https://www.proconference.org/index.php/gec/article/view/gec37-00-005
DOI: 10.30890/2709-1783.2025-37-00-005

UDC 519.7

ALGEBRAIC REPRESENTATION OF SIMPLE WORD-COMBINATIONS (ON THE EXAMPLE OF THE UKRAINIAN AND ENGLISH LANGUAGES)

Yakimova N. A.
c. t. s., as. prof.
Odessa I.I. Mechnikov National University,
Odessa, Dvoryanska, 2, 65029, Ukraine

Abstract. *In modern linguistics, the task of mathematical formalization of natural language for the purpose of its further computer processing is becoming more and more relevant. The mathematical apparatus of vector logical algebra, which is a development of the algebra of finite predicates, makes it possible to develop computer linguistics. This article shows the fundamental possibility of mathematically formalizing the simplest speech constructions using the basic operations of vector logical algebra.*

Keywords: *logical scalar, logical vector, sentential connections, Boolean operations, simple word-combinations of natural language.*

Introduction.

A piece of text consists of sentences, and each sentence is composed of statements that may contain "sub-statements" and ultimately consist of words [1]. The grammar of each natural language defines the way in which words and expressions can be combined into sentences. In each simple word-combination, the grammatical presence of the main word and dependent word is provided. Each of them, in turn, can be represented by different parts of speech. The types of grammatical connection between these words can also be different. The grammars of different groups of natural languages provide for different types of such a connection. Thus, the atomic level of any natural language is a single word, and the simplest construction is a simple word-combination consisting of two words.

Main text.

Let's consider sentences from a mathematical point of view, that is, as some constructions of vector logical algebra [2]. We will conduct research on the examples of the Ukrainian and English languages, which are representatives of different groups of natural languages. As an example, let's take the object «***flower***» («***квітка***»), which has the quality «***blue***» («***синя***»). Let the object «***flower***» correspond to the vector ***l*** of some logical space ***L***, and the quality «***blue***» to the scalar α [3]. Then the expression «***blue flower***» («***синя квітка***») corresponds to the product of the vector $\boldsymbol{l}$ by the scalar α. That is, this statement is formalized by the expression $\alpha\, \boldsymbol{l}$. Let's check the fulfillment of the axioms of logical algebra for the selected vector space and the selected scalar field. In this example, the set of all objects is taken as a vector space, and the set of all qualities of objects is taken as a scalar field. Boolean operations of inversion, disjunction, and conjunction are specified on a set of vectors. Then for word-combinations of the Ukrainian language:

$$\text{«не квітка»} = \text{«}\overline{\text{квітка}}\text{»} = \bar{\boldsymbol{l}},$$

and for English word-combinations:

$$\text{«not a flower»} = \text{«}\overline{\text{flower}}\text{»} = \bar{\boldsymbol{l}}.$$

Let the object «***handkerchief***» («***хустка***») correspond to some vector $\boldsymbol{g} \in \boldsymbol{L}$. Then in the Ukrainian language we have:

$$\text{«квітка або хустка»} = \text{«квітка»} \vee \text{«хустка»} = \boldsymbol{l} \vee \boldsymbol{g},$$

$$\text{«квітка та хустка»} = \text{«квітка»} \wedge \text{«хустка»} = \boldsymbol{l} \wedge \boldsymbol{g}.$$

For the English language, these formulas look like this:

$$\text{«a flower or a handkerchief»} = \text{«flower»} \vee \text{«handkerchief»} = \boldsymbol{l} \vee \boldsymbol{g},$$

$$\text{«a flower and a handkerchief»} = \text{«flower»} \wedge \text{«handkerchief»} = \boldsymbol{l} \wedge \boldsymbol{g}.$$

The same Boolean operations are specified on the set of elements of the scalar field. Then for the Ukrainian language

$$\text{«не синій»} = \text{«}\overline{\text{синій}}\text{»} = \bar{\alpha},$$

and for the English language

$$\text{«not blue»} = \text{«}\overline{\text{blue}}\text{»} = \bar{\alpha}.$$

Let the quality «***large***» («***великий***») correspond to the logical scalar β. Then for the Ukrainian language

$$\text{«синій або великий»} = \text{«синій»} \vee \text{«великий»} = \alpha \vee \beta,$$

$$\text{«синій та великий»} = \text{«синій»} \wedge \text{«великий»} = \alpha \wedge \beta.$$

For the English language, we have the following results:

$$«\textit{blue or large}» = «\textit{blue}» \vee «\textit{large}» = \alpha \vee \beta,$$

$$«\textit{blue and large}» = «\textit{blue}» \wedge «\textit{large}» = \alpha \wedge \beta.$$

Therefore, the operations of disjunction, conjunction and inversion of objects and their qualities given in this way satisfy all axioms of the logical field [4]. Sentential relations are the names of the corresponding operations. The auxiliary parts of speech **«or»** (**«або»**), **«and»** (**«і (й, та)»**) and **«not (no)»** (**«не (ні)»**) are considered respectively as the names of the operations of disjunction, conjunction and inversion [5].

It is obvious that the axioms connecting the product of logical scalars and logical vectors also hold. In particular, on the example of the Ukrainian and English languages:

<u>*associativity*</u>

$(\alpha\ \beta)\ l$ = **«(*синя* та *велика*) *квітка*»** =

= **«*синя*» та («*велика квітка*»)** = $\alpha\ (\beta\ l)$;

$(\alpha\ \beta)\ l$ = **«(*blue* and *large*) *flower*»** =

= **«*blue*» and («*large flower*»)** = $\alpha\ (\beta\ l)$;

<u>*distributivity about the disjunction of scalars*</u>

$(\alpha \vee \beta)\ l$ = **«(*синя* або *велика*) *квітка*»** =

= **«*синя квітка*» або «*велика квітка*»** = $\alpha\ l \vee \beta\ l$;

$(\alpha \vee \beta)\ l$ = **«(*blue* or *large*) *flower*»** =

= **«*blue flower*» or «*large flower*»** = $\alpha\ l \vee \beta\ l$;

<u>*distributivity about the disjunction of vectors*</u>

$\alpha\ (l \vee g)$ = **«*синя* (*квітка* або *хустка*)»** =

= **«*синя квітка*» або «*синя хустка*»** = $\alpha\ l \vee \alpha\ g$.

$\alpha\ (l \vee g)$ = **«*a blue* (*flower* or *handkerchief*)»** =

= **«*a blue flower*» or «*a blue handkerchief*»** = $\alpha\ l \vee \alpha\ g$.

Thus, the set of objects can be considered as a vector logical space defined over a scalar field, the elements of which are the qualities of objects.

If we take a set of objects as a scalar field, and a set of qualities of objects as a space of vectors given above it, then, obviously, all axioms of the logical space will also be fulfilled. That is, the set of qualities of objects can also be considered as a vector logical space over a scalar field, the elements of which are objects. Therefore, the formalization of these word-combinations can occur in any direction: both from the main word to the dependent one, and vice versa, from the dependent word to the main one. The considered examples illustrate the case when the dependent word in the phrase is an adjective, and the main one is a noun. However, regardless of which part of the language the main and dependent words in a simple word-combination are given, according to the specified algorithm, this word-combination can be represented by a mathematical formula, namely a formula of the vector logical algebra [2], regardless of the direction of formalization. Also, there are different types of grammatical connection between the main and dependent words in word-combinations. According to the considered scheme of formalization, any grammatical relationship between the main and dependent words can be presented.

Let us now consider the word combination as a part of the sentence, bearing in mind that the word-combination is the result of dismembering the sentence into units that have some meaningful integrity. Recently, this direction has gained popularity in linguistics. In every language there are algorithms for selecting the main members of a sentence [6]. Each sentence describes some relation expressed by a predicate. At the same time, the subject defines the subject, that is, some object of the real world. The predicate, in turn, expresses some feature (action, state, property, quality) of the object described by the subject. But in any case, the relationship between the subject and the predicate, both of which are the main members of the sentence, can be formalized similarly to the example discussed above, regardless of which of these members of the sentence is to be taken as the main word in the studied phrase, and which one is the dependent word.

Summary and conclusions.

From all that has been said, it follows that any word-combination of natural language can be formalized in any direction by means of vector logical algebra. The results obtained on the example of the Ukrainian language can undoubtedly be extended to the group of Slavic languages with a grammar close to Ukrainian. But, as it was shown, similar research methods can be implemented for any other languages. These studies show that by means of vector logical algebra, which is based on the apparatus of algebra of finite predicates, it is possible to formalize an arbitrary syntagm of any natural language.

References:

1. Gvozdinskaya N.A., Dudar Z.V., Shabanov-Kushnarenko Yu.P. (1998). On mathematical description of sense of natural-language texts. *Problems of bionic*. Vol.48, pp. 141 – 149.

2. Якімова Н.А. (2019). *Логічна алгебра: методичний посібник*. Одеса: «Освіта України», 40с.

3. Gvozdinskaya N.A., Dudar Z.V., Poslavskiy S.A., Shabanov-Kushnarenko Yu.P. (1997). On logical spaces. *ACS and automation devices*. Vol. 106, pp. 21 – 30.

4. Yakimova N.A. (2000). Simple word-combination of natural language as a logical algebra formula. *Problems of bionic*. Vol. 52, pp. 111 – 115.

5. Koltzov V.P., Shabanov-Kushnarenko Yu.P. (1990). On the meaningful interpretation of the algebra of ideas. *Problems of bionic*. Vol.45, pp. 10 – 17.

6. Korneychuk T.B., Chudina A.F. (1992). Data system for text analysis in the field of determining the main members of the sentence. *ACS and automation devices*. Vol.98, pp. 76 – 83.

The article has been sent: 12.02.2025.