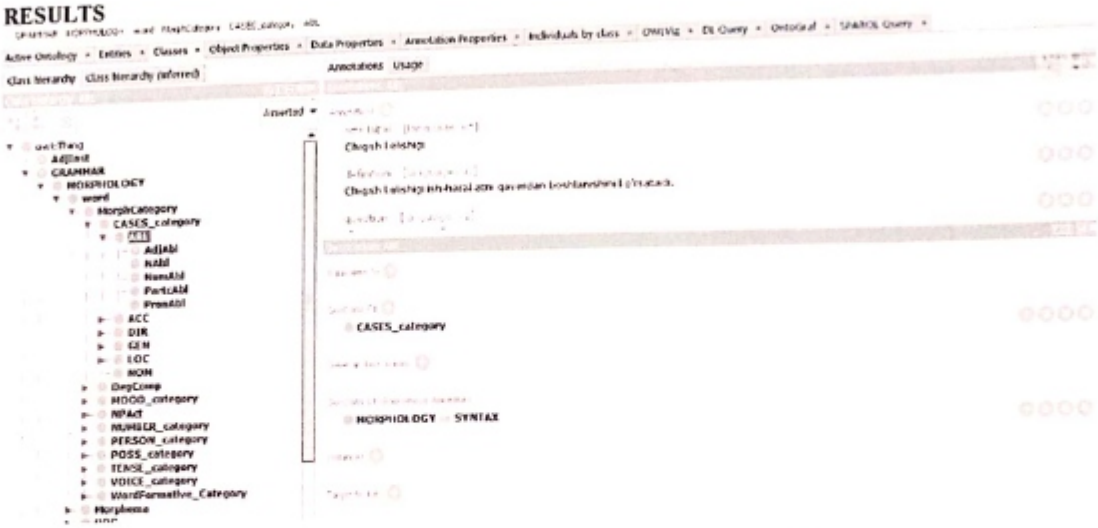
OM	Yal ın hali	Bo sh kelishik	-	-	Ki m? Ne?	Kim? Nima?	e v	u y
EN	Ta mlayan durumu	Qa ratqich	- ning	ın(- in, -un, ün)	Ki min? Neyin?	Kimning ? Nimaning?	e vin	u yning
IR	Yö nelme hali	Jo' nalish	- ga	a(-e)	Ki me? Nereye	Kimga? Nimaga?	e ve	u yga
CC	Beli rtme (Yükleme) hali	Tu shum	- ni	ı, (-i, -u, -ü)	kim i? Neyi	Kimni? Nimani?	E vi	u yni
OC	Bul unma (Kalma) hali	O' rin-payt	- da	de, (-da, -ta, -te)	Ner ede? Ne zaman?	Kimda? Nimada?	E vde	u yda
BL	Ayr ılma (çıkma) hali	Ch iqish	- dan	den, (-dan, -tan, -ten)	Ki mden? Nereden?	Kimdan? Nimadan?	E vden	u ydan
NST	Ara çlılık Durumu			ile, ile	Ki mle? Neyle?		a rabayla	

In spite of diversity of languages, there is commonness of grammatical rules among the Turkic languages. Entities inputted in Protégé as classes including object properties, data properties, individuals, annotation etc. We hope that Ontology grammatical rules of Turkic languages (Uzbek, Kazakh, Tatar, Turkish, Kyrgyz) will be implemented correspondingly and it will service for computational language processing in perspectives.

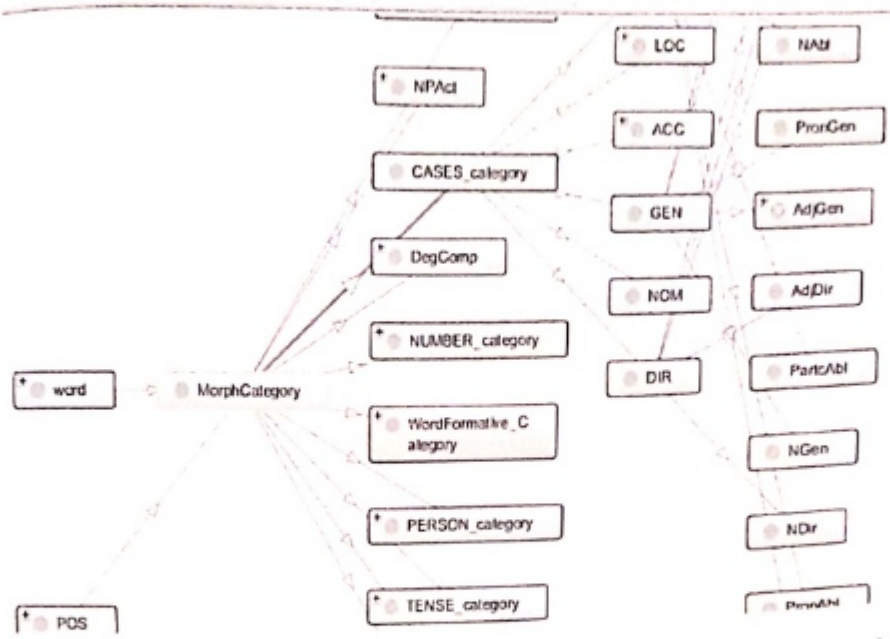
As results of source and target languages ontological models represented in Protégé application ti synchronize properly as linguistic database.

This table indicates that each grammatical models included in classes and thir properties given with annotation. Comprising of information both languaged thhen synchronized when query asked through the pre-process of MT.

RESULTS



Here represented in protégé grammatical features as ontological models:



CONCLUSION

This work developing in the frame of Turkic languages Uzbek, Kazak, Kirgiz, Turkish, and Tatar. Creation of Meta language for Turkic languages is crucial in order to NLP and other special purpose text analysis. **All classes and subclasses with attributes input Protégé** program to work further work. In spite of diversity of languages, there is commonness of grammatical rules **among the Turkic languages. Entities inputted in Protégé** as classes including object properties, data properties, individuals, annotation etc. We hope that Ontology grammatical rules of Turkic languages will be implemented correspondingly and it will service for computational language processing in perspectives.

REFERENCES

1. **Kemal Altıntaş TURKISH to CRIMEAN TATAR MACHINE TRANSLATION SYSTEM, THESIS2001,** -P. 41
2. A. Guneyd Tantug, Eshref Adali, Kemal Oflazer Machine Translation between Turkic Languages / Conference: ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic https://www.researchgate.net/publication/220874083_Machine_Translation_between_Turkic_Languages
3. Catherine Roussey, Francois Pinet, Myoung Ah Kang, and Oscar Corcho. An Introduction to Ontologies and Ontology Engineering (G. Falquet et al., Ontologies in Urban Development Projects, Advanced Information 9 and Knowledge Processing 1, DOI 10.1007/978-0-85729-724-2_2, © Springer-Verlag London Limited 2011)
4. Abdurakhmonova N., Aripov M., Sharipbay A., Razzoqova B.Ontology of grammar rules as example of noun of Uzbek and Kazakh languages AL-Khorazmy international conference // 13-15 September, 2018. P. 37-38
5. Abdurakhmonova N. Ontological model of Uzbek language (as **example morphology) «Computational** Models **in Language and Speech» international conference, Kazan, TEL**-2018. –P. 5-11
6. **Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya), Toshkent: Muharrir, 2018, 180** b.