



DOI 10.22363/2312-9220-2019-24-1-151-155

Филологи и математики: организация взаимодействия (о перспективах междисциплинарного подхода к созданию компьютерных систем ситуационного моделирования для решения проблем медиа)

В.П. Кальян, Ш.Н. Кадырова

Российский университет дружбы народов
Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 10, корп. 2

Анализ вариаций экранного образа героев и ситуативной динамики в кино-сюжетах, оценка перспектив медиарынка и медиапланирование с помощью систем искусственного интеллекта — примеры типовых задач, которые нужно решить в самое ближайшее время. Задачи настолько «перезрели», что инициатива их решения исходит уже от рядовых работников СМИ и преподавателей филологических факультетов.

«Ситуация, язык, речь. Модели и приложения (“Situation, Language, Speech. Models & Applications” — SLS 2018)» — так называлась лингвистическая секция в рамках 12-й Международной конференции «Интеллектуализация обработки информации» (Италия), организованной Федеральным исследовательским центром «Информатика и управление» (РАН) совместно с научно-координационным центром «Цифровые методы анализа данных» и Лабораторией машинного интеллекта МФТИ. Среди инициаторов и участников секции — преподаватели филологического факультета РУДН.

На конференции была сделана попытка объединить ранее изолированные исследования в сфере распознавания речи и анализа текстов. За общим круглым столом собрались специалисты трех направлений — лингвисты, математики и теории журналистики, занятые поиском семантических связей и их коммуникативных значений в тексте.

Удалось договориться о расширении взаимодействия в области моделирования ситуации по тексту и речи, которые приведут к созданию принципиально новых систем для широкого круга применений — от университетского образования до автоматизированных динамических новостных информационных систем, востребованных медиарынком, или систем машинного анализа и прогнозирования развития ситуаций, в которых заинтересованы крупные государственные заказчики.

Форуму предшествовали острые дискуссии с идеологами математических и лингвистических конференций в России и за рубежом. Математики были готовы извлекать знания из данных об экономических, логистических, исторических и любых других процессах, литературных произведениях, данных о здоровье людей и методах лечения из историй их болезней, данных о судебных и следственных

делах, процессах, развивающихся в СМИ, о том, что происходит в социальных сетях, — данных о чем угодно, используя математические методы, но не привлекая филологические, лингвистические, культурологические и другие необходимые знания из предметных областей. При этом изыскания в области компьютерной лингвистики основательны, проекты обладают значительным потенциалом, который, к сожалению, не востребован широким кругом филологов-журналистов, поскольку нет налаженной практики междисциплинарных встреч и не разработаны компьютерные системы решения задач анализа текстов для широкого круга проблем филологов и журналистов.

Очевидно, что специалисты, исповедующие сугубо технические методы, осознают некоторую исчерпанность своих подходов. На совместных конференциях лингвистов и математиков в последнее время встречались доклады о тональности текста, о поисках в речи иронии, сарказма, распознавании модальности эмоций и даже юмора. Однако радикальной научной повестки дня, расширяющей круг возможных применений усилий математиков и лингвистов, пока не намечалось, и это имеет довольно глубокие причины.

Первая из них — изоляционизм, свойственный любой технической предметной области, крепко привязанной к определенному инструментарию. Вторую причину можно увидеть в узкой специализации образования и связанной с этим недостаточности эрудиции, часто определяющей глубоко мотивированную непроходимость барьера между гуманитарным и техническим знанием. Между тем число специалистов, преодолевших этот барьер в международном научном сообществе, превысило «критическую массу».

Что касается полемики внутри отечественного журналистского научного сообщества, она уже почти десять лет ведется вокруг идеи применения математического аппарата для анализа СМИ и в связи с этим — об иерархии наук в изучении медиатекстов. В частности, поводом для споров являются публикации, посвященные кибернетике и информатике как технологическому базису новой науке о медиа — медиалогии.

Прошедшая в Италии конференция стала своеобразным объединяющим ответом на проблемную ситуацию: она была посвящена новой парадигме в изучении машинного понимания смысла текста и спонтанной речи. В центре внимания оказалось то, что упускали как математики, так и лингвисты — последняя часть семиотической триады — прагматика. Здесь понадобился весь арсенал теоретических наработок филологов и журналистов, который раньше имел по большей части методическое значение для обучения студентов разных специальностей филологических факультетов. Кроме того, нужен был широкий обзорный охват подходов и инструментов извлечения знаний из данных именно со стороны практиков, занимающихся журналистикой данных.

Традиционным гуманитариям для вхождения в новое междисциплинарное направление нужно было совсем немного — ознакомиться с опытом работ «смежников», иметь представление об инструментарии лингвистов и математиков, ведущих исследования в этой области. Необходимо было получить представление о корпусных инструментах в грамматических исследованиях русского языка, автоматической разметке семантических ролей в тексте, вероятностном тематическом моделировании.

Докладчики готовились с полной ответственностью, обновляя и актуализируя собственные знания. Знакомство с методологией «смежников» позволило авторам докладов увидеть как производительность применяемого инструментария, так и некоторые ограничения, узость поставленных задач, и, определив пробелы в существующих разработках, найти свое место в творческом процессе формирования новой научной отрасли, встроиться в тематику междисциплинарных исследований, представленных на конференции, понять ее дух, выступить с содержательным докладом.

В докладе «Морфология ситуации. Структура момента и модели динамики» (В.П. Кальян, РУДН) был представлен опыт разработки системы построения моделей ситуации по тексту с использованием методов математической и лингвистической текстологии. Была поставлена задача построения языка описания широкого класса ситуаций с выявлением их универсальной морфологии. Как один из первых подходов к ее решению оценивались документальные базы предметных областей с хорошей терминологической оснащенностью. По группе узких специальных текстов, таких как, к примеру, медицинские словари, выявлялись именованные сущности, имеющие значения морфем описания нормативных ситуаций в этих областях. Полученный инвентарь ролей был применен к широкому классу ситуаций. По исходным текстам и экспертным разметкам семантических ролей с использованием средств корпусной лингвистики восстанавливались их свойства, качества, взаимоотношения между сущностями, отслеживались этапы изменения ролей, строились модели ситуации в виде последовательности таблиц и графов отношений, выявлялась динамика изменения ситуации.

Большое внимание уделялось возможности описания интенций субъектов. Было построено математическое описание потенциалов движения к цели, оценки сопротивления среды в виде противодействия субъектов и подсчета вероятности достижения или не достижения цели, наступления тех или иных событий. Результат был представлен в виде проекта мультимодальной экспертной системы построения динамики ситуационных зависимостей и оценки вариантов развития событий. Описанный в докладе подход даёт возможность построения по текстам и речи модели не только ситуации, но и явлений языка, мышления, восприятия, культуры. Это открывает новые перспективы для решения широкого круга задач — от построения компьютерных систем более точного литературного перевода текста с одного языка на другой до выявления причин возникновения описанных в текстах драматических ситуаций.

В продолжение темы участникам секции был представлен доклад «Бедная “русская мысль”: получим ли мы автоматизированное извлечение нарратива? Краткий обзор русскоязычных приложения» (Ш.Н. Кадырова, РУДН). Был описан круг проблем, с которыми столкнулся коллектив исследователей кафедры массовых коммуникаций и кафедры теории и истории журналистики в работе над семантико-синтаксическим анализом материалов телевизионной авторской программы обозревателя Д. Киселева. Дело в том, что исследования трансмедийного сторителлинга и нарратологии в аспекте новостных историй, которыми занимается группа, показали эффективность изучения текстов СМИ с точки зрения драматургии. Была проведена ручная разметка семантических ролей героев дела о Скри-

палях в изложении Киселева с опорой на типологии персонажей сначала В. Проппа, а затем Ч. Филмора. В результате авторы и пришли к выводу о необходимости ресурса, который осуществлял бы такого рода разметку автоматически.

Автоматизированная разметка текста — это одно из актуальных и востребованных направлений междисциплинарного сотрудничества математиков и лингвистов. Однако мониторинг имеющихся в настоящее время ресурсов на базе существующих корпусов русского языка показал, что ни один из них не позволяет произвести полноценный анализ семантических ролей героев новостных повествований. Таким образом, участникам конференции было представлено техническое задание на создание программы, где комплексная типология на основе разметок Проппа и Филмора служит базовыми параметрами при автоматизированном семантико-синтаксическом анализе корпусов новостных текстов.

Ответом стало предложение о сотрудничестве от Лаборатории компьютерной лингвистики ИСА ФИЦ ИУ РАН, где уже несколько лет при поддержке РФФИ ведется работа по созданию «Машины РСА» — инструмента лингвостатистических и корпусных исследований.

Вот темы некоторых других докладов представителей филологического факультета РУДН: «Феномен интерязыка машинного текста» (С.А. Дерябина), «Клиповое мышление: реальность и перспективы» (И.И. Митрофанова). Доклады представителей других подразделений РУДН и других научных организаций: «Что такое ситуация?» (А.В. Глазков, МПГУ), «Семантика текста как модель ситуации», (А.А. Харламов, ИВНД и НФ РАН), «Кластеризация документов с помощью нейронных сетей» (Н.Т. Габдрахманова, МИ РУДН), «Сетевой дискуссионный текст как источник информации о сценарии» (И.В. Смирнов, ИСА РАН, РУДН).

Видеозаписи фрагментов ключевых докладов и дискуссий доступны по ссылке: <https://youtu.be/saqDjy0WdZ8>

Главный итог — перспектива совместных проектов филологического факультета РУДН с Лабораторией машинного интеллекта МФТИ, Лабораторией компьютерной лингвистики ИСА ФИЦ ИУ РАН, корпорацией «Яндекс», компанией «Форексис». Активно распространяет и популяризирует итоги конференции за рубежом Российский центр науки и культуры в Риме.

© Кальян В.П., Кадырова Ш.Н., 2019



This work is licensed under a Creative Commons Attribution 4.0 International License

История статьи:

Дата поступления в редакцию: 21 октября 2018

Дата принятия к печати: 31 октября 2018

Для цитирования:

Кальян В.П., Кадырова Ш.Н. Филологи и математики: организация взаимодействия (о перспективах междисциплинарного подхода к созданию компьютерных систем ситуационного моделирования для решения проблем медиа) // Вестник Российского университета дружбы народов. Серия: Литературоведение. Журналистика. 2019. Т. 24. № 1. С. 151—155. DOI 10.22363/2312-9220-2019-24-1-151-155

Сведения об авторах:

Кальян Виктор Петрович, преподаватель-исследователь кафедры массовых коммуникаций филологического факультета Российского университета дружбы народов. *Контактная информация*: e-mail: vkalyan@mail.ru

Кадьрова Шуанат Набиевна, кандидат филологических наук, старший преподаватель кафедры теории и истории журналистики филологического факультета Российского университета дружбы народов. *Контактная информация*: e-mail: kadyrova_shn@pfur.ru

**Philologists and mathematics: setting-up the interaction
(on the prospects of the interdisciplinary approach
in the development of situation modeling
computer systems for solving media problems)**

V.P. Kalyan, Sh.N. Kadyrova

Peoples' Friendship University of Russia (RUDN University)
10 Miklukho-Maklaya St., bldg. 2, Moscow, 117198, Russian Federation

Article history:

Received: 21 October 2018

Revised: 25 October 2018

Accepted: 31 October 2018

For citation:

Kalyan V.P., Kadyrova Sh.N. (2019). Philologists and mathematics: setting-up the interaction (on the prospects of the interdisciplinary approach in the development of situation modeling computer systems for solving media problems). *RUDN Journal of Studies in Literature and Journalism*, 24(1), 151–155. DOI 10.22363/2312-9220-2019-24-1-151-155

Bio notes:

Kalyan Victor Petrovich, Lecturer, Researcher, Department of Mass Communications, Philological Faculty, Peoples' Friendship University of Russia (RUDN University). *Contacts*: vkalyan@mail.ru

Kadyrova Shuanat Nabievna, Candidate of Philology, Assistant Professor of the Department of Theory and History of Journalism, Peoples' Friendship University of Russia (RUDN University). *Contacts*: e-mail: kadyrova_shn@pfur.ru