

# Comparative analysis and definition of types of reference resolution

guljamal kalman (✉ [guljamal14@gmail.com](mailto:guljamal14@gmail.com))

Eurasian State University L N Gumilev: L N Gumilyov Eurasian National University <https://orcid.org/0000-0001-8863-9447>

---

## Research Article

**Keywords:** Anaphora, cataphora, co-reference, reference resolution, Pronoun resolution

**Posted Date:** June 3rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1638743/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

This article describes new methods of solving reference resolution. The main method used here is, first of all, the method of solving by reference types, carrying out a theoretical study of methods for solving reference relations. Second, by a formal analysis of all types of references in Kazakh and Russian, English, a comparative study of all types of references in the three languages is conducted. For the Kazakh language, the absence of certain types and the definition of anaphoric and cataphoric relations in the language, taking into account the linguistic features.

The purpose of solving reference resolution in this method is to highlight previously undeveloped and fully solved methods, and to search the literature for methods of solving reference types and convenient ways of solving reference resolution using methods of theoretical investigation, formal analysis, and verification of developed research papers in the field.

A comparison of the eight methods of solving reference resolution among themselves in this study with the new method of solution is shown in Table 7. And by relating the nine reference types to the eight methods that conducted this theoretical research work, we can see that the complete solution of the reference types can only be solved in our method.

The article studies the function of pronouns in the word and taking into account the linguistic features of Kazakh Russian and English, from which we can extract specific types of reference types by eliminating interlingual differences.

## 1. Introduction

The reference resolution in text analysis is one of the most difficult tasks of computer linguistics, the solution of which is necessary in areas such as machine translation, question and answer systems, automatic text abstracting, information retrieval. the essence of the referential phenomenon is to compare language expressions, objects or real world events presented in a linear text or dialogue.

Having considered in this paper a new method of solving reference resolution, we identified all types of reference types of Kazakh and Russian English and discovered the possibility of complete resolution of reference relations.

These possibilities will be a new starting point for the Kazakh language to address reference resolution. Proceeding from the possibilities of computer technology, this direction in the developing linguistics is of great importance for a broad study and analysis at different levels of linguistic patrimony as a bright mirror of the spiritual and material culture of the people.

The solution of reference resolutions, as for the Russian language, still needs a large number of studies, the study of which will be the beginning of research for the Russian language through work.

The results of multilingual research allow us to see how much research is developed between languages, explaining the differences between languages, comparing the research focus of natural language and the work done in other languages.

## 2. Literature Review And Problem Statement

The etymology of the term anaphora "anaphora" comes from ancient Greek, "anaphora" (αναφορά) is a complex word consisting of separate words ana (back) and pheri (lift). In computational linguistics, anaphora has been studied since the 1960s, and various definitions have been given. Halliday and Hasan (1976) gave a classical definition based on the concept of fusion: anaphora is defined as a unit indicating the previous item (element) .

Anaphora was studied by Graham Hurst in the 80-90s, he also wrote in his book "Anaphora in natural language understanding" (Hirst 1981), which gave the following definition to anaphora. An anaphora is an instrument of reference to a certain object (or objects) in discourse, an anaphora of reference, which is the referent or ANTECEDENT of the object (or object) to which it refers. In Mitkov's stud "Anaphora Resolution", an anaphora is defined as a reference "back" or described as a reference to elements mentioned earlier in the discourse.

In this process, which is one of the most difficult tasks of computational linguistics and has not yet been fully solved, many researchers have presented their research. Let's look at the latest research papers.

The solution of the anaphora in Bangla and India was solved with an accuracy of 63.6%. In their research, they used different ways of presenting information in the news on the same topic. Automatic words were created for an anaphora capable of denoting all nouns and pronouns by five different criteria (number, person, status, gender and POS) [1].

In recent years, machine learning methods have been widely used to differentiate anaphora and have achieved significant success. In [2], a noun phrasal system was proposed that expands the work of Soon (2001) et al. based on machine learning, MUC6 and MUC-7 coreferences give the best results to date in the resolution datasets-they reached values of 70.4 and 63.4 respectively.

As a result of in-depth training in solving the referential resolution problem [Zhao et al. (2019) a] Pronoun Resolution, the Entity CR subtask was clearly defined and the most effective method for solving deviations in CR solutions was proposed.

Note-In-depth training for the main rating models (Deep Reinforcement Learning for Mention-Ranking Coreference Models) was used in Chinese and English texts, and they used a reward rescaled max-margin objective in English 65.73 in Chinese 63.88{4}

Solving refrain relations for the Russian language still requires a large amount of research, [5] describes a multivariate approach to solving correlation in the process of extracting information from natural language texts based on the complement of ontology, and [6]. With the automatic analysis of the text and the addition of the ontology of the subject area, an approach is proposed to solve the reference ambiguity that arises in the process of obtaining information.

On [7] PAWS, a multilingual parallel tree bank with coreference annotation. It consists of English texts of the Wall Street magazine, translated into Czech, Russian and Polish. The main focus is on the coreference annotation in PAWS and its linguistic differences, thanks to the extensive interpretation of syntax, semantic roles, basic reference relationships and identification based on the difference between languages, it can become the basis for many different linguistic studies.

In [8], several configurations of deep neural networks focused on the basic reference solution for the Polish language are presented, in the work all systems are evaluated according to the data of the Polish Coreference Corpus with a token 540K and a note 180K. The best option is to improve the last level of the Polish language by 0.53 points and reaches 81.23 points of the CoNLL indicator.

In the following work [9], a method for solving the bridging anaphore using large databases via the QA system is proposed. They showed a new way to create a large volume of "quasi-composite" training data.

[10] Here is an example that was pre-prepared in this dataset and fine-tuned in a smaller number of datasets in the domain, showed that two Bridging get new state-of-the-art results for combining anaphora resolution (ISNotes (Markert et al., 2012) and BASHI (Rosiger, 2018)).

### 3. The Aim And Objectives Of The Study

The purpose of the research is a theoretical study aimed at developing new methods for solving reference relations, solving reference relations using methods of formal research and verification of references in Kazakh, Russian, English.

To achieve this goal, the following tasks are performed.

Task 1. A theoretical research aimed at developing new methods for solving reference relations.

Task 2. Definition of reference types in three languages using formal analysis and verification methods.

### 4. Materials And Methods Of Research

In this research work for the first time, we conducted a complete research work on the methods of solving references, thanks to which it is possible to see which types of references were investigated and determined within the framework of this research work.

#### 4.1 A theoretical research aimed at developing new methods for solving reference relations.

In fact, a lot of work has been done since the 1980s to solve anaphora and conferences, but it is impossible to cover all the work on AR in one chapter. In this section we will discuss some of the work on solving anaphora and conferences.

##### 1. Hobbs Algorithm

One of the first (and best) methods based mainly on syntactic knowledge is the Hobbs algorithm (Hobbs 1976; 1978). The appeal of this algorithm is that its simplicity provides decent performance. However, the apparent simplicity hides some non-trivial assumptions about the semantic knowledge provided by the system in which the algorithm works. [11].

##### 2. Mitkov's work in solving anaphora

Mitkov made a great contribution to the solution of anaphoric relations and identified several types of anaphora and proposed ways to solve them. Ruslan Mitkov (1997) discussed the importance of a combination of factors in his work. Mitkov gave an example based on two different views, using the same set of factors. According to him, although there are a number of ways using a similar set of factors, computational strategies may differ, so the results will also differ. He distinguishes computational strategies by counting and controlling the previous ones. [12,13,14].

Mitkov (1997) compared two different methods using the same factors. The first method is the method of complex anaphoric decision (IA), which uses constraints to exclude potential candidates, and then uses advantages to evaluate the most likely candidates (Mitkov, 94a). The second method is the method to substantiate uncertainty (IA) (Mitkov, 95), which uses the assumption that natural language understanding systems based only on ambiguity based on incomplete information are not capable of fully understanding the input information.

Mitkov's algorithm used a handwritten corpus consisting of 113 paragraphs. The result showed that IA (Integrated Anaphora Resolution method) was 83%, and URA (uncertainty proof method) was 82%.

##### 3. A Mention-Synchronous Coreference Resolution Algorithm Based on the Bell Tree (Luo et al, 2004)

(Luo et al, 2004) [15] developed a new method for solving correlation relations, calling it the Bell tree method. The method uses the Bell tree algorithm to find the best path from the root of the tree to the leaf node, and then uses the maximum entropy algorithm to calculate the probability of these paths.

##### 4 BART: A Multilingual Anaphora Resolution System

BART (Eversley and others, 2008) [16] is a highly modular set of tools for solving basic correlations, which supports the latest statistical approaches and allows you to effectively design capabilities. This system is designed for German, English and Italian.

As for the architecture of the BART system, it consists of five main parts: the preprocessing pipe, the factory, the capability output module, the decoder and the code. In addition, the independent Language Plugin handles all language-related information and is accessible from any component.

The multilingual system is designed taking into account the language characteristics of each language.

For the English language, we used analysis from the modern component parser (Petrov et al., 2006) and extracted all the main noun phrases as notes. For German, the SemEval dependency tree is converted to a composite representation, except that the noun is in a syntactic position without reference (for example, the expletive "es", predicates in copula constructs). For Italian, we use EMD Pipeline and Mention Factory. The typhoon (Zanoli et al., 2009) and dementia (Biggio et al., 2009) systems were used to recognize warnings in the test set.

As a result, for each language, BART was able to show good results in solving anaphora. For the German language, BART showed better performance than all other systems on a regular track. For English, the only language focused on all systems, BART shows good performance for all indicators in a typical setting, usually exceeding systems configured only for a specific metric. The Italian version of BART shows reliable figures for the main reference resolution, taking into account the alignment problem discussed above.

## 5. Syntactically annotated PTD corpus

In the following work [17], a diagram of the arrangement of the coreference in the syntactically annotated corpus of the PTD of Czech texts is presented. Three stages of marking are considered—the designation of grammatical correlations, in which the antecedent is calculated on the basis of the grammatical rules of this language, the designation of pronominal text correlations and an extended scheme for the designation of nominal text correlations and associative anaphora. The designation of grammatical and pronominal links was made on the entire PTD body by marking the nominal coreference and the associative anaphora.

6. Natalia N. Modjeska., Katja Markert, Malvina Nissim research team proposed a method for solving other-Anaphora criticism by machine learning for the English language, where, in addition to Morpho-syntactic, semantic features based on lexical knowledge reserves, this algorithm

Receives additional semantic knowledge from the internet. The internet is searched for criticism Other-Anaphora using special lexical and syntactic models. Adding this innovative feature will lead to an improvement in the classifier by 11.4 percentage points.

## 7. The next work is a model of reference analysis of Kibrik.

This method is based on the multi-factorial nature of reference (Kibrick, 1996, 1999, 2011). The main task is not only to find and study the factors that affect referential relationships, but also to reduce their number to the required minimum (labor reduction). [19,20.]

it also includes some common discursive features.

1) Characteristics of the referent: living or inanimate, protagonism (meaning of the referent in discourse), gender and number;

2) Signs of a leader: whether it is a member of a direct speech, the type of syntactic group, grammatical function, the form of reference, the duration of the leader in a word, the number of leaders in a sequence from the current position to the full name of the phrase.

3) Signs of anaphora: whether the first/first is not mentioned in the discourse, whether it is a direct speech member, the type of syntactic group, the grammatical role, the number of references to the referent in the sequence.

4) Distance between anaphora and antecedent: linear distance in words, linear distance in sentences, linear distance in sentences rhetorical distance of elementary discursive units, distance in paragraphs. The rhetorical distance, which is the length of the line between text fragments along the constructed rhetorical network, is considered an important factor in the reference selection, since it allows us to take into account the relationship between text fragments far from each other at a linear distance, but close in terms of presentation structure.

## 8. Solving anaphoric relationships from Twitter posts

The following [22] work was used to solve anaphora, where he developed a predictive model that could identify depressed users from Twitter posts and instantly identify textual content related to mental health topics. The model can also solve the problem of anaphoric solution and highlight anaphoric interpretations. The result of this work was 92%, and the model was far better than alternative predictive models, ranging from classic machine learning models to deep learning models.

Definition of reference types in three languages using the methods of formal analysis and verification we have considered each type separately using the method of formal analysis of all reference types in Kazakh, Russian, and English and determine the inter-language differences by the method of verification.

Various types of anaphora have been described (Hirst, 1981) and (Lap pin and Leass, 1994). In the latter, these anaphoric types were further expanded to include specific cases (Mitkov, 1999; Ng, 2010; Jurafsky and Martin, 2009).

Different types of anaphora play an important role in solving coreferential relationships, and some types of anaphora can also be repeated in solving coreferential relationships[23] so it is important to distinguish a specific type of type.

In this section, we list different types of anaphores and cataphores in Kazakh, Russian and English, give a brief explanation of their unique characteristics and distinguish them by information types.

*In the course we will consider the following types of anaphora:*

### 1. The anaphoric pronoun

Among the anaphoric types, one of the most common types is the third form of the pronoun, and the first-person singular and plural pronouns are usually used in the deictic.

The anaphora of pronouns includes all types of pronouns, but since the work studied is related to three languages, it can be seen that some types of pronouns are not anaphoric cataphoric in a sentence, due to the peculiarities of the function of pronouns in a word, personal, numerical and gender (person, number and gender PNG), all types of pronouns in English and Russian are affected by PNG. There is no gender category in the Kazakh language.

These rules are able to filter out potential candidates for anaphora (pronouns) using PNG information in languages.

We consider personal pronouns, possessive and reflexive pronouns (The Dependent form of the classification pronoun and the independent pronoun for the Kazakh language) and the independent pronouns together, these three types form one category, since they are more related to each other than the rest of the pronouns, According to Quirk et al. (2012: 345-346), these three types of pronouns form one category because they belong to each other more than do the remainder of pronouns.

These "central pronouns" are not only combined with PNG characteristics, but also play a key role in finding the antecedent, since anaphores and their predecessors usually have to indicate agreement in them. three features therefore, a person, number, and gender are of great importance for solving anaphora. In addition, Quirk et al. (2012: 335-336, 346) notes that central pronouns are the most important of all pronouns, especially personal pronouns, due to their frequency and grammatical features.

personal pronoun. - classification some features of pronouns:

- Person, number, and gender.
- anaphoric use.
- cataphore use.

Classification pronoun forms in all three languages are divided into personal plural and personal gender categories for English and Russian.

The Person number and gender classification of the classification pronoun is shown in Table 1 below.

Person	Number	Singular			plural		
		Kazakh language	Russian language	English language	Kazakh language	Russian language	English language
1 <sup>st</sup>	Мен	Я меня мне		I /me	Біз,біздер	мы	We /us
2 <sup>nd</sup>	Сен /сіз	Тебя тебе		you	Сендер / сіздер	Вам вас	you
3 <sup>rd</sup>	Ол	fem	он	She/her	Олар	Их им	they / them
		masc	Она	He/ him			
		neuter	оно	it			

Table 1 Classification Person number and gender of pronoun

Anaphoric and cataphoric use.

It is known that classification pronouns retain the meaning of expressing a particular side both individually and in context. For this reason, they are constantly used in connection with the person, that is, with the concepts of Speaker, listener and outsider. classification can be either anaphoric or cataphoric in a noun phrase. You can see the example below.

John had to go to a meeting, so he decided to have a shave.

...Асқардан сұра, ол менен гәрі сауатты ғой, дұрыс жөн нұсқар.

Космонавт вернулся на борт станции. Он сообщил, что чувствует себя нормально.

In the example above, "Космонавт", "John", "Асқар" becomes the antecedent, and the third form of the classification pronoun "he", "ол", "он" becomes anaphore.

Cataphoric use of the classification pronoun. The cataphore, in turn, is an indicator of a phenomenon or object contained in the next segment of the text. Thus, anaphora and cataphora differ from each other only in the reference vector. For clarity, let's turn to examples that illustrate the phenomenon of cataphora:

Finally, I was **her**. **she** was lying there, on the cold floor. It was **Katherine**

As we can see from the example, Katherine can be the antecedent of the object, **she, her** classification pronoun cataphora.

The cataphora also functions in Kazakh and Russian languages.

Как только она приехала в Москву, Маша позвонила

Ол мектептегі ең үздік оқушы, Айнуір үздік дипломға лайықты.

"Маша/Айнуір" in this example perform the function of antecedent, "она/ Ол" can be cataphor.

**Possessive pronoun** - Possessive pronoun there is no singular pronoun in the Kazakh language, but there is a dependent pronoun for the Kazakh language, where I used this dependency.

A possessive pronoun is divided into two classes, which are determinants and independent functions based on the function performed in the word.

The forms of determinative participle pronouns are as follows: my, your, his, her, its, our, your and their/ мой, твой, его, ее, наш, твой и их/ менің, сенің, оның, біздің және олардың.

Personal possessive pronouns take the following forms:: mine, yours, his, hers, its, ours, yours and theirs./ мое, твое, его, ее, наше, твое и их./ менікі, сенікі, оныкі, олардікі.

- Person, number and gender

Possessive pronouns distinguish personality, number, and gender through personal pronouns. Table 2 shows the PNG form of three dependent verbs.

Person	Number	Singular			Plural		
		Kazakh language	Russian language	English language	Kazakh language	Russian language	English language
1 <sup>st</sup>	Менің	мой, моя, мое, мои		my / mine	Біздің біздердің	наш, наша, наше, наши	our / ours
2 <sup>nd</sup>	Сенің сіздің	твой		your / yours	Сендер / сіздер	ваш, ваша, ваше, ваши	your / yours
3 <sup>rd</sup>	Оның	Masc	его	his	Олардың	Их	their / theirs
		fem	ее	her / hers			
		Neuter	Его ,ее	its / its			

Table 2 Person number and gender classification of possessive pronouns

- Anaphoric / cataphoric use.

Possessive pronouns, such as the classification pronoun, have anaphoric and cataphoric use in the sentence. A possessive pronoun can be anaphor in all aspects, a situation that non-anaphoric use, that is, it non-anaphoric use when referring to anyone/anything that is not explicitly mentioned in the text.

We do not like **their** programme.

Нам не нравится их программа.

I don't like your stuff.

As you can see from the example, the phenomenon that exactly refers to is unknown.

An example of using anaphor.

Since my **car** did not work, I borrowed **yours**.

We went to Arman and saw **his** new house.

**Reflexive pronouns** -in the Kazakh language, self-pronoun refers only to “өз/self” word. This pronoun is often used in a separate and common dependent form, such as өзім, өзің, өзіңіз, өзі, өзіміз;

In English, Reflexive pronouns are made by connecting the singular to the classifying pronoun – self, and in plural pronouns – selves. The possessive pronoun indicates that each person has done a certain thing himself. myself, yourself, himself,

herself, itself, oneself, ourselves, yourselves, themselves;

In Russian, Reflexive pronouns do not have gender, number, and personality categories, and there is no nominative case form.

Reflexive pronouns in Russian: себе, себя, собой, собою, сам собою, само собой, к(по) себе, от себя и т. д

- Person, number and gender (English only)

Similar to the pronouns mentioned above, in the pronoun of your own, pronouns also distinguish personality, number, and gender.

In the table below Reflexive pronouns also distinguish personality, number, and gender. pronouns also distinguish personality, number, and gender. Table 3 shows the PNG form of the Reflexive pronoun.

Person	Number singular			Plural			
	Kazakh language	Russian language	English language	Kazakh language	Russian language	English language	
1 <sup>st</sup>	өзім	Я Сам	myself	өзіміз	Сами	ourselves	
2 <sup>nd</sup>	өзің	Ты сам	yourself	Өздерің, өздеріңіз	вы сами	yourselves	
3 <sup>rd</sup>	өзі	Masc	Он сам	himself	өздері	Они сами	themselves
		Fem	Она сама	herself			
		Neuter	Оно само	itself			

Table 3 Classification of Reflexive pronouns Person number and gender

- anaphoric and cataphoric use.

In relation to anaphoric use, the Reflexive pronoun is similar to the classification pronoun, since only third-party pronouns usually perform anaphoric function.

However, both the first and second parties of reflexive pronouns can be considered as indicating some form of anaphoric reference. For example, Stirling and Huddleston (2010: 1485) note that these pronouns perform both deictic and anaphoric functions. In their opinion, self-pronouns are deictic on the one hand, since both the first and second sides refer to the addressee. On the other hand, self-pronouns are simultaneously anaphoric because they are associated with the previous word I, we, or you.

Example (a) shows the first person, example (b) shows the second person

(a) I carried the bags myself.

(b) You cannot carry the bag yourself

Table 4. Anaphoric use of central pronouns

Я посоветовал сынуі купить себеі велосипед.

It was for themselves that the friends organised the party

An overview of the most important aspects of central pronouns, as described above, is given in the tables below. They indicate whether specific features belong to an element (marked with a "+" sign) or (marked with a "-" sign).). The tables summarize information, especially related to distinguishing between anaphoric and non-anaphoric use and determining the correct antecedent of each anaphore.

Cataphoric use of central pronouns

Type of central pronouns	Language			Nubmer		Person/gender			Antecedent	
				singular	plural	male	female	neuter gender	subject	object
Personal pronouns	Kazakh language	Russian language	English language							
	ол	она	She	+	-	-	+		+	+
		он	He	+	-	+	-		+	+
		оно	it	+	-	-	-	+	+	+
	Оны/оған	Его ему им о нем	him	+	-	+	-		+	+
		Ее ей	her	+	-	-	+		+	+
		Тек жансыз заттарға	it	+	-	-	-	+	+	+
	біз	мы	we	-	+	+	+	+	+	+
	Бізді, бізге	Нас /нам /нами/о нас/	us	-	+	+	+	+	+	+
	олар	они	they	-	+	+	+	+	+	+
Оларды / оларға	Им/ их / о них	Them	-	+	+	+	+	+	+	
Determinative possessive pronouns	Ол	его	his	+	-	+	-	-	+	+
		ее	her	+	-	-	+	-	+	+
		Его ее	its	+	-	+	+	+	+	+
	біз	наш, наша, наше, наши	our	+	+	+	+	+	+	+
	Олар	их	their	+	+	+	+	+	+	+
	ол	Его ее	his/her, his or her	+	-	+	+	-	+	+
Independent possessive pronouns	маған	мой, моя, мое, мои	mine	+	+	+	+	+	+	+
	саған	твой	yours	+	+	+	+	+	+	+
	оның	его	his	+	+	+	-		+	+
Independent possessive pronouns		ее	hers	+	+	-	+	+	+	+
		Его ,ее	its	+	+	+	+	+	+	+
	Біздің сіздің	наш, наша, наше, наши	ours	+	+	+	+	+	+	+
	Олардың	их	theirs	+	+	+	+	+	+	+
	оның	Его ,ее	his/hers, his or hers	+	-	+	+	-	+	+
Reflexive pronouns:	өзі	Сам, себе, себя	himself	+	-	+	-	-	+	+
		Сама,	herself	+	-	-	+	-	+	+
		Сам ,само ,сама себе себя	itself	+	-	+	+	+	+	+
	Өзіміз, өздері	Себе,себя, сами	themselves	+	-	+	+	+	+	+
	Өзіміз, өзің	Себя,себе, сами, собой	ourselves	+	+	+	+	+	+	+
	Өзіміз, өзің	Себя,себе, сами, собой	ourselves	+	+	+	+	+	+	+

Personal pronouns	Language			Integrated				Non-integrated		
	Kazakh language	Russian language	English language	In a subordinate clause	In subordinate position within a noun phrase	At the beginning of a sentence		For rhetorical effect	The following clause	Right dislocation
						In prepositional phrases	As head of noun phrases (on their own)			
Personal pronouns	ол	она	She	+	+	+	-	+	-	+
		он	He	+	+	+	-	+	-	+
		оно	it	+	+	+	-	+	+	+
	Оны/оған	Его ему им о нем	him	+	+	+	-	+	-	+
		Ее ей	her	+	+	+	-	+	-	+
		Тек жансыз заттарға	it	+	+	+	-	+	-	+
	біз	мы	we	+	+	+	-	+	-	+
	Бізді, бізге	Нас / нам / нами/о нас/	us	+	+	+	-	+	-	+
	олар	они	they	+	+	+	-	+	-	+
	Оларды / оларға	Им/ их / о них	Them	+	+	+	-	+	-	+
Determinative possessive pronouns	Ол	его	his	+	+	+	-	+	-	-
		ее	her	+	+	+	-	+	-	-
		Его ее	its	+	+	+	-	+	-	-
	біз	наш, наша, наше, наши	our	+	+	+	-	+	-	-
	Олар	их	their	+	+	+	-	+	-	-
	ол	Его ее	his/her, his or her	+	+	+	-	+	-	-
Independent possessive pronouns	маған	мой, моя, мое, мои	mine	+	+	+	-	+	-	-
	саған	твой	yours	+	+	+	-	+	-	-
	оның	его	his	+	+	+	-	+	-	-
		ее	hers	+	+	+	-	+	-	-
		Его ,ее	its	+	+	+	-	+	-	-
Independent possessive pronouns	Біздің сіздің	наш, наша, наше, наши	ours	+	+	+	-	+	-	-
	Олардың	их	theirs	+	+	+	-	+	-	-
	оның	Его ,ее	his/hers, his or hers	+	+	+	-	+	-	-
Reflexive pronouns:	өзі	Сам, себе, себя	himself	-	-	+	+	-	-	-
		Сама,	herself	-	-	+	+	-	-	-
							+	+	-	-
	Өзіміз, өздері	Себе,себя, сами	themselves	-	-	+	+	-	-	-
	Өзіміз, өзің	Себя,себе, сами,	ourselves	-	-	+	+	-	-	-

Table 5. Cataphoric use of central pronouns

**Demonstrative pronoun**

Many researchers have pointed out that the demonstrative pronoun can be anaphoric and cataphoric.

Demonstrative pronouns have their own peculiarities to make the thought in a sentence indefinite, context-dependent and to link complete sentences. They can indicate both anaphoric and cataphoric directions[12]. Himmelmann (1996) the demonstrative pronoun indicates that there is an Endophoric (anaphoric and cataphoric) use in the sentence.

Demonstrative pronouns

language	Kazakh language	Russian language	English language
Demonstrative pronouns	ана, анау, эне, бұл, мына, мынау, міне, ол, осы, осынау, сол, сонау	ЭТОТ, ТОТ, ТАКОЙ, ТАКОВ, ТОТ-ТО, ТАКОЙ-ТО, СТОЛЬКО, СТОЛЬКО-ТО	This, that, these, those

Table 6 Demonstrative pronouns in three languages

Example of anaphoric demonstrative pronouns:

**A woman** entered the room. **This woman** (the woman) I have already seen.

Труба түбіндегі жапырық тас үй – **мехцех**. **Бұл** - әншейін келешегіне қарай қойылған ат, әйтпесе нобайы түзу бір механизм жоқ.

Хочу сказать о *Феде Иванове*. **ЭТОТ** мальчик стал плохо учиться.

This is an example of “this {woman}/ **Бұл/ ЭТОТ**” demonstrative pronouns. *and anaphor use.*

*Also, in the example below, the pronoun refer can be cataphoric.*

**This** is the best news I have heard so far today: **The TV set is working again**(Cataphoric)

Демек **мынадай**. Жас жауынгер Өтегенов взводтағы артта қалушылардың бірі болатын.

Кедергілі қашықтықты игеруі «үш» деген бағаның ол жақ, бұл жағында.

Алысқа да нашар жүгіреді. Қимылы әлі күнге шабан. (Н.Ақышев).

-Я же не говорю по-фински. — Это неважно. Главное — улыбайся.

This is an example of “**мынадай / This**”The name refers to the last case.

**Relative pronoun**

**Relative pronoun** (Latin pronomina relativa) - a relative pronoun that forms a sentence in English and Russian. Relative pronouns have an anaphoric function and have the same reference in the main sentence as in the subordinate clause. Also, the relative name can express additional information about the assigned object: its soul, inanimate, etc.

Relative pronoun cannot be used cataphorically

**Their main form is in English:** who, whom, which, whose, that and zero that<sup>1</sup>. (1Huddleston, Pullum & Peterson (2010: 1034) speak of “bare relatives” in the case of zero that.)

In Russian: *который, какой, кой, кто, что, сколько, чей, каков.*

To determine the anaphoric role of relative pronouns, it is necessary to distinguish elements of relative pronouns from related expressions. At first, it is useful to distinguish between relative and positive sentences, while positive sentences are not anaphoric. In the example below, it is not anaphoric.

*had the impression **that** she was badly ill.*

У меня было впечатление, **что** она была очень больна. (**Appositive Clauses**)

A pronoun that appears in a comparative sentence is considered anaphoric. (Quirk et al. 2012: 1119-1120).

The earthquake caused the shed to collapse, which means we need to clean it up now

## 2. Zero Anaphora

A zero anaphora is a mandatory participant in this situation, denoted by zero.

The concept of zero anaphora or zero expression is found in the works of E.V. Paducheva [20; 65], A.A. Kibrik [14;317].

The zero anaphora can also be found in Paducheva's work, where the zero anaphora was called an elliptical anaphora.Y.V. Paducheva [20;30]

This type of anaphora is particularly common in prose and decorative English and was first introduced (Fillmore, 1986). This is one of the most common types of AR tasks that uses a space in a phrase or sentence to refer to a previous word.

Willie paled and  $\emptyset$  pulled the sock up quickly

Я только что встретил Лену. Сказала, что их отдел скоро закроют

I just met Lena.  $\emptyset$  She said that their department will be closed soon.

Мен жаңа ғана Ленаны кездестірдім,  $\emptyset$  Жақын арада олардың бөлімі жабылатынын айтты

The words *Лену* / *Willie*/ in the example can be anaphor, and their antecedents are secretly given in the second part of the sentence.

## 3. One Anaphora

This type of anaphora is found in English, which means that the word "One" is used to refer to the antecedent. This type of anaphora (Ng and Cardie, 2002b) was solved on the basis of machine learning

An example of the "one anaphora" is shown in where the word "one" refers back to the word "rose":

The girl planted a red rose next to a yellow one.

## 4. Associative anaphora (bridging)

bridging – a type of anaphora in which the elements associated with anaphora are not the main reference ones.

Bridging anaphora since the 90s, empirical studies of bridging anaphora have been conducted in different genres and in different languages (Fraurud, 1990; Poesio and Vieira, 1998; Poesio, 2004; Nissim et al., 2004; Rosiger ", 2018; Poesio et al., 2018).

I came to **a room**. **The walls** were white.

В **автобусе** начался пожар. **Водитель** {автобуса} сам потушил огонь.

Мен **бөлмеге** келдім. **Қабырғалары(бөлме)** ақ түсті.

## 5. Discontinuous Sets (Split Anaphora)

This type of anaphora (Mitkov, 2014) was determined by the fact that the pronoun Bund can refer to several preceding ones, that is, it is used in cases where the antecedent is two or more. Pronouns that usually refer to several pronouns: they, them, us, both, etc. in Russian они , Нас /нам /нами/о нас/ Им/ их / о них in Kazakh олар/ Оларды /оларға/ Бізді, бізге және т.б .

## 6. Noun phrase anaphora

Typical cases of anaphora of a certain noun-a noun in which the preceding word is formed by designating a certain noun with a phrase denoting the same concept (repetition) or concepts close in meaning (for example, synonyms, subordinate words), was called anaphora of the phrase.

The noun phrase anaphora can be found in the works of E. Baha, H. Campa, B. Parti.

An example of a noun phrase anaphora.

Юрий Гагарин вернулся на борт станции. Космонавт сообщил, что чувствует себя нормально.

Юрий Гагарин борт станцияға келіп қонды, ұшқыш өзін жақсы сезінетінін мәлімдеді.

Computational linguists from many different countries attended the tutorial. The participants found it hard to cope with the speed of the presentation.

In this example, "participants/ Космонавт / ұшқыш " refers back to "Computer scientists/ Юрий Гагарин " which is a definite noun phrase that represents a semantically close concept.

## 7. Other-Anaphora.

The next type of anaphora is other anaphora or a type of anaphora that comes from the pronouns Reciprocal pronoun in English.

**other** - anaphora «**other**» or «**another**» a reference with modifiers and non-structural antecedents is NP.

This type of anaphora can be used in both anaphoric and cataphoric because it is associated with a reversible pronoun.

Mutual pronouns are cataphoric concepts. This is limited to structures in which the mutual pronoun and antecedent are added to the same noun. This is shown in the example below,

The film is about the betrayal of **each other** of **the friends**.

I have found some of the **documents**, but where are the **others**?

## 5. Research Results

As part of this research paper, we conducted a theoretical study aimed at developing new methods for solving referential relationships and conducted a study of all types of anaphors in three languages, focusing on the functions of pronouns for each language in a sentence as well as the PNG form, as they are able to filter potential anaphora candidates (pronouns) using PNG information in languages.

### 5.1 theoretical research aimed at developing new methods of solving reference resolutions.

A general characteristic of methods in solving reference resolutions is presented in the following table. The most basic methods for solving reference resolutions were theoretical studies on traditional and static methods. As a result of the study, we have found an effective and convenient way to solve the reference relations by mutual comparison of the works done so far.

Table 7  
below provides a comparative analysis of theoretical research work on methods of solving reference resolutions.

Authors Types of reference	J. Hobbs,	Mitkov, R.	Luo et al, 2004	Versley және Т.б., 2008	Natalia N. Modjeska., Katja Markert, Malvina Nissim	Nedoluzhko A	Kibrik Andrej A	Akkapon Wongkoblaj; Miguel A Vadillo, Vasa Curcin	Our method
Pronominal anaphora	+	+	+	+	-	-	+	+	+
Definite noun phrase anaphor	+	+	+	+	-	-	+	+	+
Discontinuous Sets (Split Anaphora)	+	+	+	+	-	+	+	+	+
Bridging anaphor	-	+	-	-	-	+	-	-	+
One-anaphora	-	+	-	-	-	-	-	-	+
Zero anaphora (elliptical anaphora)	-	-	-	-	-	-	+	-	+
Other-Anaphora.	-	-	-	-	+	-	-	-	+
cataphora	-	-	-	-	-	-	-	-	+
Coreference	-	-	+	+	-	+	+	-	+

Table 7 below provides a comparative analysis of theoretical research work on methods of solving referencerelations.

The table shows the types of reference and the researched works on these types, where the sign + means the presence of work performed on this type, and - the lack of research work on this type. Mitkov, R. In the works of null anaphora, Other-Anaphora, cataphora and correlation were not studied, and in the works of Nedoluzhko A. only co-reference and associative anaphora (bridging), pronoun anaphora were studied. In our work there is a study of all types of reference.

5.2 As a result of determining the types of references in the three languages using formal methods of analysis and verification, we obtained the following result.

The study, conducted in three languages, analyzed 9 types of anaphora, cataphoric types of pronoun anaphora.

In the table below we have listed anaphoric pronouns and cataphoric pronouns. In Kazakh, there are no related and dependent pronouns.

Table 8  
Types of references

types of references		Kazakh language	Russian language	English language
Pronominal anaphora	Personal pronoun	An + cat	An + cat	An + cat
	Demonstrative pronoun	An + cat	An + cat	An + cat
	Reflexive pronoun	An + cat	An + cat	An + cat
	Possessive Pronoun	An + cat	An + cat	An + cat
	Relative pronouns	-	An + cat	An + cat
Definite noun phrase anaphor		+	+	+
Discontinuous Sets (Split naphora),		+	+	+
Bridging anaphor		+	+	+
One-anaphora		+	+	+
Zero anaphora (elliptical anaphora)		+	+	+
Other-Anaphora.		An + cat	An + cat	An + cat
cataphora		+	+	+
Coreference		+	+	+

The table denotes "an-anaphora" and "kat-cataphora", "an + kat" means anaphoric and cataphoric, and "an-kat" means only anaphoric, the symbol + was used for this language in the meaning of reference type. For Kazakh, the relative pronoun is not cataphoric in anaphoric.

## 6. Discussion Of Results

In the course of the research work, all types of reference relations were analyzed and a theoretical study of methods for their solution was carried out. Here is an analysis of the plural type of reference in the three languages, among which the pronoun anaphora and the cataphor for pronoun anaphora are presented in Tables 4 and 5 for the three languages, where we can see the anaphoric and cataphoric function of each of the pronouns of the language. Second, Table 8 shows that the eight methods formulated in the theoretical research method were not solved for all types of reference.

In this paper, the limitation is due to the difference of languages, i.e. the lack of gender classification for the Kazakh language, i.e. the pronoun "ол" for the Kazakh language can meet the pronouns "she he it" i.e. for male and female other subjects for all this pronoun. The main shortcoming in the work is the neglect of the factors influencing the referential relations, we consider this shortcoming as the main task of the forthcoming work.

## 7. Discussion

*Our method gave a positive solution for all types when solving reference relations. In contrast to the methods for solving referentials in the link and more recently, they do not have a fully solved version of the types.*

*In our theoretical research work we found an application of Hobbes' algorithm in solving pronoun anaphora, Mitkov's Anaphora Ratio, and suggested ways to solve them, identifying several types of anaphora. A Luo et al, (2004) proposed a new approach to solving correlation relations and contributed to solving correlation Versley et al, 2008, Nedoluzhko A., and Natalia N. Modjeska, Katja Markert, Malvina Nissim's research group presented a method of critical Other-Anaphora solving by machine learning for English Kibrick's model of referential analysis, which is based on the multifactorial nature of referencing, aimed at solving anaphoric relations*

Using deep learning, Twitter developed a predictive model that could identify users with depression and instantly identify textual content related to mental health topics. The model was also able to solve the anaphoric resolution problem and highlight anaphoric interpretations.

Compared to these eight theoretical studies we reviewed, our method can find solution paths for all types of reference types, which is the main difference in our method, a solution method that includes a set of reference types in three languages.

## 8. Conclusions

In this paper, we conducted a theoretical study aimed primarily at developing new methods of solving referential relations. Here, when we analyze the work on solving referential relations of the past and recent times, identifying all the problems solved in it, as a result of determining on what level the referential relation was constructed, there is a new possibility for the invention of our method. As a result, it can be seen that there was not a complete study of all referential types, as shown in Table 8.

Secondly, as a result of determining the types of reference in the three languages, using formal methods of analysis and verification, we obtained the following result. In the course of the study conducted in the three languages, 9 types of anaphoric, cataphoric forms of anaphoric pronouns were analyzed. Table 8 shows anaphoric pronouns and cataphoric pronouns. In Kazakh, there are no related and dependent pronouns.

Our method has yielded positive results in solving referential relations by type, we also consider ways of solving referential relationships by identifying factors that influence referential types.

## Declarations

### Acknowledgments

"This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No AP09057872)"

### Conflicts of interests/Competing interests

Conflict of Interest

The authors declare that they have no conflict of interest

### Information about authors:

**Kalman Gulzhamal** – doctoral student of specialty [8D06103](#) "Information system" at L.N.Gumilyov Eurasian National University, Nur-Sultan, 010000, Kazakhstan. E-mail: [guljamal14@gmail.com](mailto:guljamal14@gmail.com) , <https://orcid.org/0000-0001-8863-9447>;

## References

1. Wohiduzzaman, Kazi, and Sabir Ismail. (2018) "Recommendation System for Bangla News Article with Anaphora Resolution." In Proceedings of the the International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), pp. 177–182
2. Improving Machine Learning Approaches to Coreference Resolution Vincent Ng and Claire Cardie
3. Zhang H, Song Y, Song Y (2019a) Incorporating context and external knowledge for pronoun coreference resolution. arXiv preprint arXiv:190510238
4. Deep Reinforcement Learning for Mention-Ranking Coreference Models Kevin Clark Christopher D. Manning 2016
5. Мульти-агентный подход к разрешению кореференции на основе многофакторного сходства при пополнении онтологий Н. О. Гаранина, Е. А. Сидорова, А.С. Серый 2017
6. Подход к разрешению референциальной неоднозначности текста при пополнении онтологии Е. А. Сидорова Н. О. Гаранина И. С. Кононенко 2017
7. Paws: a multi-lingual parallel treebank with anaphoric relations Anna Nedoluzhko, Michal Novak, Maciej Ogrodniczuk 2018
8. Bartłomiej Nitoń, Paweł Morawiecki, and Maciej Ogrodniczuk. 2018. [Deep Neural Networks for Coreference Resolution for Polish](#). In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA).
9. Yufang Hou. 2020. [Bridging Anaphora Resolution as Question Answering](#). In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1428–1438, Online. Association for Computational Linguistics.
10. Muller, Mathias, Annette Rios, Elena Voita, and Rico Sennrich. 2018. A large-scale test set for the evaluation of context-aware pronoun translation in neural machine translation. In Proceedings of the Third Conference on Machine Translation, Volume 1: Research Papers, pages 61–72, Brussels, Belgium, October. Association for Computational Linguistic
11. Hobbs, J. R. (1978): Resolving pronoun references. *Lingua*, 44, pp. 339–352.
12. Mitkov, R. (1994a): An Integrated Model for Anaphora Resolution. COLING '94
13. Mitkov, R. (1995b): An uncertainty reasoning approach for anaphora resolution.
14. Mitkov, R. (1997): Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches, In Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution, pp. 14–21. Madrid, Spain.
15. Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing et al. A Mention-Synchronous Coreference Resolution Algorithm Based on the Bell Tree In Proc. of the ACL. -- 2004. -- P. 135–142.
16. Samuel Broscheit, Massimo Poesio, Simone Paolo Ponzetto, Kepa Joseba Rodríguez, Lorenza Romano, Olga Uryupina, Yannick Versley, Roberto Zanolli: BART: A Multilingual Anaphora Resolution System. [SemEval@ACL 2010](#): 104-107
17. Недолужко а разметка кореференции на синтаксически аннотированном корпусе чешских текстов [charles university in prague, prague, czech republic;dialogue, 2009](#)
18. Natalia N. Modjeska, Katja Markert, and Malvina Nissim. 2003. [Using the Web in Machine Learning for Other-Anaphora Resolution](#). In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, pages 176–183.
19. Kibrik Andrej A. (2011). Reference in discourse. Oxford: Oxford University Press.
20. Kibrik, Andrej A. 1996. Anaphora in Russian narrative discourse: A cognitive calculative account. — In: Fox (ed.), *Studies in anaphora*, 255-304. Amsterdam: Benjamins.
21. Kibrik, Andrej A. 1999. Cognitive inferences from discourse observations: Reference and working memory. — In: *Discourse studies in cognitive linguistics*. Proceedings of the 5th International cognitive linguistics conference, ed. Karen van Hoek, Andrej A.Kibrik, and Leo Noordman, 29-52. Amsterdam: Benjamins.

22. Akkapon Wongkoblak, Miguel A Vellido, Vasa Curcin, Deep Learning With Anaphora Resolution for the Detection of Tweeters With Depression: Algorithm Development and Validation Study, JMIR Mental Health, August 2021.
23. Nikolaos Stylianou, Ioannis Vlahavas, A Neural Entity Coreference Resolution review, Expert Systems with Applications, Volume 168, 15 April 2021