

МРНТИ 20.23.25; 16.31.21; 28.23.39

Разработка единого метаязыка морфологии тюркских языков

Шарипбай А.А., НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: sharalt@mail.ru
Гатиатуллин А.Р., Институт прикладной семиотики Академии наук Республики
Татарстан, г. Казань, Татарстан, Россия, E-mail: ayrat.gatiatullin@gmail.com
Ергеш Б.Ж., НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: b.yergesh@gmail.com
Қажымұхан Д.А., НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: askarovna0105@mail.ru

В настоящее время в связи с резким увеличением объема информации на естественных языках в интернете и социальных сетях исследование и разработки в области вычислительной лингвистики становятся чрезвычайно актуальными. Как известно, вычислительная лингвистика является новым научным направлением и входит в состав вычислительной науки - информатики. Вычислительная лингвистика включает в себе компьютерную обработку естественных языков (ОЕЯ) – Natural Language Processing (NLP). Создание унифицированной системы разметки для тюркских языков (UniTurk) является актуальной задачей для обработки тюркских языков. Единая система разметок позволит унифицировать разметки, облегчить их понимание и использовать общее программное обеспечение, а также проводить различные исследования по лингво-статистическому сравнительному анализу среди тюркских языков. В статье представлены некоторые результаты, полученные в ходе выполнения запланированных работ над международным проектом по созданию многоязычной онтологии и единого метаязыка морфологии тюркских языков. С помощью онтологических моделей формализованы морфологические правила тюркских (казахского, киргизского, татарского, турецкого и узбекского) языков. Результат этих работ могут быть использованы для решения задач NLP, например, для разметки корпусов, в системах извлечения знаний, системах информационного поиска, машинного перевода и т.д.

Ключевые слова: морфология, тюркские языки, метаязык, тезаурус.

Түркі тілдері морфологиясының бірегей метатілін әзірлеу
Шәріпбай А.Ә., «Жасанды интеллект» ғылыми-зерттеу институты,
Л. Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: sharalt@mail.ru
Гатиатуллин А.Р., Қолданбалы семиотика институты, Татарстан Республикасының
ғылым академиясы, Казань қ., Татарстан, Ресей, E-mail: ayrat.gatiatullin@gmail.com
Ергеш Б.Ж., «Жасанды интеллект» ғылыми-зерттеу институты,
Л. Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: b.yergesh@gmail.com
Қажымұхан Д.А., «Жасанды интеллект» ғылыми-зерттеу институты,
Л. Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: askarovna0105@mail.ru

Соңғы кездегі Интернет пен әлеуметтік желілердегі табиғи тілдегі ақпараттың көлемінің күрт артуы компьютерлік лингвистика саласындағы зерттеулер өте өзекті мәселе. Компьютерлік лингвистика информатика сияқты есептеу ғылымының жаңа бағыты екені белгілі. Компьютерлік лингвистика табиғи тілдерді өңдеу (Natural Language Processing-NLP) есебін пешумен айналысады. Түркі тілдерінің бірегей белгілеу жүйесін әзірлеу түркі тілдерін компьютерлік өңдеу саласының өзекті есебінің бірі. Бірегей белгілеу жүйесі белгілеулерді бірегейлендіріп, оларды түсінуді жеңілдетеді және түркі тілдеріне ортақ қосымшаларды пайдалануға, түрлі лингвистикалық-статистикалық зерттеулер жүргізуге және салыстырмалы талдау жасауға мүмкіндік береді. Бұл мақалада түркі тілдерінің көптілі онтологиясын және бірегей ортақ метатіл құру бойынша орындалып жатқан халықаралық жобада жоспарланған жұмыстардың кейбір нәтижесі көрсетілген. Онтологиялық моделдер арқылы түркі (қазақ, қырғыз, татар, түрік, өзбек) тілдерінің морфологиялық ережелері формалданған. Бұл жұмыстың нәтижелерін NLP есептерін шешу үшін қолдануға болады. Мысалы, корпустарды белгілеуде, білімдерді шығару жүйелері, ақпараттық іздеу жүйелері, машиналық аудары және басқалары.

Түйін сөздер: морфология, түркі тілдері, метатіл, тезаурус.

Development of an unified meta language of the turkic languages morphology

Sharipbay A., Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: sharalt@mail.ru

Gatiatullin A., Institute of Applied Semiotics of the Academy of Sciences Republic of Tatarstan,

Kazan, Tatarstan, Russia, E-mail: ayrat.gatiatullin@gmail.com

Yergesh B., Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: b.yergesh@gmail.com

Kazhymukhan D, Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: askarovna0105@mail.ru

Currently, due to the sharp increase of information amount in natural languages on the Internet and social networks, research and development in the field of computational linguistics is becoming extremely relevant. As is known, computational linguistics is a new scientific field and part of computer science. Computational linguistics includes the Natural Language Processing (NLP). Creating a unified metalanguage for Turkic languages (UniTurk) is an important task for processing Turkic languages. An unified metalanguage system will allow to unify tags, facilitate their understanding and use common software, as well as conduct various studies on linguistic-statistical comparative analysis among the Turkic languages. The article presents some of the results obtained on an international project to create a multilingual ontology and unified metalanguage of the Turkic languages morphology. Using ontological models, the morphological rules of the Turkic (Kazakh, Kyrgyz, Tatar, Turkish, and Uzbek) languages are formalized. The result of these works can be used in the NLP applications, for example, for corpus tagging, in knowledge extraction systems, information retrieval systems, machine translation, etc.

Key words: morphology, turkic languages, metalanguage, thesaurus.

1 Введение

Для компьютерной обработкой любых естественных языков требуются формализация их грамматических (морфологических и синтаксических) правил, разработка алгоритмов анализа и синтеза слов и предложения по этим правилам, программная реализация всех этих алгоритмов, создание тезаурусов по предметным областям аналогично WordNet [1], построение текстовых корпусов (база данных размеченных текстов) и аудио корпусов (база данных размеченных аудиозаписей) и других программ для анализа и обработки текстов.

Начиная с 2013 года усилиями ученых по компьютерной лингвистике из Казахстана, Татарстана, Кыргызстана и Турции проводятся ежегодные международные научно-практические конференции по компьютерной обработке тюркских языков «TurkLang», в которых обсуждаются проблемы компьютерной обработки тюркских языков. В рамках этой конференции организуется традиционный круглый стол UniTurk, где в 2014 году была принята резолюция о разработке унифицированной системы грамматической разметки для электронных текстов на тюркских языках. Тогда же был представлен рабочий вариант унифицированной морфологической разметки, который ориентируется на морфемную структуру тюркской словоформы и должен максимально полно отразить морфологические особенности всех тюркских языков [2].

2 Обзор литературы

Метаязык – язык, предназначенный для описания другого языка, называемого объектным языком. Метаязык – язык лингвистики. метаязык некоторый искусственный язык описывающий естественный язык. Систематическая формализация научного языка [3]. Метаязык обладает следующими свойствами: с помощью его языковых средств можно выразить все, что выразимо средствами объектного языка, и обозначить все знаки, выражения объектного языка, для которых имеются имена; на метаязыке можно говорить о свойствах выражения объектного языка и отношениях между ними; на нем можно сформулировать определения, обозначения, правила образования и преобразования для выражений объектного языка [4]. Имеются известные системы разметки, такие как Penn Treebank [5], система разметки CLAWS [7], который применяется для разметки Британского корпуса [6], система разметки Брауновского корпуса, в американском национальном корпусе [8, 9] используются несколько систем разметки, про системы разметки более подробно описаны в [10, 11]. Все эти системы в основном применяются для разметки английского языка.

В настоящее время существует больше десяти электронных корпусов для тюркских языков: корпус казахского языка [12, 13, 14]; корпус татарского языка «Туган тел» [15, 16], корпус турецкого языка [17], корпус крымскотатарского языка [18], чувашского языка [19] и др. Одним из основных компонентов этих корпусов является система морфологической, синтаксической и семантической разметок, базовой среди которых является система морфологической разметок, которая разнятся в некоторых корпусах. Поэтому необходимость создания единой системы разметок не вызывает сомнения. Единая система разметок позволит унифицировать разметки, облегчить их понимание и использовать общее программное обеспечение, а также проводить различные исследования по лингво-статистическому сравнительному анализу среди тюркских языков [20, 21, 22]. Создание унифицированной системы разметки (метаязыка) для тюркских языков (UniTurk) является актуальной задачей.

3 Материал и методы

Для создания такой системы разметок необходим общий ресурс с которым могли бы работать все разработчики тюркских электронных корпусов. Такой ресурс мог бы выполнять роль справочной системы как для разработчиков, так и для

пользователей тюркских электронных корпусов. Наиболее подходящими компонентами такого ресурса, которые соответствуют требуемым условиям являются онтологические модели грамматики тюркских языков. Положительной чертой онтологической модели является возможность поэтапного уточнения ресурса, начиная с описания категорий и концептов, добавляя к ним систему формализованных свойств функций и отношений между ними, а также добавление к ним тезаурусов. В настоящее время существует множество тезаурусов, наиболее известным среди которых является Принстонский WordNet [23]. В этом тезаурусе присутствуют большие фрагменты с описанием лингвистических единиц и лингвистических категорий (Рисунок 1), однако эти фрагменты тезауруса невозможно использовать для работы напрямую с тюркскими языками, так как они не отражают в полной мере все структурные и семантические особенности тюркских языков.



Рисунок 1: Фрагмент Принстонского WordNet

Особенность нашей работы заключается в том, что параллельно ведется работа по пяти тюркским языкам: казахский, татарский, киргизский, узбекский и турецкий. Такая параллельность в процессе построения онтологии грамматики позволяет производить сравнительный анализ лингвистических концептов этих языков и создавать единые концепты для описания их грамматик. Ожидаемым результатом этой работы должен стать единый, многоязычный тезаурус грамматики тюркских языков, в основе которого лежат их лингвистические концепты с системами отношений между ними. Текстовыми входами тезауруса являются лексемы по грамматике тюркских языков.

В таблице 1 представлен фрагмент единого многоязычного тезауруса. В левой части таблицы представлено обозначение концептов в виде названия на английском языке, а также обозначения тэга. Обозначение тэга используется для разметки электронных

корпусов тюркских языков. В центральной части таблицы представлены название тэгов на тюркских языках.

Таблица 1: Фрагмент единого многоязычного тезауруса

Tag	English	Kazakh	Tatar	Kyrgyz	Uzbek	Turkish	Russian
N	Noun	Зат есім	Исем	Зат атооч	От сўз туркуми	Isim (Ad)	Имя существительное
SIMP	Simple Noun	Дара зат есім	Гади исем	Жөнөкөй	Оддий		Простые существительное
CMPL	Complex Noun	Күрдөлі зат есім	ясалма исем	Татаал	Мураккаб		Сложные существительное
FUSW	Fused words	Біріккен зат есім	Кушма исем	Бириккен	Кўшма сўз		слитные (сложные) существительное
PAIR	Pair Noun	Қосарланган зат есім	Парлы исем	Кош	Жуфт сўз		парные существительное
CMPN	Compound Noun	Тіркесті (құрама) зат есім	Тезмә исем	Кошмок	Мураккаб сўз		составные существительное
ABBR	Abbreviations	Қысқарған зат есім	кыскартылма исем	Кыскартылган	Қисқартмалар		сокращенные существительное
UNDR	underivatives Noun	Негізгі зат есім		Тубаса негиз	Туб сўз	Basit isim	непроизводные существительное
DRVT	derivatives Noun	Туынды зат есім	ясалма исем	Туунду негиз	Ясама сўз	Türemis ad	производные существительное
COMP	compound					Birlesik	
ANIM	animate Noun	Жанды зат есім	жанлы исем	Адамзаттык	Жонли	Canlı	одушевленные существительное
INAM	inanimate Noun	Жансыз зат есім	Жансыз исем	Адамзаттык эмес	Жонсиз	Cansız	неодушевленные существительное

Текстовая часть на тюркских языках состоит не только из названия концепта, а

также из описания, вопросов и примеров на этих языках. В таблице 2 представлен фрагмент описания концептов на казахском языке. Аналогичное описание концептов создано для каждого из тюркских языков, представленного в проекте.

Таблица 2: Фрагмент описания концептов на казахском языке

Part of speech	properties	types	types	definition	Questions	example
Зат есім				заттың, құбылыстың, оқиганың атын білдіреді	кім? не?	Адам, бала, кітап
	Құрамы					
	жалаң	Дара зат есім		Бір сөзден жасалған зат есім		бала, ағаш, буын, орақ, жауын, жасау
	күрделі	Күрделі зат есім		кемінде екі я одан да көп түбір морфемадан құралған зат есімдер		ата-ана, ақтас, теміржол
	біріккен		Біріккен зат есім	кемінде екі түбір морфемадан бірігу арқылы жасалған зат есімдер.		Жаңаарқа, Алтынтөбе, Ақтөбе, теміржол
	Қосарланған		Қосарланған зат есім	екі түбір морфемадан қосарланып жасалған зат есімдер		Балашаға, ата-ана
	тіркесті		Тіркесті (құрама) зат есім	кемінде екі я онан астам түбір сөздерден әр қилы жолмен тіркесу арқылы жасалған зат есімдер.		ауылдың маңы, үйдің жаны

Компьютерная реализация онтологии грамматики выполнена в системе Protege (Рисунок 2), одном из наиболее известных инструментов для создания и редактирования онтологий общего назначения [24, 25]. Данный продукт является разработкой Центра биомедицинской информатики Стэнфордского университета (Stanford Center for Biomedical Informatics Research) и распространяется под свободной

лицензией. Программа имеет две реализации: в виде кроссплатформенного десктопного приложения и в виде web-приложения WebProtege.

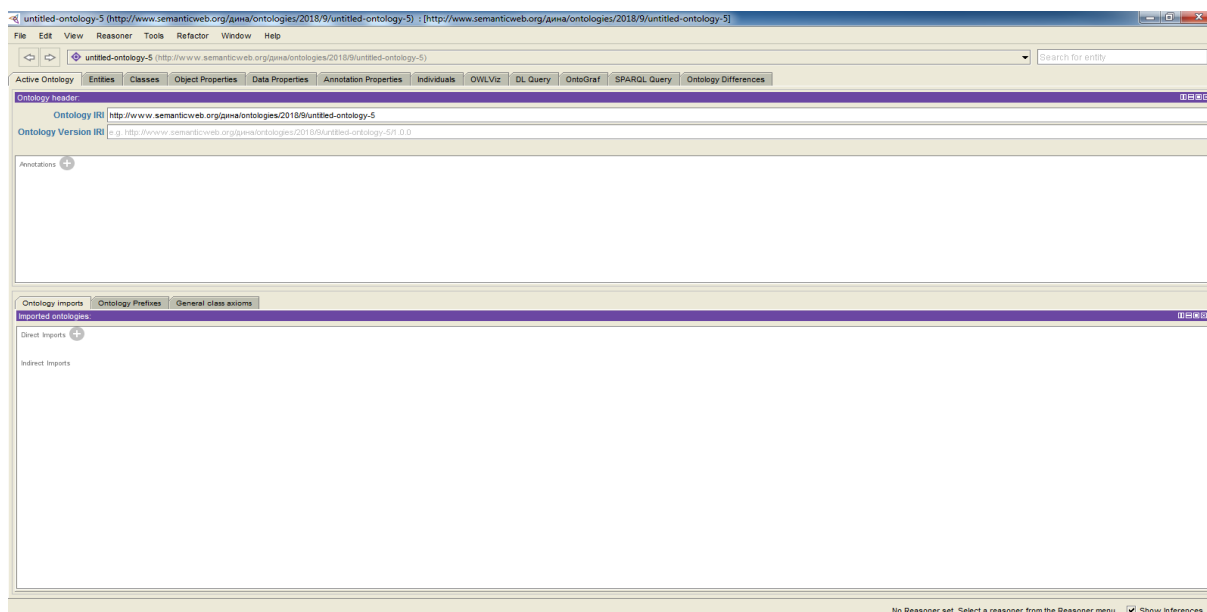


Рисунок 2: Фрагмент интерфейса системы Protege

4 Результаты и обсуждение

В результате исследования получена система единой разметки морфологии тюркских языков(казахский, татарский, киргизский, узбекский и турецкий). В создании метаязыка принимали участие ученые, молодые исследователи из Кыргызстана, Татарстана, Турции и Узбекистана. Также с помощью единого метаязыка разработаны онтологические модели морфологии этих языков. В будущем планируется создание единого метаязыка синтаксиса и их онтологических моделей, а также разработка тезауруса тюркских языков.

5 Заключение

Работа над созданием многоязычного тезауруса грамматик тюркских языков показывает насколько тюркские народы отдаляются друг от друга в создании научной терминологии. Создаваемый многоязычный информационный ресурс в Интернете должен, с одной стороны, способствовать взаимопониманию терминологии между тюркскими языками, а с другой стороны, стать многоязычной базой данных, которая будет использоваться в системах многоязычного поиска, машинного перевода между тюркскими языками, автореферирование тюркских текстов, а также в информационно-справочных и обучающих системах. Машинный перевод между тюркскими языками, многоязычный семантический поиск и извлечение знаний приведет к усилению экономических, образовательных, социальных и культурных

связей между тюркскими странами, так как будет уменьшаться языковой барьер. Тесное научное сотрудничество между коллективами ученых тюркских стран приведет к еще большему расширению научных и академических связей. Однако, также понятно, что создаваемые онтологические модели грамматики тюркских языков являются лишь небольшой частью тех лингвистических ресурсов, которые необходимы для полноценной реализации, перечисленных программных продуктов и технологий. Созданный информационный ресурс будет открытым и возможно его расширение путем добавление концептуальных описаний грамматик других тюркских языков.

6 Благодарности

Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Министерством науки и образования Республики Казахстан (грант № AP05132249, 2018-2020 годы)

Список литературы

- [1] *Wordnet. A large lexical database of English* [Электр.ресурс]. – 2018. – URL: <https://wordnet.princeton.edu> (дата обращения: 10. 09. 2018)
- [2] *Резолюция научно-практического семинара «Унификация систем грамматической разметки в корпусах тюркских языков (семинар UniTurk)* [Электр.ресурс]. – 2014. – URL: <http://www.turklang.tatar.ru/резолюция/> (дата обращения: 10.10.2018)
- [3] *Мельчук И.А.* Курс общей морфологии. Том 1. – Москва - Вена: «Языки русской культуры», Венский славистический альманах, Издательская группа «Прогресс», 1997. – 416 с.
- [4] *Залевская А.А.* Введение в психолингвистику: учебник. – М.: Российск. гос. гуманит. ун-т, 2000. – 382 с.
- [5] *Marcus, Mitchell P.* Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of English: the Penn Treebank. // *Computational Linguistics*. – 1993. – 19(2). С. 313–330.
- [6] *British National Corpus* [Электрон.ресур]. – 2018. – URL: <http://www.natcorp.ox.ac.uk> (дата обращения: 10.09. 2018)
- [7] *Garside, R.* The CLAWS Word-tagging System // In: R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach*. – London: Longman, 1987.
- [8] *The Open American National Corpus* [Электр.ресурс]. – 2018. – URL: <http://www.anc.org> (дата обращения: 10.10.2018)
- [9] *Ide, N.* *The American National Corpus: Then, Now, and Tomorrow* // Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages, Cascadilla Proceedings Project, Somerville, MA, - 2008.
- [10] *Jurafsky, Daniel and James H. Martin.* : *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. – 2nd Edition. – Prentice-Hall, 2009. – 988 p.
- [11] *Nitin Indurkha and Fred J. Damerau.* *Handbook of Natural Language Processing*. – 2nd ed. – Chapman and Hall, 2010. – 704 p.
- [12] *Kazakh Language Corpus* [Электрон.ресур]. – 2018. – URL: <http://kazcorpus.kz/> (дата обращения: 10.09.2018)
- [13] *Makhambetov O., Makazhanov A., Yessenbayev Zh., Matkarimov B., Sabyrgaliyev I., and Sharafudinov A.* 2013. Assembling the Kazakh Language Corpus // In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013. – p. 1022–1031.
- [14] *Мадиева Г.Б., Уматова Ж.М.* Об Алматинском корпусе казахского языка // *Вестник КазНУ. Серия «Филология»*.– Алматы, 2015. – №5 (157). – С. 99-103.

- [15] *Татарский национальный корпус «Туган тел»* [Электр.ресурс]. – 2018. – URL: <http://tugantel.tatar> (дата обращения: 10.10.2018)
- [16] *Galieva A., Khakimov B., Gatiatullin A.* On the Way to the Relevant Grammatical Tagset for the Tatar National Corpus // EPiC Series in Language and Linguistics . Volume 1, CILC2016. 8th International Conference on Corpus Linguistics, 2016. – P.121–129.
- [17] *Turkish National Corpus (TNC)* [Электрон.ресурс]. – 2018. – URL: <http://www.tnc.org.tr/> (дата обращения: 10.10.2018)
- [18] *Kubedinova L., Gatiatullin A.* Morphological tagging of crimean tatar electronic corpus //Proceedings of the international conference «Turkic languages processing» TurkLang-2015. – Kazan, Tatarstan, 2015. – P.331-337
- [19] *Zhelto P.* Morphological annotation system for the national corpus of the chuvash language //Proceedings of the international conference «Turkic languages processing» TurkLang-2015. – Kazan, Tatarstan, 2015. – P. 328-331.
- [20] *Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G.* Ontology modeling of morphological rules of the kazakh and turkish languages // Abstract of the VI international conference «modern problems of applied mathematics and information technology - al-Khorezmiy 2018". – Tashkent, Uzbekistan, 2018. – P. 51-52.
- [21] *Zhetkenbay L., Sharipbay A., Bekmanova G., Kamanur U.* «Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages // Journal of Theoretical and Applied Information Technology, - 2016. – Vol. 91. No.2. – P. 257- 263.
- [22] *Aripov M., Sharipbay A., Abdurakhmonova N., Razakhova B.* Ontology of grammar rules as example of noun of Uzbek and Kazakh languages // Abstract of the VI international conference «modern problems of applied mathematics and information technology - al-Khorezmiy 2018". – Tashkent, Uzbekistan, 2018. – p.37-38.
- [23] *Fellbaum, Christiane.* WordNet and wordnets // In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, - Oxford: Elsevier, 2006. – p. 665-670.
- [24] *Protege* [Электрон.ресурс]. – 2018. – URL: <http://protege.stanford.edu> (дата обращения: 10.10.2018)
- [25] *Musen, M.A.* The Protege project: A look back and a look forward // AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence. - 2015. – 1(4). – P.4-12. DOI: 10.1145/2557001.25757003.

References

- [1] "WordNet. A large lexical database of English accessed September 10, 2018, <https://wordnet.princeton.edu>.
- [2] "Resolution of the scientific-practical seminar "Unification of the systems of grammatical annotation in the Turkic languages corpora (UniTurk seminar) accessed October 10, 2018, <http://www.turklang.tatar.ru/резолуция/>.
- [3] Melchuk I.A. *Kurs obschey morfologii [General morphology course]*. Moskva - Vena: «Progress», Vol 1, 1997.
- [4] Zalevskaya A.A. *Vvedenie v psiholingvistiku [Introduction to Psycholinguistics]*. M.: Rossiysk. gos. gumanit. un-t, 2000.
- [5] Marcus, Mitchell P. Beatrice Santorini, and Mary Ann Marcinkiewicz. "Building a large annotated corpus of English: the Penn Treebank." *Computational Linguistics*.19(2)(1993):313–330.
- [6] "British National Corpus", accessed September 10, 2018, <http://www.natcorp.ox.ac.uk>.
- [7] Garside, R. "The CLAWS Word-tagging System."In: R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman, 1987.
- [8] "The Open American National Corpus", accessed October 10, 2018, <http://www.anc.org>.
- [9] Ide, N. "The American National Corpus: Then, Now, and Tomorrow."Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages. Sommerville, MA, 2008.
- [10] Jurafsky, Daniel and James H. Martin.*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd Edition. Prentice-Hall, 2009.
- [11] Nitin Indurkha and Fred J. Damerau. *Handbook of Natural Language Processing*. 2nd Edition. Chapman and Hall/CRC, 2010.

- [12] "Kazakh Language Corpus accessed September 10, 2018, <http://kazcorpus.kz/>.
- [13] Makhambetov O., Makazhanov A., Yessenbayev Zh., Matkarimov B., Sabyrgaliyev I., and Sharafudinov A. "Assembling the Kazakh Language Corpus" *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing(2013): 1022–1031*.
- [14] Madiyeva G.B., Umatova Zh.M. "Ob Almatinskoy korpusy kazakhskogo yazyka[About Almaty Corpus of the Kazakh language]" *KazNU Bulletin. Philology series. no.5(157) (2015): 99-103*.
- [15] "Tugan Tel" Tatar National Corpus accessed October 10, 2018, <http://tugantel.tatar>.
- [16] Galieva A., Khakimov B., Gatiatullin A. "On the Way to the Relevant Grammatical Tagset for the Tatar National Corpus" *EPiC Series in Language and Linguistics .CILC2016. 8th International Conference on Corpus Linguistics, Volume 1 (2016): 121–129*.
- [17] "Turkish National Corpus (TNC)", accessed October 10, 2018, <http://www.tnc.org.tr/>.
- [18] Kubedinova L., Gatiatullin A. "Morphological tagging of crimean tatar electronic corpus." *Proceedings of the international conference "Turkic languages processing" TurkLang-2015(2015):331-337*.
- [19] Zheltov P. "Morphological annotation system for the national corpus of the chuvash language." *Proceedings of the international conference "Turkic languages processing" TurkLang-2015 (2015):328-331*.
- [20] Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G. "Ontology modeling of morphological rules of the kazakh and turkish languages." *Abstract of the VI international conference "Modern problems of applied mathematics and information technology - al-Khorezmiy 2018"(2018): 51-52*.
- [21] Zhetkenbay L., Sharipbay A., Bekmanova G., Kamanur U. "Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages." *Journal of Theoretical and Applied Information Technology , Vol. 91. No.2 (2016): 257- 263*.
- [22] Aripov M., Sharipbay A., Abdurakhmonova N., Razakhova B. "Ontology of grammar rules as example of noun of Uzbek and Kazakh languages." *Abstract of the VI international conference "Modern problems of applied mathematics and information technology - al-Khorezmiy 2018"(2018): 37-38*.
- [23] Fellbaum, Christiane. "WordNet and wordnets." *Encyclopedia of Language and Linguistics, ed. Brown, Keith et al., 2nd Edition, Oxford: Elsevier (2007): 665-670*.
- [24] "Protege accessed October 10, 2018, <http://protege.stanford.edu>.
- [25] Musen, M.A. "The Protege project: A look back and a look forward." *AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4) (2015): 4-12. doi: 10.1145/2557001.25757003*.