



Method of Sentiment Preservation in the Kazakh-Turkish Machine Translation

Lena Zhetkenbay¹, Gulmira Bekmanova^{1,2}(✉), Banu Yergesh¹,
and Altynbek Sharipbay^{1,2}

¹ L.N. Gumilyov, Eurasian National University, Nur-Sultan, Kazakhstan
jetlen_7@mail.ru, sharalt@mail.ru,
gulmira-r@yandex.kz, b.yergesh@gmail.com
² Nuclear University MEPhI, Moscow, Russia

Abstract. This paper describes characteristics which affect the sentiment analysis in the Kazakh language texts, models of morphological rules and morphological analysis algorithms, formal models of simple sentence structures in the Kazakh-Turkish combination, models and methods of sentiment analysis of texts in the Kazakh language. The studies carried out to compare the morphological and syntactic rules of the Kazakh and Turkish languages prove their closeness by structure. In this respect, we can assume that taking into account sentiment in machine translation for these combinations of languages will give a good result at preserving the text meaning.

Keywords: Sentiment analysis · Kazakh language · Machine translation · Production rules · Rule-based method · Morphological rules · Ontology

1 Introduction

Kazakh language belongs to the Kipchak group [1], whereas Turkish language belongs to Oguz group of Turkic languages [2]. Both of the languages alike other Turkic languages have an agglutinative property, which is characterized by the ability of forming the word formation and word forms by adding affixes (suffixes) to the each root or stem of a particular word. In this case, suffixes change the semantics (meaning) of words which is part of the semantic category that forms new words, and ending - in the structural category, that can change only the composition of words.

In general, translation from one language to another requires changing of alphabet, vocabulary and semantics of the language. Translation is a type of information service which demand will never disappear, on the contrary, it is growing every year. Problems of translation modeling, their presentation in a computer environment are the main problems of applied linguistics and artificial intelligence. Of course, the automation of translation process will increase its effectiveness, as well as expand human relations.

The machine translation is a computer translation of text from one natural language to another natural language, which is equivalent by content [3].

There are high-quality automatic translation systems from Kazakh to Turkish and vice versa, but experience shows that their quality is not very good. It has been stated that among the systems used, the results of translation systems between language

combinations, which do not significantly differ from each other, are much better than translations between language combinations completely different by structure. From such result we can see that a large number of linguistic similarities are very favorable for development of machine translation systems and contribute to successful translation.

Nowadays there are various methods and systems for machine translation of languages [4–6]. The application of any of them depends on the complexity of formalizing the language or the national language entire corpora.

The first research among the Turkic languages was made in an automatic translation from Azerbaijani to Turkish based on a dictionary [7]. The next study on machine translation between Turkic languages was carried out between Crimean Tatars and Turkish [8]. Besides, the problems of machine translation for other Turkic languages, in particular from Turkmen and Uyghur to Turkic languages, are presented in the following papers [9, 10]. The computer analysis of morphology of the Turkmen language and statistical methods were used to eliminate the imprecision between the languages. The architecture of the machine translation system has been developed among the Turkic languages. The most widely used method of two-level morphological analysis was carried out using finite state automats.

The machine translation works are based on grammar rules in languages other than the Kazakh language, including [10–12]. Bilingual data for the Kazakh-Tatar languages [13], Tatar-Bashkir languages [14], a database developed for Crimean-Tatar-Turkish and Kazakh-Russian and Kazakh-English combinations of languages in the Apertium system and a database of structural rules at the grammatical (morphological, lexical, syntactic) level. An algorithm and a model of lexical analysis for Kazakh-Russian and Kazakh-English language combinations and a technology for automated generation of machine translation grammatical rules based on a parallel corpus has been developed.

Unified tagging system (meta-language), morphological and syntactic models of Kazakh and Turkish languages have been developed for creating a Rule-based machine from Kazakh to Turkish [15–18].

The quality of machine translation depends on the quality of the translation of words expressing emotions from one language to another, and not only words, but emotions as well. For emotions translation, we need to do a sentiment analysis in one language and transfer these emotions to another one. In order to do this, both languages need to have a sentiment knowledge base with marked sentiment level. The sentiment analysis or opinion mining in natural languages is one of the most rapidly growing natural language processing technologies. The sentiment analysis is a field of research which analyzes opinions, moods, ratings, assessments, attitudes and emotions of people related to such objects as products, services, organizations, individuals, problems, events, topics and their features [19]. The sentiment analysis has been applied at various levels, starting with the level of the whole text, and then moving to sentences and/or phrases and aspects levels.

A large number of resources and systems for sentiment analyzing in texts have been developed to date for the English language, a large number of resources and systems for analyzing the polarity of texts have been developed to date [19]. A number of researches are currently being implemented to analyze sentiment for the Russian [20, 21], Turkish [22, 23], Spanish [24], Arabic [25] and other languages. An approach for subjectivity

identification in Twitter Microtexts is proposed for the Spanish language [24], which investigates the use of structured information within a social network framework. As for the Arabic language, a semantic approach is proposed for identifying user attitudes and business insights from Arabic social networks by framing ontology. The sentiment text analysis modules, described in the works [26, 27] were implemented for the analysis of Kazakh and Russian texts. In [28] described the sentiment classification approach to measure the machine translation quality and sentiment preservation.

In addition to the above-mentioned main tasks, there are other important tasks, such as creating evaluation dictionaries, emotion recognition, sarcasm detection, etc. Emotion recognition, as is known, is the process of determining human emotions. People differ from each other in terms of their emotions. Emotion recognition technology is a new area of research.

There are interesting researches on emotion recognition from videos based on people’s facial expressions, from audio information by intonation, from texts by the writing style [29, 30]. The [31] conducted a survey on the implementation of emotions in robotics and views for the future in the field of Affective computing.

Researches on the emotions recognition from texts in the Kazakh language is just beginning, before that, there were conducted researches in sentiment analysis of texts. However, simple emotions can already be recognized. As the basis of such recognition, we use sentiment analysis of texts. The sentiment analysis of a text is a group of semantic analysis methods designed for automatic identification words that bring a subtle coloring to the text, and an emotional assessment of opinion regarding an object, event, phenomenon, process or its features written in the text.

In the case with the Kazakh-Turkish machine translation, it is necessary to use the previously created semantic base of lexical units [32–34] with a ratings score in the Kazakh language and automatically generate a semantic database of lexical units with a sentiment rating score in Turkish based on the languages similarity. The obtained database is being checked by the experts for its correctness.

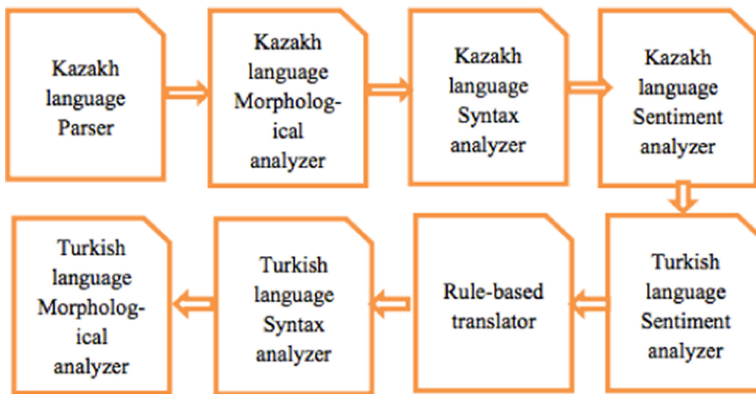


Fig. 1. Kazakh-Turkish machine translation schema

2 Method of Sentiment Analysis in the Kazakh-Turkish Machine Translation

Schema of the Kazakh-Turkish machine translation is shown in Fig. 1.

At the first stage, text processing is done by a special parser, which performs preliminary graphematic processing.

At the second stage the text is processed by morphological analyzer to determine the parts of speech and the features of each part of speech.

The syntax analyzer is engaged at the third stage; the Kazakh sentences structure is being determined.

At the fourth stage the sentiment analyzer is active, which determines the sentiment of the Kazakh text.

The sentiment analyzer for the Turkish language makes a query in the Turkish sentiment knowledge base and compares rating score at the fifth stage, choosing the words closest in terms of ratings and meaning.

The rule-based translator is engaged at the sixth stage. The main idea of such translator is based on the languages similarity. The ontological models for Kazakh and Turkish morphology and syntax are built; the sentiments of the knowledge base are compared.

At the seventh step, the Turkish parser generates the structure of the translated sentence.

The morphological analyzer finally selects the most suitable words in Turkish from the candidates at the eighth stage.

3 Morphological Analyzer

At the stage of morphological analysis, words are processed separately, their basis and changing parts, such as suffix and ending are determined. The grammatical class of words combined with a generalized meaning as grammatical and lexical units, is classified by parts of words.

The morphological analysis ensures definition of normal format of a given form of word and a set of parameters typical for a given type of word.

Up to that time the known morphological analysis methods based on the semantic neuron network [35] and cell automaton, and also algorithm applied jointly with declarative and procedural methods. The morphological analysis algorithm used to solve the tasks set in this work is as follows.

The following algorithm has been used for morphological parsing of words in the Kazakh language:

1. The word is being read.
- 2 The word root is in the words database.
- 3 If the word is found then move to the 11th step, otherwise to the 4th one.
4. Every symbol in the cycle, starting from the word's ending is read and searched in the word endings database.
- 5 If the ending is found, then the search proceeds to the word roots database.

6. Morphological information about word endings is preserved.
- 7 If the remaining part of the word is in the word roots database, then
- 8 If it is a verb, the word from the left is then read and the verb tense is determined.
- 9 If the word is an adjective, its connection is verified with the list of additional words, which may pronounce word before word and lead to its intensifying meaning, if it is within the list then the adjective will be intensifying.
- 10 Return to the 4th step in case if the word was not found, otherwise return to the 11th step.
11. Complete.

The ontological models, built in the Protégé environment for Kazakh and Turkish morphological rules are compared for every part of speech (see Fig. 2, Fig. 3).

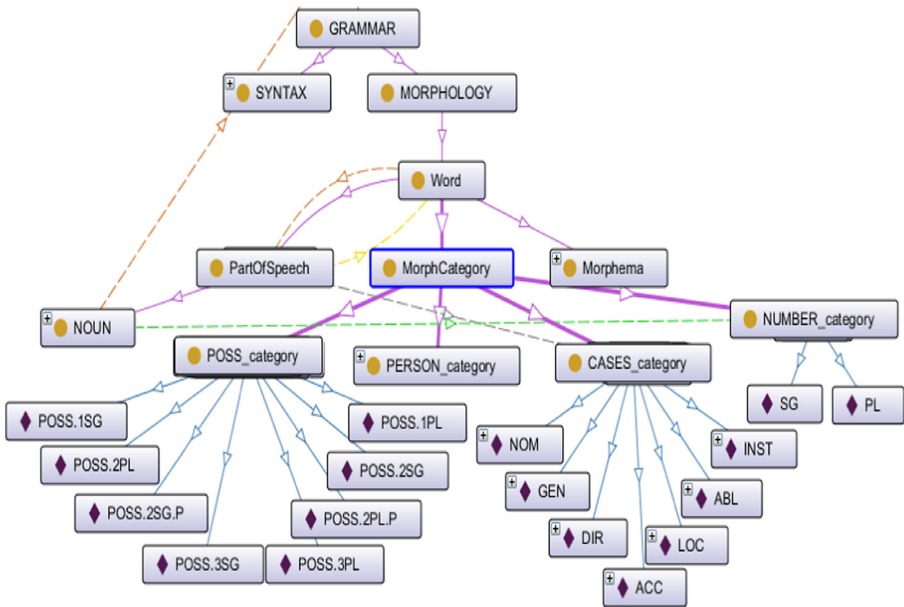


Fig. 2. Ontological model of the noun in the Kazakh language

A comparative analysis of the nouns features showed that out of 54 features 9 of them are present in the Kazakh language, and are absent in the Turkish. Consequently, 83% of nouns features are the same for both languages. Similar results brought comparisons of other parts of speech. This provides possibility to use synchronized knowledge bases, as well as to use rule-based machine translation. The unitary meta-language is used to formulate the rules.

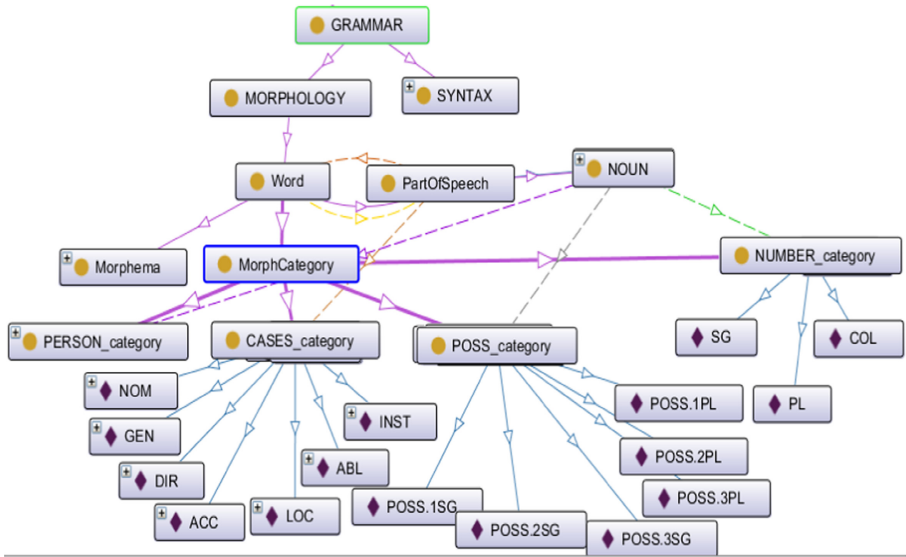


Fig. 3. Ontological model of the noun in the Turkish language

4 Syntax Analyzer

It is known that the formal Context-Free Grammar (CFG) by Chomsky is a math apparatus which allows formal description of syntax in the given language [36, 37]. Currently the CFG is the most widely used formal system for modeling composite structures in human languages.

By using this apparatus, we formalize the syntactic rules of simple narrative sentences in the Turkish language and build their component trees. But for this we need to input special linguistic marks – tags. The parsing model for the Kazakh language is presented in the paper [38]. The tags for sentence elements in the Turkish language will be indicated in the system of the unified meta-language for describing Turkic languages - UniTurk, which are presented in Table 1.

Table 1. Tags for description of the Turkish sentence structure.

Tag	English	Russian	Kazakh	Turkish
S	Simple sentence	Prostoe predlojenie	Jai sóilem	Yalın cümle
Sub	Subject	Podlejashee	Bastaýysh	Özne
Obj	Object	Dopolnenie	Tolyqtaýysh	Tümleç
Obj 1	Object	Dopolnenie		Nesine
Abr	Abbreviation	Opredelenie	Anyqtaýysh	
Adl	Adverbial	Obstoyatel'stvo	Pysyqtaýysh	Zarf tümleci
Pre	Predicate	Skazuemoe	Barandaýysh	Yüklem

The CFG consists of a set of rules or deduction rules, each of them expresses methods by which meta-symbols may be grouped and arranged together. The list of abbreviations and their meanings are given in the Table 1.

The CFG of the common G setup is determined by the following parameters [37]:

$$G = \langle N_s, T_s, R, S \rangle \tag{1}$$

where:

- N_s – a set of non-terminal symbols (variable);
- T_s – a set of terminal symbols (constants): herein $N_s \cap T_s = \emptyset$;
- R – set of rules of the form $A \rightarrow \alpha$, where A – non-terminal symbol, α – a string of symbols from the infinite set of strings $(N_s \cup T_s)$;
- S – initial non-terminal symbol.

The sentence structure may be represented as of two parts: nominal and verbal. The syntax of simple narrative sentences of the Turkish language can be described using a definite CFG.

For instance, let the following simple sentences be introduced:

1. “Samat kitap okuyor” - “Samat is reading a book”;
2. “Samat kütüphanede kitap okuyo” - “Samat is reading a book in a library”;
3. “Samat annesiyle dün geldi” - “Samat came yesterday with his mother”;
4. “En yakın arkadaşım Samat okuyor” - “My best friend Samat is reading”;

In order to describe the structures of these sentences for the CFG system parameters, we assign the following values:

$$N_s = \{S, NP, VP, Adj, Adv\}$$

$$T_s = \{S, a, d, e, g, h, i, k, l, m, n, o, p, r, t, s, s, u, ü, y\}$$

$$R = \{S \rightarrow NP|VP, NP \rightarrow N|N|Adj|Adv, VP \rightarrow N|V|Adv|NP|VP\}$$

The constituency-based parse trees of the above-mentioned sentences of the Turkish language with using rules of this grammar may be represented as in Fig. 4–7:

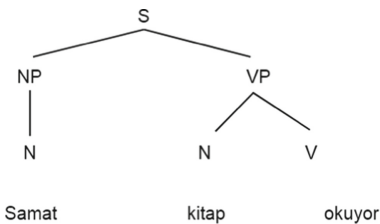


Fig. 4. Parse tree S(NP(N),VP(N,V))

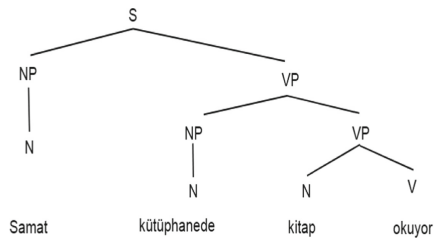


Fig. 5. Parse tree S(NP(N),NP(N,N)),VP(N,V)

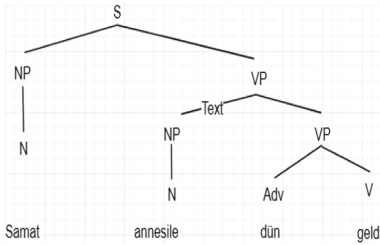


Fig. 6. Parse tree S(NP(N), VP(NP(N)

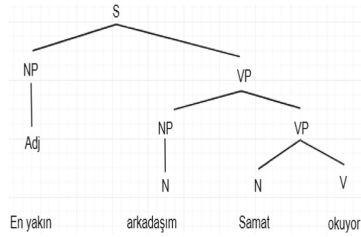


Fig. 7. Parse tree S(NP(Adj), VP(NP(N), VP(N,V)))

In order to formalize the syntactic rules for simple sentences of the Turkish language, we first built trees of components for them using Chomsky’s CFG [36, 37], and then, with consideration of their semantics, we developed ontological models in the Protégé environment, which has a free, open ontology editor and framework for building knowledge bases. The Protégé platform supports two main methods for ontology modeling: through the Protégé-Frames editor and through the Protégé-OWL editor. The ontologies built by Protégé are open ones with an easily extensible architecture due to the support of extension modules and can be exported to many formats, including RDF (RDF Schema), OWL and XML Schema.

The result of the parsing is the syntactic structure of the sentence which is presented either in the form of a tree of dependencies, or shaped as a tree of constituency. The syntax defines the rules used to organize words in a sentence based on the following elements: constituency, grammatical relations, subcategories, and dependency relations [35]. The components are word groups that may behave as a whole or a phrase. The components here are described using parts of speech and phrases. The nominal phrase consists of groups of nouns, pronouns, adjectives, numerals (my house, large room). Grammatical relations are the formalization of the sentence structure as relations between subjects, objects, and other concepts. The components are characterized by their role within the sentence, and parts of the sentence are identified. Kazakh language possesses five parts of sentence: Subject, Predicate, Attribute, Adverbial, Object. For example, ol [SUBJECT] kitapty [OBJECT] aldy [Predicate] (he took the book) [35].

5 Features for Sentiment Analysis of Text in the Kazakh Language

The feature here denotes the sentiment character or attribute of a word or phrase. In case of sentiment analysis of the text, parts of speech (noun, adjective, verb, adverb, interjection), the words of negation speech are used as features. For instance in the Kazakh language the following parts of speech bring sentiment to the text: nouns (crime, war), verb (arrest, rejoice, get angry), adjective (beautiful/attractive, good/bad), adverb (really, very, very), interjection (cheers!, excellent!, wow). The words “not”,

“no” are the words of negation in the Kazakh language. According to the research, the noun is an aspect (object) of discussion, and adjectives mainly determine the semantic orientation (polarity) of the text. The sentiment of emotion word may depend on the context and the object domain.

6 Sentiment Analyzer

The results of morphological and syntactic analysis are used in building a model for sentiment analysis of texts in the Kazakh language. For example, sentiment may be determined with the help of these phrases:

[N] · [V]

[N] · [V] · [Negation]

[ADJ] · [N]

[ADJ] · [Negation] · [N]

[ADJ] · [V]

[ADJ] · [V] · [Negation]

[ADV] · [ADJ]

[ADV] · [N];

here, ADJ – part of speech, the adjective, N- the noun, Negation – negation words “not/no”, V – the verb, ADV – the adverb, · - concatenation .

In order to carry out the sentiment analysis, it is necessary to determine the lexical units of sentiment analysis, which may be words, phrases and sentences in human language. We may calculate the sentiment of the entire text by determining the sentiment features of lexical units.

A production model has been used for modeling the sentiment analysis of texts in the Kazakh language. These models were presented in the papers [32–34].

Dictionary and formal rules based hybrid method has been developed for sentiment analysis of the Kazakh language. The database of lexical units with sentiment ratings in the Kazakh language is used as a dictionary. The database was manually created and tagged by sentiment on a 5-point scale (from –2 to 2). Besides, some words or characters may reverse sentiment rating depending on the context of text. “+ -” symbol is used for such cases. The sentiment of lexical units is measured by the following scores: –2 - very negative; –1 - negative; 0 - neutral; 1 - positive; 2 - very positive. The semantic base comprises not only a list of words and phrases, but also their interpretation and synonyms with sentiment ratings. The database consists of more than 13000 words and phrases built with words related to various parts of speech [33].

The formal rules are used as rules determining the sentiment of texts in the Kazakh language using the production model. Each rule used to determine the sentiment of a text fragment is presented in the form “IF condition, THEN conclusion”.

7 Conclusion

In late years, the influence of posts (messages) in social networks, the Internet is observed on changing of business, changing in the population mood regarding various events.

The data obtained from social networks, forums and micro-blogs is of great interest for researches and applications, due to the possibility of publishing real-time reviews and people's moods on any issues with availability of information in huge volumes and in different languages. Now we can track and analyze reviews about a product or company, identify supporters or opponents of a political party or social movement in various fields, predict financial incomes.

In this relation, the preservation of the text sentiment in machine translation from the Kazakh to Turkish language is an interesting and important task.

In this work, we review the method of text sentiment preservation in machine translation between Kazakh and Turkish.

This paper describes the features that affect sentiment determination of texts in the Kazakh language, models of morphological rules and algorithms for morphological analysis, formal models of simple sentence structures of the Kazakh-Turkish combination, models and methods of sentiment analysis of texts in the Kazakh language. The performed studies comparing the morphological and syntactic rules of the Kazakh and Turkish languages prove the similarity of their structure. In this regard we may assume that consideration of sentiment in machine translation for these language combinations will bring a good result while preserving the meaning of the text.

In the future, it is planned to involve experts to evaluate the texts translation results with emotional colors, the works to increase the amount of semantic base with rating score using the machine translator between these languages are also planned.

Acknowledgments. The work was supported by the grant financing for scientific and technical programs and projects by the Ministry of Science and Education of the Republic of Kazakhstan (Grant No. AP05132249, 2018–2020).

References

1. Kazakh grammar: Phonetics, word formation, morphology, syntax in Kazakh, Astana, Kazakhstan (2002)
2. Lewis, G.: Turkish Grammar, Oxford University Press (2001)
3. Promt, http://www.promt.ru/company/technology/machine_translation/, Accessed 15 Dec 2019
4. Koehn, F.J., Och, M.D.: Statistical phrase-based translation. In: Proceedings of NAACL-HLT, pp. 48–54. Edmonton, Canada (2003)
5. Koehn, H., et al.: Moses: open source toolkit for statistical machine translation. In: Proceedings of the ACL Demo and Poster Sessions, pp. 177–180. Association for Computational Linguistics, Prague (2007)
6. Lagarda, A.L., Alabau, V., Silva, F. R., D'íaz-de-Lianono, E.: Statistical post-editing of a rule-based machine translation system. In: Proceedings of NAACL HLT. Short Papers, Boulder, pp. 217–220. Association for Computational Linguistics, Colorado (2009)

7. Hamzaoglu, I.: Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages. MSc Thesis. İstanbul: Bogazici University (1993)
8. Altıntaş, K.: Turkish to Crimean Tatar Machine Translation System. MSc Thesis, Bilkent University, Ankara (2000)
9. Tantuğ, A.C., Adalı, E., Oflazer, K.: Computer analysis of the turkmen language morphology. In: Salakoski, T., Ginter, F., Pyysalo, S., Pahikkala, T. (eds.) FinTAL 2006. LNCS (LNAI), vol. 4139, pp. 186–193. Springer, Heidelberg (2006). https://doi.org/10.1007/11816508_20
10. Orhun, M., Tantuğ, A. C., Adalı, E.: Rule based analysis of the uyghur nouns. In: Proceedings of the International Conference on Asian Language Processing (IALP), Chiang Mai, Thailand (2008)
11. Abduali, B., Akhmadieva, Z., Zholdybekova, S., Tukeyev, U., Rakhimova, D.: Study of the problem of creating structural transfer rules and lexical selection for the Kazakh-Russian machine translation system on Apertium platform. In: Proceedings of the International Conference Turkic Languages-2015, pp. 5–9. Academy of Sciences of the Republic of Tatarstan Press, Tatarstan (2015)
12. Tukeyev, U., Zhumanov, Z., Rakhimova, D., Kartbayev, A.: Combinational circuits model of kazakh and russian languages morphology. In: Abstracts of International Conference Computational and Informational Technologies in Science, Engineering and Education, pp. 241–242. Al-Farabi KazNU Press, Almaty, Kazakhstan (2015)
13. Salimzyanov, I., Washington, J., Tyers, F.: A free/open-source Kazakh-Tatar machine translation system. Machine Translation Summit XIV (2013)
14. Tyers, F.M., Washington, J.N., Salimzyanov, I., Batalov, R.: A prototype machine translation system for Tatar and Bashkir based on free/open-source components. In: First Workshop on Language Resources and Technologies for Turkic Languages, pp. 11–14 (2012)
15. Bekmanova, G., et al.: A uniform morphological analyzer for the Kazakh and Turkish languages. In: Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts - AIST 2017, pp. 20–30. Moscow, Russia (2017)
16. Yergesh, B., Mukanova, A., Sharipbay, A., Bekmanova, G., Razakhova, B.: Semantic hyper-graph based representation of nouns in the Kazakh language. *Computacion y Sistemas* **18**(3), 627–635 (2014)
17. Yelibayeva, G., Mukanova, A., Sharipbay, A., Zulkhazhav, A., Yergesh, B., Bekmanova, G.: Metalanguage and knowledgebase for kazakh morphology. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11619, pp. 693–706. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_51
18. Zetkenbay, L., Sharipbay, A., Bekmanova, G., Kamanur, U.: Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages. *J. Theor. Appl. Inf. Technol.* **91**(2), 257–263 (2016)
19. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **5**(1), 1–167 (2012)
20. Loukachevitch, N.V., Chetviorkin, I.I.: Evaluating Sentiment Analysis Systems in Russian. *Artificial intelligence and decision-making*, 1, 25–33. Russian (2014)
21. Chetviorkin, I., Loukachevitch, N.: Extraction of russian sentiment lexicon for product meta-domain. In: Proceedings of COLING 2012, pp. 593–610 (2012)
22. Akba, F., Uçan, A., Sezer, E.A., Sever, H.: Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In: Proceedings of the 8th European Conference on Data Mining, pp. 180–184 (2014)

23. Eryiğit, G., Çetin, F., Yanık, M., Temel, T., Çiçekli, I.: TURKSENT: A sentiment annotation tool for social media. In: Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, ACL 2013, Sofia, Bulgaria (2013)
24. Sixto, J., Almeida, A., López-de-Ipiña, D.: An approach to subjectivity detection on twitter using the structured information. In: International Conference on Computational Collective Intelligence. ICCCI 2016, LNCS, vol. 9875. Springer, Cham (2016)
25. Samir, T., Ibrahim, A.-N.: Semantic sentiment analysis in arabic social media. *J. King Saud Univ. Comp. Inf. Sci.* **29**(2), 229–233 (2016)
26. Sakenovich, N.S., Zharmagambetov, A.S.: On one approach of solving sentiment analysis task for kazakh and russian languages using deep learning. In: International Conference on Computational Collective Intelligence. ICCCI 2016. LNCS, vol. 9876. Springer, Cham (2016)
27. Abdullin, Y.B., Ivanov, V.V.: Deep learning model for bilingual sentiment classification of short texts. *Sci. Tech. J. Inf. Technol. Mech. Optics* **17**(1), 129–136 (2017)
28. Lohar, P., Afli, H., Way, A.: Maintaining sentiment polarity in translation of user-generated content. *Prague Bull. Math.Linguist.* **108**, 73–84 (2017)
29. Gervasi, O., Franzoni, V., Riganelli, M., Tasso, S.: Automating facial emotion recognition. *Web. Intelligence.* **17**, 17–27 (2019). <https://doi.org/10.3233/WEB-190397>
30. Majumder, N., et al.: DialogueRNN: an attentive rnn for emotion detection in conversations. In: Proceeding of the AAAI Conference on Artificial Intelligence, 33, pp. 6818–6825. Honolulu (2019)
31. Franzoni, V., Milani, A., Nardi, D., Vallverdu, J.: Emotional machines: The next revolution. *Web Intell.* **17**, 1–7 (2019). <https://doi.org/10.3233/WEB-190395>
32. Yergesh, B., Bekmanova, G., Sharipbay, A.: Sentiment analysis of Kazakh text and their polarity. *Web Intell.* **17**(1), 9–15 (2019)
33. Yergesh, B., Bekmanova, G., Sharipbay, A., Yergesh, M.: Ontology-Based Sentiment Analysis of Kazakh Sentences. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10406, pp. 669–677. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_47
34. Yergesh, B., Bekmanova, G., Sharipbay, A.: Sentiment analysis of Kazakh phrases based on morphological rules. *J. Theor. Appl. Sci. Tech.* **2**(38), 39–42 (2016)
35. Sharipbayev, A., Bekmanova, G., Buribayeva, A., Yergesh, B., et al.: Semantic neural network model of morphological rules of the agglutinative languages. Proceeding of the SCIS/ISIS 2012, pp. 1094–1099. Kobe, Japan (2012)
36. Jurafsky, D., Martin, J.H.: *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed.). Prentice Hall PTR, Upper Saddle River, NJ, USA (2009)
37. Chomsky, N.: *Syntactic Structures*. The Hague: Mouton, 1957. (Reissue: Chomsky N. *Syntactic Structures*. – De Gruyter Mouton) (2002)
38. Sharipbay, A., Razakhova, B., Mukanova, A., Yergesh, B., Yelibayeva, G.: Syntax parsing model of Kazakh simple sentences. In: proceedings of the Second International Conference on Data Science, E-Learning and Information Systems DATA 2019, Article 54, p. 5. Dubai (2019)