UDC 658.12

Lyubov Badyorina

METHOD OF GRAMMATICAL STRUCTURE FORMALIZATION OF A NATURAL LANGUAGE

National Aviation University
Kosmonavta Komarova avenue 1, 03680, Kyiv, Ukraine
E-mail: vada@ukr.net

Abstract. *The paper considers the application of mathematical tools for formalizing of grammatical structures to facilitate the Phrases processing by computer means.*
Keywords: algebraic and logical means of natural language constructions, information systems, natural language.

1. Introduction

The development of mathematical modeling of various processes of research, industry and public life is the basis for solving complex problems of design, management and decision making. An important trend of mathematical modeling is the automation of interaction between a human and a computer in the processing of large amounts of hard formalized and structured information (Zaitseva, John 1991).

In this regard, the development of mathematical models and efficient methods and tools for processing of primary information is relevant, such as a natural language, which has not found its final solution yet.

Methods and tools used for constructing the apparatus of mathematical modeling of natural languages can be useful in other information areas which require combining large amounts of data and knowledge into a flexible system for further analysis. An example of such problems is the mathematical modeling of the functioning of complex information systems (Bondarenko et al. 2005).

All attempts to solve the problems of natural language formalization by the middle of the last century had been unable until the appearance of the required mathematical apparatus, i.e. it was suggested to use algebra of finite predicates and predicate operations.

Significant contribution to the creation of mathematical tools of predicative models and methods for in-depth studies and mathematical modeling of discrete processes and objects including natural languages have been made by N. Chomsky (1963), R. Shenk (1980), Y.P. Shabanov-Kushnarenko (2005), M.F. Bondarenko (2005), D.O. Pospelov (1981).

A natural language as a phenomenon of human intellectual activity is a very complicated subject. But having a formal description of a natural language, it is possible to implement on a computer and, thus, give the machine the ability to own its natural language.

Algebra-logical apparatus that can be found in a natural language will empower the developer who tries to create new information technologies. Thus, the conceptual and methodological approach to a natural language (from the mathematical point of view) can perceive it as some kind of algebra and texts as formulas of algebra (Bondarenko et al. 2005).

And the sense (meaning) of thoughts can be expressed in sentences and texts which we are going to consider in terms of their mathematical nature as predicates.

Later, the starting point of our considerations is that thoughts are predicates.

Thus, each sentence is thought as some function with a binary value that specifies a predicate $P(x)=\lambda$. Independent variable x of this function will be variable situation and the dependent is true variable λ.

After the substitution instead of constant variable x of the specific situation $x=a$ the given sentence becomes true ($\lambda=1$) or false ($\lambda=0$). It depends on whether the content of the sentences determines the situation a to which it is referred.

Let's cosider a variable situation $x=(x_1,x_2,...,x_m)$ as a set of subject variables $x_1,x_{2,}...,x_m$. Any permanent situation $x=a$ should be a set of some objects $x_1=a_1$, $x_2=a_2$, $x_m=a_m$.

Thus, each sentence can be expressed by some predicate $P(x_1,x_2,...,x_m)=\lambda$, which represents the dependence of the true variable λ of subject variables $x_1,x_{2,}...,x_m$.

However, any sentence in its natural linguistic form differs from the mathematical formula in such a way that it does not express the entire function $P(x_1, x_2, ..., x_m)$, but only its name. And it's true, because every time when a person makes a particular sentence in accordance to his or her opinion, he or she completes it to the predicate. However, he or she adds to it (as to the name of the predicate), absent substantive variables. Only after this the sentence becomes accessible for understanding. And, on the contrary, turning some thought into a sentence, a person excludes from it substantive variables what can transmit to the others not the very idea, but only its name (Abstracts…2009).

2. Processing of phrases with mathematical methods

Using the methods of composition and decomposition by a set of variables models, being formalized with the help of algebraic predicates, it is possible to handle natural language phrases inflectionally.

The decomposition model $\langle M, P \rangle$ for a set of variables to a set of models

$$I = \{\langle M_{\sigma_1,\sigma_2,\ldots,\sigma_i}, P_{\sigma_1,\sigma_2,\ldots,\sigma_i} \rangle\},$$
$$\sigma_k \in A, k = \overline{1,i}$$

is based on the theorem of decomposition.

Let $x_1, x_2, ..., x_i$ predicate variables and $P(x_1,\ x_2, ..., x_i\ , \ldots\ x_m)$ $\sigma_1, \sigma_2, \ldots, \sigma_i \in A$ are their values, respectively. The predicate corresponding the relation $M_{\sigma_1,\sigma_2,\ldots,\sigma_i}$ has the form

$$M_{\sigma_1,\sigma_2,\ldots,\sigma_i}(x_{i+1}, \ldots, x_m) =$$
$$= M(\sigma_1, \sigma_2, \ldots, \sigma_i, x_{i+1}, \ldots, x_m).$$

The predicate of the model $\langle M_{\sigma_1,\sigma_2,\ldots,\sigma_i}, P_{\sigma_1,\sigma_2,\ldots,\sigma_i} \rangle$ is found by the formular:

$$P_{\sigma_1,\sigma_2,\ldots,\sigma_i}(x_{i+1}, \ldots, x_m) =$$
$$= P(\sigma_1, \sigma_2, \ldots, \sigma_i, x_{i+1}, \ldots, x_m).$$

Of all the obtained relations $M_{\sigma_1,\sigma_2,\ldots,\sigma_i}$ and predicates $P_{\sigma_1,\sigma_2,\ldots,\sigma_i}$ $(\sigma_1, \sigma_2, \ldots, \sigma_i \in A)$, we can form models $\langle M_{\sigma_1,\sigma_2,\ldots,\sigma_i}, P_{\sigma_1,\sigma_2,\ldots,\sigma_i} \rangle$, forming a system

$$I = \{\langle M_{\sigma_1,\sigma_2,\ldots,\sigma_i}, P_{\sigma_1,\sigma_2,\ldots,\sigma_i} \rangle\},$$
$$\sigma_k \in A,\ k = \overline{1,i}\,.$$

It should be noted that it is not necessary to lay the mathematical model by the first *i*-variables. This method is applied to the predicate defined on the whole space U^m.

Thanks to the models decomposition with the set of variables and mathematical models of natural language it is possible to determine the variables being necessary for further modeling (Abstracts…2009).

We have modified mathematical predicate models of nouns and adjectives declination:

$\alpha_{ap} = (M_{ap}, P_{ap})$ – mathematical model of declination of possessive adjectives,

$\alpha_n = (M_n, P_n)$ – mathematical model of regular nouns declination,

$\alpha_a = (M_a, P_a)$ – mathematical model of declination of complete unpossessive adjectives, where M_i – model carrier, P_i – predicate model $i = \{a, n, ap\}$.

Using mathematical models of predicate processing of possessive adjectives we can make logical calculations.

A word form of each word is always defined and characterized by a single root end. There are six attributes that uniquely determine the choice of the required word form of possessive adjectives. For the formal process description of the inflectional processing of possessive adjectives for each grammatical and lexicographical features we have introduced a substantive variable:

x_1 – kind of word forms,

x_2 – number of word forms,

x_3 – case of word forms,

x_4 – a sign of materiality of word forms,

x_5 – an archaic sign of word forms;

c – a suffix of a word basis.

Substantive variables were introduced to indicate the type of context influence as declination, word endings and soft base of words:

r – type of influence of the context;

s – type of words declination;

z – words ending;

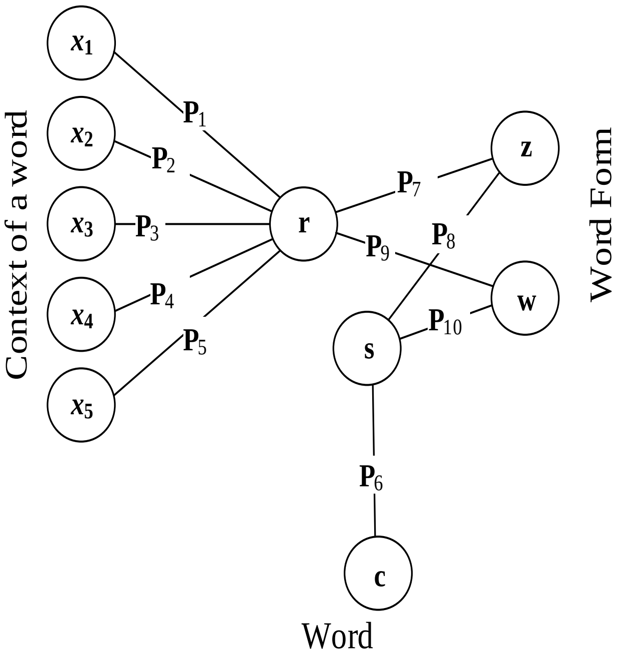
w – a sign of soft base of a word.

Sets $M_{\alpha p}$ and predicate $P_{\alpha p}$ of a mathematical model that describes the structure of declination of possessive adjectives are as follows:

$$M_{\alpha p} = M_{x_1} \times M_{x_2} \times M_{x_3} \times M_{x_4} \times$$
$$\times M_{x_5} \times M_r \times M_c \times M_s \times M_z \times M_w,$$

$$P_{\alpha p}(x_1, x_2, x_3, x_4, x_5, r, c, s, z, w) =$$
$$= P_1(x_1, r) \wedge P_2(x_2, r) \wedge P_3(x_3, r) \wedge P_4(x_4, r) \wedge$$
$$\wedge P_5(x_5, r) \wedge P_6(c, s) \wedge P_7(r, z) \wedge$$
$$\wedge P_8(s, z) \wedge P_9(w, r) \wedge P_{10}(w, s) .$$

Graphic image of Phrases processing is a scheme which is shown in Figure.



Processing of phrases

3. Conclusions

Development of methodological tools of mathematical modeling will improve the efficiency and quality of information processing in natural language systems.

Thus, algebra is considered as a research tool, not as a subject. It is an effective means of mathematical representation of information and solving logic problems for empowering and improving the efficiency in processing of natural language information. Due to high-quality processing of natural information it is possible to simplify the process of formalizing linguistic units of information systems and training of primary data to the stage of program implementation tasks.

References

Abstracts of the IV International Conference “Horizons of applied linguistics and linguistic technologies” (“MegaLing-2009”). Kyiv, UMIF Academy of Sciences of Ukraine, 2009. 104 p. (in Ukrainian).

[*Тези* доповідей IV Міжнародної конференції «Горизонти прикладної лінгвістики і лінгвістичних технологій» (“MegaLing-2009”). – Київ: УМІФ НАН України, 2009. – 104 с.]

Bondarenko, M.F.; Dudar, Z.; Shabanov-Kushnarenko, Y.P.; Chikin, V.; Protsay, N.T.; Cherkashin, V. 2005. *Algebra of predicatives and predicate operations*. Radioelectronics and Computer Science. Kharkov, KNURE. N 1: 80–86 (in Ukrainian).

[*Алгебра* предикатів и предикатних операцій / М.Ф. Бондаренко, З.В. Дударь, Ю.П. Шабанов-Кушнаренко, В.А. Чикина, Н.Т. Процай, В.В. Черкашин // Радіоелектроніка та інформатика. – Харків: ХНУРЕ, 2005. – № 1. – С. 80–86.]

Chomsky, N.; Miller, G.A. 1963. *Introduction to the formal analysis of natural languages*. Handbook of Mathematical Psychology. New York, Wiley. Vol. 2: 269–322.

Pospelov, D.A. 1981. *Logiko-linguistic models are in control system*. Monograph. Moscow, Sciences. 232 p. (in Russian).

[Поспелов Д.А. Логико-лингвистические модели в управлении: монография / Д.А. Поспелов. – Москва: Наука, 1981. – 232 с.]

Programmed Competencies and cybernetic teaching machines. Collected articles ed. A.I. Shestakov. Moscow, Sov. Radio. 1963. 247 p. (in Russian).

[*Программированное* обучение и кибернетические обучающие машины: сборник статей / под. ред. А.И. Шестакова. – Москва: Советское радио, 1963. – 247 с.]

Shenk, P. 1980. *Treatment of conceptual information.* Monograph. Moscow. 360 p. (in Russian).

[Шенк Р. Обработка концептуальной информации: монография / Р. Шенк. – Москва, 1980. – 360 с.]

Zaitseva, L.; Zakis, John D. 1991. *Course Development for Tutoring and Training Systems in Engineering Education*. Global J. of Engng. Educ. Vol. 1.

Received 17 January 2013.

Л.М. Бадьоріна. Метод формалізації граматичних структур природної мови
Національний авіаційний університет, проспект Космонавта Комарова, 1, Київ, Україна, 03680
E-mail: vada@ukr.net

Розглянуто застосування математичного апарата для формалізації граматичних структур з метою полегшення обробки словосполучень комп’ютерними засобами.

Ключові слова: алгебро-логічні засоби, інформаційні системи, конструкції природної мови, природна мова.

Л.Н. Бадёрина. Метод формализации грамматических структур естественного языка
Национальный авиационный университет, проспект Космонавта Комарова, 1, Киев, Украина, 03680
E-mail: vada@ukr.net

Рассмотрено применение математического аппарата для формализации грамматических структур с целью более качественной обработки словосочетаний компьютером.

Ключевые слова: алгебро-логические средства, естественный язык, информационные системы, конструкции естественного языка.

Badorina Lyubov. Candidate of Engineering. Doctoral student of the Department of Air Navigation Systems, National Aviation University, Kyiv, Ukraine.
Education: Faculty of Automation and Computer Engineering, National Aviation University, Kyiv, Ukraine (2004).
Researches area: intelligent information technologies, computer technologies.
Publications: 27.
E-mail: vada@ukr.net