IDENTIFYING SEMANTIC RELATIONS BETWEEN SENTENCES
BY SOLVING AN ANAPHORA

This paper presents the method of semantic relations identification by solving the anaphora, specifically to determine the nature of relations between sentences in scientific texts. Authors explored the use of the pronoun "anaphora", developed a mathematical predicate model to determine the anaphora and antecedent and described the relationships between them for fragments in the Ukrainian language. Authors conducted an experiment based on the developed model and the logical network that was build based on presented predicate model.

Keywords: semantic relations, pronoun anaphora, antecedent, Ukrainian, algebra of predicates and predicate operations, logic network, natural language processing.

Introduction

Problem Statement. The exponential increase in the amount of textual information contributes to the emergence of new tasks for natural language processing. Search engine technologies, machine automatic translation methods, text summarization, and other language processing methods need constant improvement. One of the tasks among the listed areas is to create a new method for determining semantic relations between phrases, within one sentence, and between different sentences. Scientific publications explore approaches to identifying semantic relations by defining anaphora. The determining anaphora place and role in syntactic constructions is a separate task of computational linguistics.

Analysis of Relevant Works. Semantic relations in syntactic constructions are studied via anaphora and antecedent in computational linguistics. Sentences, phrases, as well as parts of sentences, and sentence integrations are syntactic constructions that have anaphoric relations. This paper focuses on the semantic relations between sentences.

The problem of linguistic anaphora has been the subject of many studies [1 –6]. The functionality of known approaches of the system analysis to the decision of an anaphora question of Ukrainian natural-language texts is investigated in work [4]. The algorithm for calculating semantic-syntactic criteria for a pre-selected antecedent-anaphora pair is presented. Peculiarities of the semantic-syntactic structure of the pronoun anaphora its functions in the literary text are investigated by the author [5]. The publication [6] proposes an approach to solving the pronoun anaphora by constructing a classifier. Based on the syntactic and semantic properties of the anaphora and antecedent a classifier draws a conclusion about their compatibility or incompatibility.

Extraction approach for information abstracting is implemented in [7]. Anaphoric relations are used to solve the automatic generalization of textual information on the Internet. Automatic text generalization is the process of creating a summary of one or more related documents that stores only the most valuable information.

The paper purpose is to present the method of semantic relations identification by solving the anaphora: to determine the nature of relations between sentences in scientific texts; explore the use of the pronoun “anaphora”; develop a mathematical predicate model to determine the anaphora and antecedent, and describe the relationship between them for fragments in the Ukrainian language and build an appropriate logical network; to conduct an experiment based on the developed model.

Statement of basic materials

Types of relations between sentences in scientific texts

All texts can be divided into three varieties according to the relationship types between sentences [8]:
- texts with parallel (remote) relations;
- texts with serial (chain) relations;
- texts with relations or with a combined relations.

In parallel relations, the sentences are equal. Parallel relation is the use of sentences in which the same word order, the same grammatical forms of the sentence members. The main means of implementing parallel relations is syntactic parallelism. This is when the same or similar construction of sentences, which is often expressed in the same sequence of words, and the unity of temporal forms of verbs-predicates (predicates) [8].

Example of using parallel communication in scientific texts:
“Ступінь розвитку Web-простору буде визначатися технологіями роботи з величезним обсягом інформації, що накопичилась в Інтернет. Web наступного покоління буде характеризуватися переходом від мережі документів до мережі даних, що при необхідності агрегуються в семантично зв’язані документи за допомогою Web-сервісів.”
The degree of Web-space development will be determined by the technology of working with a huge amount of information accumulated on the Internet. The next-generation Web will be characterized by the transition from a network of documents to a network of data, which, if necessary, are aggregated into semantically related documents using Web-services.

The article's research object is a scientific text. Concepts and actions must be clearly and unambiguously presented in a scientific text in order of correct understanding by any researcher-reader. Correct understanding ensures the predominant use of sequential (chain) relations in academic texts.

Chain relation is used in all language styles. This is the most common way of connecting sentences, which is most consistent with the specifics of thinking. The chain relation is necessarily present where the thought develops sequentially, where each subsequent sentence develops the previous one as if formed from it. The relationship between sentences in a sequential relationship is great, so without the previous sentence, the meaning of the next one may not be clear.

Serial or chain relations exist because the complement of the previous sentence becomes the subject in the next sentence. The structural form of this relation is as follows: “complement-subject”. Other models of the sentence structure are also widely used: “subject-complement”, “complement-object”, “subject-subject”. The syntactic essence of the chain relation is expressed in these syntactic models, in the syntactic relations between neighboring members of the sentence. This is the internal, structural side of the chain relation. There are ways to embody syntactic relations in a serial relation [8]: lexical repetition, synonyms, indicative words, personal pronouns, pronoun adverbs, conjunctions, verbal omission, etc. Thus, we can distinguish: 1) chain relation by lexical repetition, 2) chain synonymous relation, 3) chain pronoun relation.

The syntactic essence of the chain relation, its structural types, models (the relation of the sentence member pairs) are preserved regardless of the means of model implementation:

“Перше покоління моделей інноваційних процесів припадає на 1950–1960-ті рр. На думку Р. Росвела, це покоління можна описати як «модель, яка підпідсвічується технологіями»;

“Процес автоматизації у період експлуатації ППГ, наприклад для поліетиленової плівки, характеризується трьома стадіями: періодом індукуції, якій відповідає стадія зародження ланцюгів; періодом, прискорення, якому відповідає стадія росту ланцюгів; періодом співвідношення, що відповідає стадії обриву ланцюгів. Відповідно до зміни властивостей полімеру, розподіляють ці процеси на агрегативні, пов’язані з розпадом макромолекул на більші структурні фрагменти”.

“The first generation of innovation processes dates back to the 1950s and 1960s. According to R. Roswell, this generation can be described as a "technology-driven model”;

“The process of auto-oxidation during the operation of PPT, for example for polyethylene film, is characterized by three stages: the period of induction, which corresponds to the stage of nucleation of chains; period, acceleration, which corresponds to the stages of growth of the chains; the deceleration period corresponding to the stage of chain breakage. According to the change in the properties of polymer raw materials, these processes are divided into aggregative, associated with cross-linking processes, and destructive, associated with the decomposition of macromolecules into smaller fragments.”

Using the pronoun “anaphora”

There are the following means of relation: words close in meaning, pronouns, adverbs, numerals, and other means, repetition of words, words that indicate the sequence of development of the content. According to this, we can distinguish the following types of anaphora: pronoun, noun, adverb, and zero [6].

The most common type of anaphora in natural language is the pronoun anaphora [3]. In computational linguistics, anaphora is usually defined as a reference to objects mentioned earlier in the discourse, or a reference to a “pointing back”. “Pointing back” (reference) is called an anaphora, and the essence to which it refers is its predecessor - the antecedent [2].

Example of pronoun anaphora:

“Найбільш вираженим у плані динаміки є, безперечно, сегмент інформації у вигляді новин. З одного боку, він має найвищий рівень оновлення, а з іншого боку – у ньому генеруються і поширюються насправді великі обсяги даних”.

(“The most pronounced in terms of dynamics is, of course, the segment of information in the form of news. On the one hand, it has the highest level of updating, and on the other hand – it generates and distributes really large amounts of data.”)

One of the most commonly used relations is between the anaphoric pronoun and the antecedent. “Anaphoric pronouns are pronouns that refer to some other words or phrase (antecedent) of this text, the semantic meaning of which they reflect” [9].

For research, we used our own text corpus, which consists of scientific articles and publications. The statistical characteristics of this corpus are presented in the Tabl. 1.

From the total number of words in the Ukrainian text corpus, 5.3% are pronouns (21923 out of 415565). Only 14 (0.5%) Ukrainian fragments out of the total
number of 2633 do not contain pronouns. With an average fragment length of 157 words, the maximum number of pronouns in a fragment is 32, the average is 9. These data indicate the \( P(x_1, x_2, \ldots, x_n) = Z \) frequent use of anaphoric pronouns.

<table>
<thead>
<tr>
<th>Statistical indicators of the corpus</th>
<th>Ukrainian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of authors</td>
<td>32</td>
</tr>
<tr>
<td>Number of documents</td>
<td>271</td>
</tr>
<tr>
<td>Number of fragments</td>
<td>2633</td>
</tr>
<tr>
<td>The average number of fragments per author</td>
<td>82</td>
</tr>
<tr>
<td>Total size (tokens)</td>
<td>415565</td>
</tr>
<tr>
<td>The average number of tokens per author</td>
<td>12986</td>
</tr>
<tr>
<td>The average number of tokens in the fragment</td>
<td>157</td>
</tr>
</tbody>
</table>

The presence of pronouns in fragments of the Ukrainian text corpus is presented in the Tabl. 2.

The solution of anaphora is based on syntactic and morphological information. According to these features, the antecedent in relation to the anaphora is identified. Antecedent and anaphora should have similar characteristics. 90% of antecedents are in the same or previous sentence as the anaphora [10].

<table>
<thead>
<tr>
<th>Types of pronouns and their number (text corpus, Ukrainian language)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of pronoun</td>
</tr>
<tr>
<td>Defining</td>
</tr>
<tr>
<td>Interrogative</td>
</tr>
<tr>
<td>Personal</td>
</tr>
<tr>
<td>Indicative</td>
</tr>
</tbody>
</table>

The mathematical model development for determining anaphora and antecedent

To describe the semantic relations between sentences and solve the anaphora, we will use the apparatus of finite predicate algebra (FPA) and predicate operations (PO) [11–12]. It is known [13–14] that the algebra of finite predicates is complete, i.e. any finite relations can be described in its language.

Algebra of finite predicates is a formal language with which it is possible to describe mathematically investigated structures and functions. It is used to formalize the description of deterministic, discrete, finite objects. Deterministic processes are processes with an unambiguous result; they do not have a factor of randomness. Discrete processes are processes in which information has the form of separate portions or quantum. They lack the factor of continuity, for example, numbers, letters, words, formulas. Finite processes are processes in which there is no infinity factor, i.e. only a finite number of information units can participate in them [11–12]. This is how we describe our task of natural language processing and defining anaphoric relations.

For our problem, a predicate based on the Cartesian product \( A_1 \times A_2 \times \ldots \times A_n \), is any function that represents the Cartesian product \( A_1 \times A_2 \times \ldots \times A_n \) sets in the mathematical set \( Z = \{0; 1\} \). Symbols 0 and 1 are called Boolean elements, \( Z \) is called the set of all Boolean elements.

The variable \( Z = \{0; 1\} \) is called Boolean and it is also the value of the predicate \( P \). The predicate \( P(x_1, x_2, \ldots, x_n) \) is a function and \( x_1, x_2, \ldots, x_n \) are the arguments of the predicate \( P \).

We formulate a rule that allows the transition from the relation to the corresponding predicates and in the opposite direction. Let \( L \) be the set of all relations on \( S \). The relations \( P \) with \( L \) and the predicate \( P \) with \( M \) are called corresponding to each other, if for any \( x_1 \in A_1, x_2 \in A_2, \ldots, x_n \in A_n \):

\[
P(x_1, x_2, \ldots, x_n) = \begin{cases} 1, & \text{if } (x_1, x_2, \ldots, x_n) \in P; \\ 0, & \text{if } (x_1, x_2, \ldots, x_n) \notin P. \end{cases} \tag{1}
\]

If we know the relation \( P \), then we can always use rule (1) to trace all the values of the predicate \( P \), i.e. to give a complete description of it (for example, to represent the predicate \( P \) in the graph form).

The inverse transition from the predicate \( P \) to the relation \( P \) can be performed using the following rule:

\[
\text{if } P(x_1, x_2, \ldots, x_n) = 1, \text{ then } (x_1, x_2, \ldots, x_n) \in P; \quad \text{if } P(x_1, x_2, \ldots, x_n) = 0, \text{ then } (x_1, x_2, \ldots, x_n) \notin P. \tag{2}
\]

Expression (2) allows for known Boolean values of the predicate \( P \) to find all sets of predicates, which are preceding the relation \( P \).

Rules (1) and (2) establish a balanced correspondence between all relations of the set \( L \) and all predicates of the set \( M \), which are given in \( S \). The set of all vectors \( (x_1, x_2, \ldots, x_n) \) that satisfy the equation \( P \) \( (x_1, x_2, \ldots, x_n) = 1 \), forms the relation \( P \), which is called the truth domain of the predicate \( P \). The predicate \( P \in M \), which is given by rule (1), is called the characteristic function of the relation \( P \in L \).

For Boolean elements as well as for predicates there are operations of disjunction, conjunction and negation.

Let’s make a formal process description of the anaphoric relation identifying. We will develop a mathematical model to determine the anaphora and the antecedent and the relationship between them for fragments.
in the Ukrainian language and build an appropriate logical network. In this work we analyze only one type of connection construction, it has 3-person pronouns and a noun in the singular.

We identify seven features that identify the antecedent (tabl. 3). The features selected are complete, consistent and unbreakable.

Table 3

<table>
<thead>
<tr>
<th>№</th>
<th>Name of attribute</th>
<th>Notation</th>
<th>Attribute value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Location in the text (regarding the pronoun (anaphora))</td>
<td>( X_1 = \left{ x_1^1, x_1^2 \right} )</td>
<td>( x_1^1 ) – before the anaphora ( x_1^2 ) – after the anaphora</td>
</tr>
<tr>
<td>2.</td>
<td>Part of speech is a noun</td>
<td>( X_2 = \left{ x_2^1, x_2^2 \right} )</td>
<td>( x_2^1 ) – noun ( x_2^2 ) – other parts</td>
</tr>
<tr>
<td>3.</td>
<td>Grammatical gender</td>
<td>( X_3 = \left{ x_3^1, x_3^2 \right} )</td>
<td>( x_3^1 ) – masculine ( x_3^2 ) – feminine ( x_3^3 ) – neutral</td>
</tr>
<tr>
<td>4.</td>
<td>Number</td>
<td>( X_4 = \left{ x_4^1, x_4^2 \right} )</td>
<td>( x_4^1 ) – plural ( x_4^2 ) – singular</td>
</tr>
<tr>
<td>5.</td>
<td>Grammatical case</td>
<td>( X_5 = \left{ x_5^1, x_5^2, x_5^3, x_5^4, x_5^5, x_5^6 \right} )</td>
<td>( x_5^1 ) – Subjective ( x_5^2 ) – Genitive ( x_5^3 ) – Dative ( x_5^4 ) – Accusative ( x_5^5 ) – Ablative ( x_5^6 ) – Prepositional</td>
</tr>
<tr>
<td>6.</td>
<td>Proper names</td>
<td>( X_6 = \left{ x_6^1, x_6^2 \right} )</td>
<td>( x_6^1 ) – proper name ( x_6^2 ) – not proper name</td>
</tr>
<tr>
<td>7.</td>
<td>The distance between a pronoun and a potential antecedent in words</td>
<td>( X_7 = \left{ x_7^1, x_7^2 \right} )</td>
<td>( x_7^1 ) – less than 30 words ( x_7^2 ) – more than 30 words</td>
</tr>
</tbody>
</table>

The value change area for a feature with 6 values of \( i \in \{ 1, 2, 4, 6, 7 \} \) is as follows:

\[
\begin{align*}
\{ x_1^1 \} \land \{ x_1^2 \} & = 0, \\
\{ x_1^1 \} \land \{ x_1^2 \} & = 1.
\end{align*}
\]

For attributes with three values of \( i \in \{ 3 \} \) we have the following equations:

\[
\begin{align*}
\{ x_3^1 \} \land \{ x_3^2 \} \land \{ x_3^3 \} & = 0, \\
\{ x_3^1 \} \land \{ x_3^2 \} \land \{ x_3^3 \} & = 0, \\
\{ x_3^1 \} \land \{ x_3^2 \} \land \{ x_3^3 \} & = 1.
\end{align*}
\]

The set of anaphora values (defining of anaphoric relation) is binary:

\[
P(x_1, x_2, ..., x_n) = Z, \quad \text{where} \quad Z = \{0, 1\}.
\]

The position in the text (before the pronoun (anaphora)) is important. The pretender to the antecedent should be before the anaphoric pronoun, preferably in the same sentence or in the next. Therefore, for \( x_1^2 = 1 \) and for any values of other attributes \( \forall x \in \{ x_2, x_3, x_4, x_5, x_6, x_7 \} \) we have:

\[
P(x_1, x_2, ..., x_n) = 0.
\]

Based on the attributes combination and the relations between them, we build a logical network, which is a graphical representation of the Cartesian product of all mentioned attributes. Logical networks are a universal, simple, and natural means of visual representation of any
object structure. From a mathematical point of view, logical networks are a system of binary predicates [14].

Any logical network consists of poles and edges. Each pole has its own subject variable (pole attribute). The pole is denoted by its own subject variable. The pole attribute change area (domain) is associated with the pole. Any pole of a logical network at any given time reflects certain knowledge of attribute value. It is called the state of the pole. It is possible to obtain the state of a network at a given moment by specifying the state of all its poles at that moment [13].

Each edge of a logical network is assigned its own binary relation, which is called the relation of this edge. Each edge is denoted by its relation number. An edge unites two poles that meet the subject variables that are related to this edge [13].

To build a logical network, perform decomposition. Decomposition plays an important role. It provides the replacement of one complex equation (task) with an equivalent system of simpler equations (tasks). Thus, decomposition serves as a powerful means of simplifying and reducing the notation of finite predicate algebra equations [12].

After performing the decomposition, we group the attributes related by type and form aggregate indicators for each type of attribute.

The first group of attributes determines the location of the word, which is analyzed for the role of the antecedent: \( X_1 \) is a location in the text (regarding the pronoun (anaphora)), \( X_7 \) is the distance between a pronoun and a potential antecedent in words, they form an aggregate indicator \( Q_1 \).

The second group of attributes analyzes the word about the possibility of performing the antecedent role: \( X_2 \) is part of speech (noun), \( X_6 \) is a proper name. These attributes form an aggregate indicator \( Q_2 \).

Morphological attributes of a potential antecedent characterize: \( X_3 \) is grammatical gender, \( X_4 \) is number, \( X_5 \) is grammatical case. They form an aggregate indicator \( Q_3 \). We divide the values of each aggregate indicator into the following values.

The indicator \( Q_1 = \{q_1^1, q_1^2\} \) has the following values: \( q_1^1 \) relation is possible, \( q_1^2 \) relation is impossible. The logical network of attributes for this indicator is presented in Fig. 1.

We have the following equation for a logical network of attributes \( Q_1 \):

\[
q_1^1(x_1^1, x_1^2) = 1
\]

The indicator \( Q_2 = \{q_2^1, q_2^2\} \) has the following values: \( q_2^1 \) – can play the role of an antecedent, \( q_2^2 \) cannot play the role of an antecedent (Fig. 2).

\[
q_2^1(x_1^1, x_1^2 \lor x_2^2 \lor x_2^2_2 \lor x_3^3_2 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1) = 1
\]

There is the equation for a logical network of attributes \( Q_2 \):

\[
q_2^3(x_1^1, x_1^2 \lor x_2^2 \lor x_2^2_2 \lor x_3^3_2 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1) = 1
\]

The indicator \( Q_3 = \{q_3^1, q_3^3\} \) has the following values: \( q_3^1 \) – potential antecedent and anaphora have similar morphological attributes; \( q_3^3 \) – morphological attributes of a potential antecedent do not meet the attributes of an anaphoric pronoun (Fig. 3).

This is the equation for \( Q_3 \):

\[
q_3^1(x_1^1, x_1^2 \lor x_2^2 \lor x_2^2_2 \lor x_3^3_2 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_4^4_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1 \lor x_5^5_2_1) = 1
\]

According to the aggregate indicators, we have a binary value of the indicator \( R \) (result), which shows the presence or absence of an anaphoric relation. The process of creating aggregate indicators in general is presented in Fig. 4.
Let’s enter a set of values of $R$ (result), which shows the presence or absence of an anaphoric relation:

$$R = \{r_1, r_2\},$$

where $r_1$ – an anaphoric relation has been identified, $r_2$ – there is no anaphoric relation.

Thus, using the apparatus of finite predicate algebra (FPA) and predicate operations (PO), we created a mathematical model that describes the process of solving anaphora and antecedent in scientific articles in Ukrainian.

**Experiments and Results**

All experiments were carried out on our corpus described in the “Using the pronoun “anaphora” section. As a result of processing texts from this corpus, we received the following statistical information of the pronouns which are used for resolving the anaphora.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fragments</td>
<td>2633</td>
</tr>
<tr>
<td>Average number of sentences</td>
<td>11</td>
</tr>
<tr>
<td>Average number of pronouns per sentence</td>
<td>9</td>
</tr>
</tbody>
</table>

The Tabl. 4 shows that scientific texts often contain pronouns. Fig. 5 shows the most frequent pronouns that are found in our corpus. These pronouns have frequency more than 500 in our corpus.

These are example sentences from our corpus, the semantic relations are found between them.

**Electroenergetics** – це базова галузь національної економіки, стабільність роботи якої для розвитку країни має особливе значення. **Вона** впливає не тільки на розвиток національного господарства, а і на територіальну організацію продуктивних сил.

(Electricity is a basic branch of the national economy, the stability of which is of particular importance for the development of the country. It affects not only the development of the national economy, but also the territorial organization of productive forces.)

Our method identified the pronoun “Вона” (“She”) and found the corresponding antecedent “Електроенеретика” (“Electroenergetics”).

The experiments showed that 96% of semantic relations (using the pronominal anaphora) were found in our corpus. Cases when the model did not work related to the syntactic features of sentences, errors in the work of the morphological analyzer, and the distance between the anaphora and antecedent.

**Conclusions and Recommendations**

The authors analyzed various means of communication: words close in meaning, pronouns, adverbs, numerals and other means, repetition of words, words that indicate the sequence of development of the content. The main attention in the paper is paid to the pronoun anaphora, for which a predicate mathematical model was developed. This model defines anaphora and antecedent, and the relationship between them. All research was conducted on the Ukrainian corpus of scientific publications created by the authors. The relationship between anaphora and antecedent for fragments in Ukrainian was described using algebra of finite predicates and predicate operations.

With the help of selected features, it was possible to form aggregate indicators for each type and set of these attributes. Based on this, a logical network is built, which is a graphical representation of the Cartesian product of all the attributes considered. The authors conducted an experiment using the developed model and a logical network built on the basis of the presented predicate model. As a result of the experiments, it was determined that about 96% of semantic relations were found in the text, cases when the model did not work were related to the syntactic features of sentences and their distance. The developed predicate model can be used in various applications that are related to the processing of natural language, namely the summarization task, semantic similarity of sentences, etc.

**References**


2. Mitkov, R. (1999), *Anaphora resolution: the state of the art*, School of Languages and European Studies, University of Wolverhampton, pp. 1-34.


2. Mitkov R. Anaphora resolution: the state of the art / R. Mitkov. – School of Languages and European Studies, University of Wolverhampton, 1999. – P. 1-34.


Список літератури


ВІЗНАЧЕННЯ СЕМАНТИЧНИХ ЗВ’ЯЗКІВ МІЖ РЕЧЕННЯМИ ШЛЯХОМ ВИРІШЕННЯ АНАФОРИ
Ю.М. Гліччева, О.В. Каніщева, М.А. Вовк

Одним із завдань обробки природної мови є створення нових методів з визначення семантичних зв’язків між словосполученнями, в межах одного речення та між різними реченнями. У даній статті представлено метод ідентифікації семантичних зв’язків шляхом визначення анафори, а саме, для визначення характеру відношень між реченнями у наукових текстах. Для опису семантичних зв’язків між реченнями та вирішення анафори було використано алгебру скінчених предикатів та предикатних операцій. Це обумовлено тим, що алгебра скінчених предикатів є повною, більшою природної мови, а саме задачі реферування, семантичної близькості речень та ін.

У роботі авторами виділено наступні засоби зв’язку: близькі за змістом слова, займенники, прислівники, числівники та інші засоби, повторення слів, слова, що вказують на послідовність розгортання змісту. Найпоширенішим типом анафори в природній мові є займенникова анафора, тому автори зосередили своє дослідження саме на використанні таких засобів.

За допомогою алгебри скінчених предикатів та предикатних операцій було описано відношення між анафорою та антецедентом. Всі дослідження проводились на створеному авторами українському корпусі наукових публікацій. За допомогою алгебри скінчених предикатів та предикатних операцій було описано відношення між анафорою та антецедентом для фрагмента українською мовою.

За допомогою використання ознак було сформульовано правила відносно виявлення початкових ознак анафори. На основі цього побудовано логічну мережу, яка представляє собою графічне представлення декартового добутку усіх можливих комбінацій анафори та антецеденту. В результаті проведених експериментів було виявлено, що більш 96% семантичних зв’язків було знайдено у тексті, випадки коли модель не спрацьовувала стосувались синтаксичних особливостей речення та їх відстанню.

ОДИМ ЗАВДАНЬ АЛГОБРИ СКІНЧЕНИХ ПРЕДИКАТИВІВ ТА ПРЕДИКАТНИХ ОПЕРАЦІЙ
Ю.М. Гліччева, О.В. Каніщева, М.А. Вовк

В данной статье представлен метод определения семантических связей путем определения анафоры, в частности, для определения характера отношений между предложениями в научных текстах. Авторы исследовали использование местоимения "ананора", разработали предикатную математическую модель для определения анафоры в антецеденте, и описали отношения между ними для фрагментов в украинском языке. Авторы провели эксперимент на основе разработанной модели и логической сети, построенной на основе представленной предикатной модели.

Ключевые слова: семантические отношения, местоименная анафора, антецедент, украинский язык, алгебра предикатов и операций предиката, логическая сеть, обработка естественного языка.