Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ALGORITHM FOR SOLVING PRONOMINAL ANAPHORA IN THE KAZAKH LANGUAGE

GULZHAMAL KALMAN[1], YERZHAN ZHUMABAY[2], ELMIRA NURGALIEVA[3], ASSEL KUANYSHEVA[4], MUSATAY ESMAGANBET[5]

[1,3,4]L. N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

[2] Astana International University, Astana, Republic of Kazakhstan

[5]A. Myrzakhmetov Kokshetau University, Kokshetau, Republic of Kazakhstan

E-mail: [1]guljamal14@gmail.com, [2]erzhan_93kz@list.ru, [3]Elmira.nurgalieva.1992@mail.ru, [4]Asel.kuanysheva@yandex.ru, [5]esmaganbet_m@mail.ru

ABSTRACT

In this article, we will present methods for solving pronominal anaphora in the Kazakh language. The goal is to identify and solve the difficulties in solving anaphoric relations in the Kazakh language. in the course of our research, we identify some issues related to anaphora annotation specific to the Kazakh language, such as the general position of the candidate antecedent, the distance between words, etc. in the study, we use an annotated text corpus and determine the types and number of pronouns in the text with the help of a morphological analyzer and a syntactic analyzer.

Methods used during the research The first method uses Support Vector Machine as training and classification algorithms, the second method uses Decision Tree Inductor. we use an annotated corpus based on machine learning for these methods.

We tested the anaphora solving methods with the anaphora solving system for the Kazakh language. The results are encouraging.

Keywords: *Anaphora, Support Vector, Kazakh Language, Antecedent, Kazakh Pronouns.*

1. INTRODUCTION

One of the urgent tasks facing computer linguistics is to extract information about various objects in textual documents: people, organizations, events, places, etc., as well as the relationships between them. Each information object (entity) corresponds to a certain concept/relationship of the subject domain and has a certain structure. This issue in natural language processing will be related to the notion of referentiality. If the referring expression in the sentence refers to the expression mentioned before it, then an anaphoric relation (anaphora) is established between them, the last of the expressions is called an anaphora, and the preceding one is called an antecedent.

Solving anaphora is one of the major problems of natural language processing. Resolving referential relations in natural language is important for both deep natural language study, language machine learning, and language learners. A lot of researches have been conducted for solving the anaphoric relationship for other languages, for example: English, Czech languages. German, Arabic, Turkish, Russian.etc

The problem of solving the anaphora for the Kazakh language has just begun, [1] the first research work on solving the reference in the Kazakh language, where the authors [4] tried to integrate it into the proposed method [2,3]and proposed a model for solving reference relations in a multilingual system.

In this paper, we conduct the following research in solving the anaphora of pronouns in the Kazakh language. First: research on the basis of machine learning using the annotated case.

Second: The impact of semantic roles and their features on solving anaphora in the Kazakh language is considered.

We solve the pronoun anaphora only by comparing statistical and inductive methods of the solution tree based on a vector machine for the classification noun reference noun and proper nouns (see Table 1 for the complete set of pronouns treated). As a training and testing database, we used annotated corpora: National Corpus of Kazakh language (NKL).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: *Types Of Anaphora Treated.*

Pronominal Anaphora in English			Pronominal Anaphora in Kazakh		
Personal	Demonstrative	Reflexive	Personal	Demonstrative	Reflexive
He She It His Her Him Its They Them Their	This That These Those Others	Himself Herself Itself themselves	Ол Олар	Ана Анау Әне Бұл Мына Міне Осы Осынау Сол Сонау	Өзім Өзің Өзіңіз Өзі Өздері Өзіміз

2. THE KAZAKH LANGUAGE

The Kazakh language belongs to the Kipchak group of Turkic languages, including the Kypchak-Nogai branch, together with the Karakalpak, Nogai, and Nogai languages. In addition, it is close to the Kyrgyz, Tatar, Bashkir, Novy-Balkar, Kumyk, Karaim, and Crimean languages. More than 13 million people speak Kazakh in the world.

As for the grammatical features of the Kazakh language:

i) The Kazakh language has nine-word groups, one of the most used is pronouns.

ii) There is no gender category in the Kazakh language

he, she, it – “ол”

iii) sentence order is mostly stable,

iiii) In the Kazakh language, not all pronouns have an anaphoric function. The most common ones, the personal pronouns, demonstrative pronouns and reflexive pronouns. (See Table 1 for the complete set of pronouns treated)

2.1 National corpus of Kazakh language (nkl)

The Corpus site contains an electronic text fund of the Kazakh language. The volume of text in the corpus is 21 million. The texts are collected from 5 styles of the Kazakh language (artistic style, scientific style, journalistic style, paper style, speech style). You can search the corpus by word, word form (word transformation) and see the list of sentences where the word you are looking for is used and their source. Any found word/word form or any word in the examples is provided with information relevant to all levels of the language. https://qazcorpus.kz/

The term "notation" is used in the corpus. The label displays detailed information about the word you searched for in the corpus. There are 5 types of notations in the corpus: morphological, word-formative, semantic, lexical, phonetic-phonological.

Morphological segmentation of Kazakh words [26] Here we propose to use the theory about the system of affixes of the Kazakh language to describe the process of morphological analysis. This method, developed by me in collaboration with U.A. Tukeev, can serve as a special case for describing the morphological model of the Kazakh language based on the theory of the mapping method. Using the method of multivalued mappings allows us to improve the efficiency and significantly increase the performance of the segmentation system.

This method is also based on the works of Professor K.Bektaev[27], who identified more than 700 endings of the Kazakh language. our morphological analyzer gives morpheme-by-morpheme segmentation along with their respective meanings.

Semantic and pos tagsets in the corpus. Kazakh is an agglutinative Turkic language, in which word forms are generated by means of the affix inflection. When different affixes are attached to the root word, the meaning of the word changes, so distinguishing between pos and semantic tags makes it easier for us to solve this anaphoric relationship. about the attachment of affixes in the Kazakh language, the main linguistic properties defined in the Kazakh grammar, their codes, and main signs, that is, several values they take, are shown in Table 2 and the following examples. [28] the authors described it in detail in the following works.

Table 2: Main Linguistic Properties Defined In Kazakh Grammar.

№	Linguistic property	Code	Cardinality
1	Animacy	A	2
2	Number	N	2
3	possessiveness	S	10
4	Person	P	8
5	Case	C	7
6	Negation	N	2
7	Tense	T	3
8	Mood	M	4
9	Voice	V	5

An example related to the attachment of affixes.

In this example personal pronoun "I- мен" and preposition "to- ке" are "hidden" in the affixes of case and person, i.e.:
Банк(noun= a bank)+ке(dative case= to bank)
Бар (verb, imperative=go)+ды(affix ,past tense=he/she went)+м(affix, 1st person=I went).

As the example shows, the combination of affixes affects the pronouns in the word.

3. LITERATURE REVIEW

Numerous studies have been conducted on the solution of pronoun anaphora. Important works include Hobbes [5,6] [7] [8], Mitkov [9,10,11],Tetreo [14] and Trouille [12,13] applies. If we look at recent research,

In the next work Yu-Hsiang Lin and Tyne Liang [15] a method was used to solve the pronoun anaphora using UMLS ontology and SA / AO (subject-action / action-object) models, where an F-value of 92% was obtained.

(Saoussen Mathlouthi Bouzid et al 2021),[16] they proposed a SVM method of self-learning based on a set of patterns and linguistic criteria to solve an ellipse with a pronoun in the Arabic language, For the solution step, they developed a new hybrid method that combines an enhanced learning method with Word implementation templates. The learning enhancement method used an adapted version of the Q-learning algorithm to find the optimal combination of capabilities. It uses a set of morphological and syntactic features. The Word-based approach uses word display templates to test the semantic validity of candidates. Evaluation of the identification method gives an accuracy of 99.23% for pronominal anaphores and 94.33% for ellipses.
In the next study, [17] Support Vector Machine (SVM) was the most effective of several methods for solving the pronoun anaphora using the machine learning method for the first time for the Nepali language.

[18] In this work the solution of the pronoun anaphora in the Turkish language is provided, the lgorithms used here are: J48, Voted Perceptron, SVM (support vector machine), Naive Bayes and k-nearest neighbours.
[19] In the next study describes a quantitative analysis performed to compare two different methods on the task of pronoun resolution for Swedish. they are Mitkov's algorithm and SVM (support vector machine), A a result of SVM-based methods significantly outperformed the implementation of Mitkov's algorithm. [20] They are Using the SVM (support vector machine) method, solved the pronoun anaphora in the Holy Quran, based on the evaluation results, the system can find the front row of an anaphor with the best accuracy value of 88.08%.

[21] In this paper, the authors proposed a learning approach for solving noun phrases in unconstrained text using an annotated corpus, and the task involves solving general noun phrases, not just pronouns.
As a result of the study, they got a good result that encouraged them.
[22] the paper presents a joint feature selection and ensemble learning model to solve anaphora in Indian language. They used Conditional Random Field (CRF) and Support Vector Machine (SVM) as machine learner. The result was better than the current models, and they predicted that future work would be entropy and support vector machine.
[23,24] In the following work, the author explained in detail the machine learning methods for solving Russian pronoun anaphora and got a satisfactory result.

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[25] In the research work, the solution of the reference relationship of clinical narratives is described. The method used here is the vector machine and decision tree methods. as a result, vector machine learning outperformed the traditional decision tree baseline.

4. MATERIALS AND METHODS

The following methods were used during the writing of the article: formal analysis method, classification algorithms, Support Vector Machine and decision tree method.

Fig 1 shows the overview of the Kazakh anaphora resolution processes.

In the training data, we used the annotated corpus, because: automatically performs pos tagging, sentence segmentation, deep morphological analysis,etc

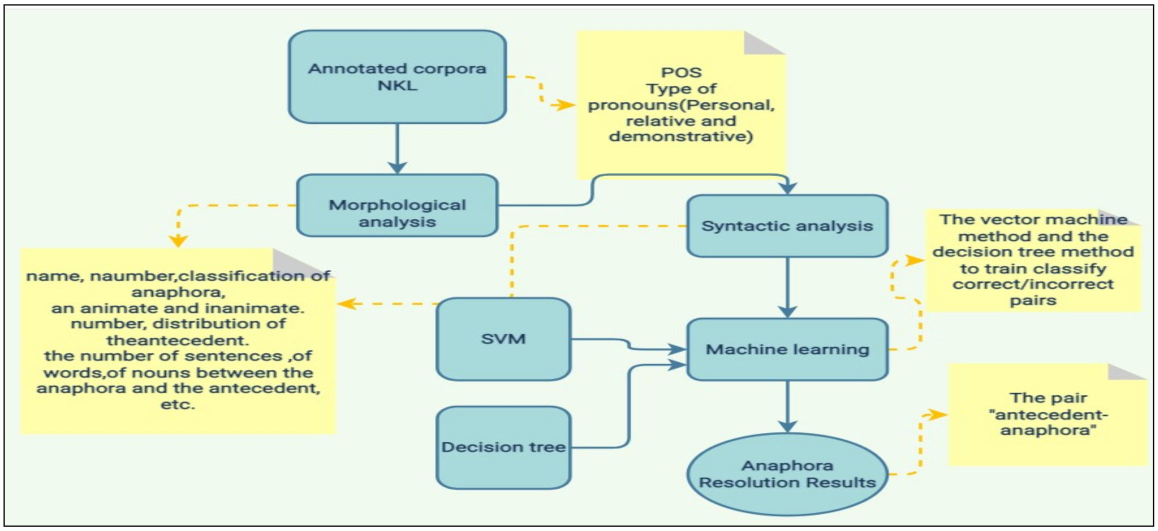
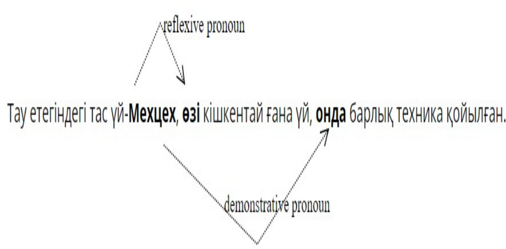


Figure 1: The Kazakh Anaphora Resolution Processes

4.1 THE PRONOUN ANAPHORA

The most common type of anaphora is the pronoun anaphora. The most common of the pronouns being the personal pronouns, demonstrative pronouns, and reflexive pronouns. From the following examples we can see the anaphoric function of the reference demonstrative noun.

For example:



This syntactically complex unit that connects the first sentence and the second sentence is an anaphoric relationship, where the word **мехцех** in the first sentence is repeated with pronoun **онда** in the third sentence. Where the word **Өзі** reflexive pronouns.

For example:

In the example of the first sentence it is anaphoric to have the Personal pronoun **Елжас** and **бразилияда** in the second sentence the word ***ол*** and **сол елден** is repeated through it in third person form and repeated with the same demonstrative pronoun.

4.2 SOLVING THE PRONOUN ANAPHORA

Solving anaphora - the problem of determining the correct pair of "antecedent-anaphora" in our

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

study we consider only the classificatory, reference, personal pronouns.
The antecedent must match the anaphora in number and spelling. The distance between the anaphora and the antecedent in a word must not exceed a given value, depending on the text. In order to solve the anaphoric relationship, we classify the anaphor and the antecedent according to the following features.

Morphological and syntactic features:
1) name, number, classification of anaphora
2) number, distribution of the antecedent;
3) juxtaposition of an anaphora and an antecedent between an animate and inanimate (if there are nouns);
4) the number of sentences between the anaphora and the antecedent;
5) the number of words between the anaphora and the antecedent;
6) the number of nouns between the anaphora and the antecedent;
Semantic features:
7) semantic roles of anaphora;
8) semantic roles of the antecedent;
In the morphological analysis anaphora corresponds to the antecedent, i.e. the correspondence of number, surname, classification, animate and inanimate (if nouns) in classifications 1-3, signs in classifications 4-6 give information about the scale of distance between anaphora and antecedent.
Spacing in words. For each candidate the distance to the pronoun in words is calculated. Depending on this distance the vector is filled in units. There are three gaps to attach them to the vector:

- count from 10 words; vector [1, 0, 0];
- 10 to 30 words; vector [0, 1, 0];
- more than 30 words; vector [0, 0, 1].

Only one vector with a description in vector form can correspond to a candidate.

4.3 ALGORITHM OF CREATING A TRAINING DATABASE

1) Find the pair "antecedent-anaphora" in the set of texts.
2) Find all nouns and nouns between the anaphora and the antecedent.
3) All nouns and pronouns found in step 2 are incorrect hypothetical antecedents.
4) If the correct antecedent is not in the search area, it will not be included in the textbook.
5) Follow steps 1-4 for each processed example.

4.4 ANAPHORA SOLUTION ALGORITHM

1. Find the first anaphora without an antecedent. If no anaphora is found, the algorithm terminates.
2. Find all nouns or pronouns that are anaphors between the anaphora and the antecedent. The anaphora must match their number and quantity. The search area is limited to a predetermined number of words.
3. Add them to the set of hypothetical antecedents.
4. Match the semantic roles of the antecedent to each noun in the set of hypothetical antecedents.
5. Calculate the probability that each hypothetical antecedent will be the correct antecedent using the classification method.
6. Choose the most likely antecedent and connect it to the appropriate anaphora. Skip to step 1.
The search area for the hypothetical antecedent is limited to step 2, since the anaphora usually represents the nearest hypothetical antecedent. This value was calculated in our experiments.

The figure 2 shows a block diagram of finding an “antecedent-anaphora”pair.

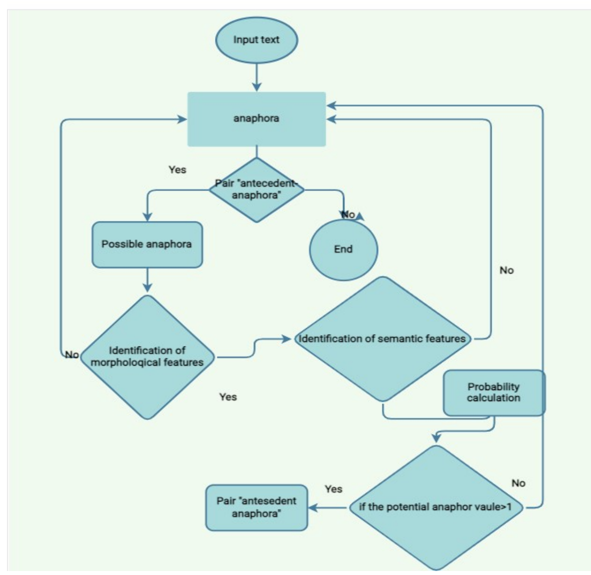


Figure 2: Block Diagram Of Finding An Antecedent-Anaphora Pair.

4.5 SUPPORT VECTOR MACHINE

In this study, we used the method of data classification by Support Vector Machine (SVM). Such classification has a wide range of applications: from pattern recognition or creating spam filters to calculating the

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

distribution of hot aluminum particles in rocket emissions.

The task of classification is to find a classification line where the SVM algorithm correctly distinguishes positive instances from negative instances in an anaphoric relation, which is called the optimal classification line.

The main task in training SVM is to solve the following quadratic optimization problem:

y ∈{1,−1}, i∈ X i | i, =1,2…,

$$min_a f(a) = \frac{1}{2} a^T Q a - e^T a \quad (1)$$

$$0 \le a_i \le C, i = 1, \dots i, y^T\ a = 0,$$

Where e- is a vector of all units; c-is upper limit of all variables; Q -is a symmetric matrix.

4.6 DECISION TREES

We used the decision tree method for the possibilities of correctly finding the anaphor antecedent pair.

- The advantages of the decision tree allow us to see why this method is the most flexible when it comes to choosing solutions.
- Easy to use.
- Several people can work on the model at once, which simplifies the task.

Anaphoric expressions that can have an antecedent are divided into three classes: anaphors or reflexives - reflexive pronouns and reciprocal pronouns; pronominal's - personal pronouns; referential expressions are names that directly name an object in the surrounding world.

The structure of a sentence in a generative grammar is usually depicted schematically as a "tree of components", which illustrates the relationship between elements. On Fig. 3 shows a diagram of a simple sentence.

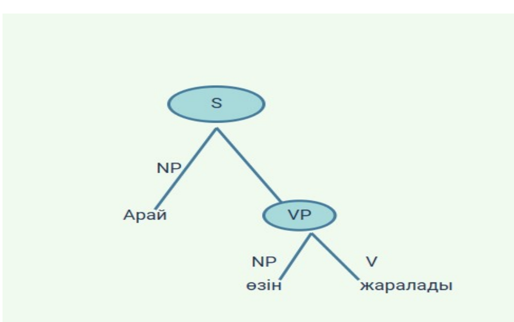


Figure 3: Decision Trees: Simple Sentence Scheme

The subject group (Арай) and the predicate group (өзін жаралады) are on the same level of the constituent tree and are called structural sisters. Through the concept of structural sisterhood, one of the possible relations between the elements of a syntactic group is defined, the relation of SI-command (constituent-command, lit. component control), which is central to the theory of binding (2).

SI -command: The expression α C-commands the expression β if α is the structural sister of the expression containing β.

In the example in Fig. 3 noun phrase (NP - noun phrase) *Арай* si-commands his "structural sister" - the verb group (VP - verb phrase) *өзін жаралады*, and vice versa. At the same time, Arai si-commands each of the components of the verb group, including the reflexive pronoun itself. In the same time,

the pronoun itself does not si-command the noun phrase *Арай*, since it is only included in the sister category, but in itself is not a "sister" of the noun phrase.

Simply put, the pronoun itself is not high enough on the constituent tree to si-command the noun phrase *Арай*. A schematic representation of the SI -command relationship is shown in Fig. 4.

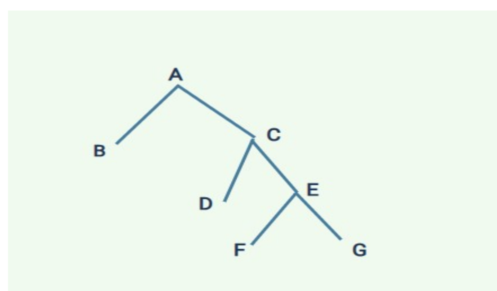


Figure 4: SI-Command Relationship Diagram. A-Not Si Commands Any Category, Because A Dominates All Other Nodes.

B-si commands C,D,E,F,G, but not over A.
C-si commands B.
D-si in commands E,F,G
E-si commands only D
F and P-si command only together

In the original version, formulated by N. Chomsky in the book "Lectures on control and binding" [3], the theory of binding was based on three basic principles (3)

(3) Standard binding theory:

Principle A: the reflexive is bound in its governing category.

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Principle B: the pronominal is free (not bound) in its governing category.
Principle C: referential expression is free (not bound)
Governing category: β is a governing category for α if and only if β is, first, the minimum category, containing α; second, it controls α; and third, is a subject suitable for α
Binding: α binds β if α c-commands β, and α and β co-indexed (have the same index, i.e. an anaphoric dependence is established between them), while co-indexing is possible only if there are no differences in grammatical characteristics.
We will consider examples.
Арай өзін жаралады. (Arai injured himself) (1)
Марал айтты, Арай өзін жаралады. (Maral said that Aray would injure himself) (2)

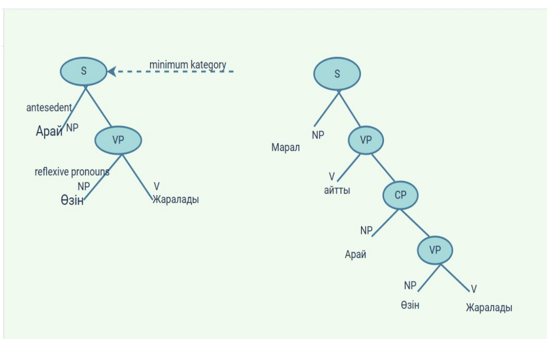


Figure 5. The Illustration Of The Parse Tree Of Sentence (1,2) Andthe Algorithm Working On It

Example (1) is grammatical because it satisfies the requirements of Principle A: the pronoun itself and its antecedent Masha are in the same minimum category, and Masha si-commands herself, which is clearly demonstrated in the component tree in Fig. 4 (left diagram).
The interpretation offered in example (2), on the contrary, is not grammatical, since (see the component tree diagram in Fig. 4 on the right) herself and Masha are not in the same minimum category, which means that the application of Principle A is impossible.

5 CORPUS ANALYSIS AND RESULTS

Quantitative data about pronouns in the corpus are shown in Table 3. Due to our research objective, we only presented information about three types of pronouns in the corpus, and there are some cataphoric relations and anaphoric relations in the corpus, which are 0.25% cataphoric and 12% non-anaphoric respectively.

Table 3: Quantitative Measure From The NKL Corpus

Measure	Count	%
Personal pronouns	23,563	69.7%
Relative pronouns	3753	14.20%
Demonstrative pronouns	1116	3.84%
Total	28432	
non-anaphoric	2101	12%
Cataphors	45	0.26%
Total	28127	86%

Table 4 shows statistical data about pronouns in the NKL corpus

Table 4: Statistics Of The Various Types Of Pronouns In The NKL Corpus

Pronoun types	Count	%
Person		
1st person	4560	17.85%
2nd person	5645	24.72%
3rd person	12766	60.38%
Gender		
Masculine	13876	56.62%
Feminine	2533	10.62%
Others	3145	14.76%
Number		
Singular	7045	22.48%
Dual	458	3.51%
Plural	13424	63.01%

Table 5 shows anaphora and antecedent distances between lines, sentences, words, segments. According to the results of the research, 52.6% of anaphoras were found between the preceding words, 17.5% of anaphoras in the next clause, and only 1% of the anaphora pair matched in the last clause.
Looking at this result, we realized that the distance influences the "antecedent anaphor" pair.

Table 5: Distances Between Anaphor And Antecedent

Distance	Occurrence		
	Verse	Word	Segment
0	6793	120	63
1 to 9	5725	6855	5327
10 to 19	378	2506	2356
20 to 29	56	1302	989
30 to 49	18	1023	678
50 to 100	8	958	549
100 to 200	0	235	456
200 to the end (1060)	0	154	312

during the testing of the reading data, the detectors for finding the anaphor-antecedent pair are shown in the table below.

Table 6: "Antecedent Anaphor" Pair For Different Methods

Feature set	SVM	REPTree
The first set	0.76	0.76
The second set	0.64	0.58

6 DISCUSSION OF THE RESULTS OF THE STUDY OF THE ALGORITHM FOR SOLVING THE PRONOUN ANAPHORA BASED ON MACHINE LEARNING

During the study, all types of Pronominal anaphora relations were analyzed and a theoretical study of methods for their solution was carried out. Works [4–25] consider the referential choice only between full name groups and anaphoric pro-nouns, in addition, studies were conducted Pronominal anaphora relations (the English, Russian languages).

In our study, we used the vector machine(SVM) REPtree method [Chang and Lin, 2014] and the
decision tree method [Waikato University, 2014] to train and classify correct/incorrect pairs.
This makes it possible to design a solution of pronominal anaphora research in the field of computational linguistics and automatic text processing.
The peculiarity of the Algorithm for solving pronominal anaphoric communication in the Kazakh language we calculated the optimal distance for each data set covering 90%. An experimental study of the resolution of Pronominal anaphora relations was carried out. As part of the research work, we analyzed the texts of more than 1000 different topics from the corpora, in the quantitative evaluation of the results of the study we obtained the following data. In the first set of texts 170 texts were examined, 367 pairs of "antident-anaphora" were found, in the second set of texts 200 texts were examined, 467 pairs of "antident-anaphora" were found.

7 CONCLUSION AND FUTURE WORK

The results of the research showed that the method used to solve pronoun anaphora in the Kazakh language was used for the Personal pronouns 81%, Relative pronouns 79%, Demonstrative pronouns 78.3%. these indicators showed that machine learning method and decision tree methods were effective for the Kazakh language.

Thus, the methods used in the course of the experiments showed effectiveness in solving the Kazakh anaphora. as a future work, we will improve the way we define all kinds of pronoun anaphora and hypothetical antecedents.

ACKNOWLEDGEMENT

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP14972834).

REFERENCES:

[1] Zhumabay, Y., Kalman, G., Sambetbayeva, M., Yerimbetova, A., Ayapbergenova, A.., & Bizhanova, A. Building a model for resolving referential relations in a multilingual system. Eastern-European Journal of Enterprise Technologies, (2022). 2(2 (116), 27–35.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] Garanina, N. O., Sidorova, E. A., Seryi, A. S. (2018). Multiagent Approach to Coreference Resolution Based on the Multifactor Similarity in Ontology Population. Programming and Computer Software, 44 (1), 23–34. doi: https://doi.org/10.1134/s0361768818010036
[3] Sidorova, E. A., Garanina, N. O., Kononenko, I. S. (2018). Mnogomestnye ontologicheskie otnosheniya v zadache razresheniya koreferentsii. Shestnadtsataya natsional'naya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2018. 279-287
[4] Kibrik, A. A. (1999). Reference and Working Memory. Current Issues in Linguistic Theory, pp 129. doi: https://doi.org/10.1075/cilt.176.04kib
[5] Hobbs JR (1978) Resolving pronoun references. Lingua 44: 339-352.
[6] Hobbs JR (1976) Pronoun resolution. Technical Report, Department of Computer Science, City College, City University of New York 76-1. 339-352
[7] Lappin S, Leass HJ (1994) An Algorithm for Pronominal Anaphoric Resolution. Computational linguistics 20: 535-561
[8] Kennedy C, Boguarev B (1996) Anaphora for everyone: pronominal anaphora resolution without a parser. 16th International Conference on Computational Linguistics (COLING'96) Denmark, 113-118.
[9] Mitkov R (1998) Robust Pronoun Resolution with Limited Knowledge. 18th International Conference on Computational Linguistics, Canada. 869–875
[10] Mitkov R (2001) Outstanding Issues in Anaphora Resolution. Lecture Notes in Computer Science 2004: 110-125.
[11] Mitkov R (2007) Anaphora resolution: to what extent does it help NLP applications? Anaphora: Analysis, Algorithms and Applications. SpringerVerlag, Berlin Heidelberg, 179-190.
[12] Tetreault J, Allen J (2003) An Empirical Evaluation of Pronoun Resolution and Clausal Structure. 2003 International Symposium on Reference Resolution. 119–126
[13] Tetreault J, Allen J (2004) Dialogue Structure and Pronoun Resolution. Lecture Notes in Computer Science 1793: 515-525.
[14] Trouilleux F (2002) Insertions et interprétations des expressions pronominales. Actes de l'Atelier Chaînes de référence et résolveurs d'anaphores. TALN 2002 Nancy.
[15] Yu-Hsiang Lin and Tyne Liang. 2004. Pronominal and Sortal Anaphora Resolution for Biomedical Literature. In Proceedings of the 16th Conference on Computational Linguistics and Speech Processing, pages 101–109, Taipei, Taiwan. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).
[16] Saoussen Mathlouthi Bouzid, Chiraz Ben Othmane Zribi.2021,Efficient Learning Approach for Pronominal Anaphora and Ellipsis Identification and Resolution in Arabic Texts IEEE/ACM Transactions on Audio, Speech, and Language Processing ,pp 3335 - 3348
[17] Senapati, A., Poudyal, A., Adhikary, P., Kaushar, S., Mahajan, A., & Nath Saha, B. (2020). A Machine Learning Approach to Anaphora Resolution in Nepali Language. 2020 International Conference on Computational Performance Evaluation (ComPE). pp 436-441
[18] H. Abaci, M. Eminagaoglu, I. Gor, and Y. Kılıcaslan, "Pronoun Resolution for Tweets in Turkish", Int J Intell Syst Appl Eng, vol. 10, no. 2, pp. 207–215, May 2022.
[19] C Ahlenius. 2020. Automatic Pronoun Resolution for Swedish. - 2020 - diva-portal.org
[20] Ardiyani, A. D., Bijaksana, M. A., & Huda, A. F. (2020). Anaphora Resolution on Al-Quran with Indonesian Translation. Indonesia Journal on Computing (Indo-JC), 5(2), 99-106.
[21] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A Machine Learning Approach to Coreference Resolution of Noun Phrases. Computational Linguistics, 27(4):521–544.
[22] Kumar Sikdar, U., Ekbal, A., & Saha, S. (2016). A generalized framework for anaphora resolution in Indian languages. Knowledge-Based Systems, 109, 147–159. doi: 10.1016/j.knosys.2016.06.033
[23] Tolpegin P. V. (2006), The new methods and algorithms of automated third person pronominal reference resolution of Russian texts, Kom-kniga, p-88.

[24] Tolpegin P. V., Vetrov D. P., Kropotov D. A. (2006), Automated third person anaphora resolution algorithm based on machine learning methods. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2006" pp. 504–507.

[25] Zheng, J., Chapman, W. W., Miller, T. A., Lin, C., Crowley, R. S., & Savova, G. K. (2012). A system for coreference resolution for the clinical narrative. Journal of the American Medical Informatics Association, 19(4), 660–667. doi:10.1136/amiajnl-2011-000599

[26] Gulila Altenbek and WANG Xiao-long. Kazakh segmentation system of inflectional affixes. In Joint Conference on Chinese Language Processing, 2010pages 183–190. CIPSSIGHAN.

[27] K.Bektaev, large Kazakh-Russian Russian-Kazakh dictionary - Altyn Kazna, Almaty, 1999 pp 70.

[28] Olzhas Makhambetov, Aibek Makazhanov, Zhandos Yessenbayev, Bakhyt Matkarimov, Islam Sabyrgaliyev, and Anuar Sharafudinov. 2013. Assembling the Kazakh Language Corpus. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1022–1031, Seattle, Washington, USA. Association for Computational Linguistics.

[29] Chomsky, N. Lectures on Government and Binding. / N. Chomsky — Dordrecht: Foris Publications, 1981. — 371 c.