МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Филологический факультет
Кафедра русского языка
Лаборатория общей и сибирской лексикографии
ИНСТИТУТ РУССКОГО ЯЗЫКА ИМ. В.В. ВИНОГРАДОВА РАН
ГОСУДАРСТВЕННЫЙ ИНСТИТУТ РУССКОГО ЯЗЫКА ИМ. А.С. ПУШКИНА
ПОД ЭГИДОЙ РОССИЙСКОГО ОБЩЕСТВА ПРЕПОДАВАТЕЛЕЙ
РУССКОГО ЯЗЫКА И ЛИТЕРАТУРЫ (РОПРЯЛ)

ЛЕКСИКОГРАФИЯ ЦИФРОВОЙ ЭПОХИ

Сборник материалов Международного симпозиума
(24–25 сентября 2021 г.)

Томск
Издательство Томского государственного университета
2021

DOI: 10.17223/978-5-907442-19-1-2021-118

*В.А. Мишланов¹, Л.А. Каджая^{1, 2},
В.А. Салимовский¹, И.В. Смирнов³
Valery A. Mishlanov, Liudmila A. Kadzhaya,
Vladimir A. Salimovskiy, Ivan V. Smirnov*

**Словарь предикатных слов как инструмент
автоматической обработки текстов методом
реляционно-ситуационного анализа***
**Predicate word dictionary as a tool for automatic text processing
by situational-relational analysis**

¹ Пермский гос. национальный исследовательский университет, Пермь –
Perm State University

vmishlanov@yandex.ru, salimovskiy@rambler.ru

² Шаньдунский университет, Шаньдун (Китай) – Shandong University, Weihai, China

³ Институт проблем искусственного интеллекта ФИЦ «Информатика
и управление» РАН, Москва – Russian Artificial Intelligence Research Institute,
FRC CSC RAS, Moscow
ivs@isa.ru kadzhaya1@icloud.com

Аннотация. Обсуждаются вопросы совершенствования структуры словаря предикатных слов, используемого при решении задач приобретения знаний и анализа текстов. Раскрывается принцип открытой архитектуры словаря, учитывающий стилевую дифференциацию речи и предполагающий описание подсистем предикатных слов, функционирующих в отдельных речевых разновидностях.

Summary. This paper discusses the issues of improving a predicate word dictionary structure that is used in solving problems of knowledge acquisition and text analysis. The principle of open dictionary architecture is shown. It takes into account the stylistic differentiation of speech and involves the description of predicate word subsystems functioning in separate speech varieties.

Ключевые слова: реляционно-ситуационный анализ, предикатное слово, словарь предикатных слов, функциональная дифференциация языка

Keywords: relational-situational analysis, predicate word, text, functional differentiation of language

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-07163 мк.

Знаменуя современный этап в развитии науки и техники, цифровизация, сопровождающаяся все большим проникновением искусственного интеллекта в самые разные сферы деятельности, оказывает определенное влияние на социальные отношения, порождая новые проблемы гуманитарного характера. В частности, настоящим вызовом общественным институтам становится проблема сохранения человеческого капитала в новых социальных условиях, когда человек все чаще оказывается в психологически дискомфортной ситуации нового профессионального выбора. Принятие решения в кризисной ситуации, разумеется, остается задачей самого индивида, но проблема поиска и обработки необходимой для этого информации решается ныне методами искусственного интеллекта, что существенно повышает значимость теоретических исследований, направленных на оптимизацию программ автоматического анализа текстов.

Исследования в области автоматического анализа занимают важное место в проблематике специфического направления в современной математике, которое метонимически было обозначено словосочетанием *искусственный интеллект* (о предмете и методах этой дисциплины см. [2]). Извлечение информации из текста с помощью компьютерных программ обеспечивается ныне на всех языковых уровнях, в то же время алгоритмы содержательного анализа высказываний, как правило, недостаточны не только для «чтения» имплицитных смыслов, которые в реальной коммуникации без труда фиксируются адресатом (благодаря особым компетенциям, отсутствующим пока у искусственного интеллекта – [2. С. 9]), но и для разрешения полисемии языковых выражений.

Для преодоления этих трудностей лингвисту необходимо выявить максимально широкий круг речевых признаков воплощаемых в тексте смыслов, при этом в оптимальном варианте во внимание должен быть принят фактор речевой системности: алгоритмический выбор между значениями в условиях полисемии проводится с учетом принадлежности текста к определенному дискурсу и жанровой модели. Все формальные маркеры представляются в виде «лингвистических шаблонов», которые далее используются в программных средствах содержательного анализа текстовых массивов.

Автоматический анализ, имеющий целью идентификацию текстов по заданным параметрам, оптимально основывать на такой модели, которая допускает использование единого инструментария, модифицируемого в той или иной мере в зависимости от речежанровых и дискурсивных особенностей анализируемых текстов. Одной из перспективных является, на

наш взгляд, модель, основанная на методе реляционно-ситуационного анализа (РСА) [3] и давшая хорошие результаты при разработке алгоритмов распознавания воплощаемых в тексте ментальных действий [4]. Означенный метод в лингвистическом плане опирается на концепцию коммуникативного синтаксиса Г.А. Золотовой, в которой поверхностно-синтаксическим формам русского языка (синтаксемам) ставятся в соответствие определенные семантические роли, отражающие структуру описываемой ситуации. «Полномочным представителем» ситуации является, как известно, предикат, поэтому основным инструментом РСА-метода оказывается формализованный словарь предикатных слов. В оптимальном варианте такой словарь должен иметь открытую архитектуру, позволяющую не только расширять, по мере необходимости, список предикатных слов, но и углублять его типологическую структуру – за счет ветвления семантических типов, образующих инвариантную типологическую схему. Данная особенность словаря обусловлена тем, что язык как система реализуется в ряде функциональных стилей, представленных множеством речевых жанров. Дифференциация значений и форм языкового знака (в частности, предикатного слова и его аргументов) обусловлена тем, что каждая из разновидностей речи, в соответствии с характерными для нее задачами общения, реализует лишь часть лексического и грамматического потенциала языка [2].

Полисемия, характерная для большинства предикатных слов, снимается перечислением в словарной статье уникального для того или иного значения предиката набора синтаксем (аргументов), представленных именами определенных семантических разрядов во всех возможных для текстов данного жанра и стиля поверхностно-синтаксических формах. Так, глагол *быть* выступает в русском тексте как знаменательный предикат в бытийном или посессивном значении, а также как связка при именном сказуемом в характеризующих и локативных высказываниях (*Она была в белом платье / в театре*), поэтому в словаре он описывается как минимум в четырех словарных статьях, дифференцируясь не только сегментными, но и супraseгментными маркерами (ср.: *В селе Ольховка есть действующая церковь / Действующая церковь есть в селе Ольховка*).

Литература

1. Кожина М.Н. Речеведение. Теория функциональной стилистики. М.: Флинта; Наука, 2014. 624 с.
2. Осипов Г.С. Методы искусственного интеллекта. М.: Физматлит, 2011. 296 с.

3. Осипов Г.С., Смирнов И.В., Тихомиров И.А. Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. 2008. № 2. С. 3–10.

4. Салимовский В.А., Девяткин Д.А., Каджая Л.А., Мишланов В.А. Автоматическое распознавание ментальных действий, реализуемых в научных эмпирических текстах // Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки. 2019. Т. 10, № 3. С. 74–88.

DOI: 10.17223/978-5-907442-19-1-2021-119

Е.А. Неведова, Е.А. Ковригина
Elena A. Nefedova, Elena A. Kovrigina

**О проекте электронного идеографического словаря
говоров архангельского региона
About project of an electronic ideographic dictionary
from Arkhangelsk dialect**

Московский государственный университет им. М.В. Ломоносова, Москва –
Moscow State University, Moscow
eanefedova@gmail.com, el.kovrigina7@yandex.ru

Аннотация. «Идеографический словарь архангельских говоров» призван отразить как языковые особенности исследуемого региона, так и народные представления о важнейших сферах жизни и деятельности человека. Основная задача Словаря – представить все разнообразие и богатство традиционной крестьянской культуры и языка жителей Русского Севера. Электронный формат Словаря облегчает доступ к нему не только специалистов в области диалектологии, этнолингвистики, фольклористики и культурологии, но и более широкого круга пользователей.

Summary. The “Ideographic Dictionary of Arkhangelsk Dialects” is intended to reflect both the linguistic characteristics of the region and popular ideas about the most important spheres of human life and activities. The main task of the Dictionary is to present all the diversity and richness of the traditional peasant culture and language of the inhabitants of the Russian North. The electronic format of the Dictionary facilitates access to it not only for specialists in the field of dialectology, ethnolinguistics, folkloristics and cultural studies, but also for a wider range of users.

Ключевые слова: лексикография, архангельские говоры, идеографический словарь

Keywords: lexicography, dialects of the Arkhangelsk region, ideographic dictionary