

**МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ ПО КОМПЬЮТЕРНОЙ
ЛИНГВИСТИКЕ И ИНТЕЛЛЕКТУАЛЬНЫМ ТЕХНОЛОГИЯМ
«ДИАЛОГ 2009»
(Бекасово, 27 – 31 мая 2009 г.)**

27 – 31 мая в Бекасово (Подмосковье) прошла очередная международная конференция «Диалог 2009», организованная Институтом лингвистики РГГУ, Институтом проблем информатики РАН, Институтом проблем передачи информации РАН, компанией АBBYU, компанией Яндекс, РосНИИ искусственного интеллекта и филологическим факультетом МГУ.

В конференции приняли участие не только сотрудники академических институтов и учебных заведений, но и представители различных фирм. В программу были включены 93 доклада, прошедшие предельно жесткий отбор. Доминантой «Диалога 2009» была объявлена проблема создания и использования лексических баз данных. Стоит отметить исключительный уровень организации конференции. Строго выдерживался регламент, согласно которому каждому докладчику давалось полчаса на доклад и дискуссию, и докладчик мог распорядиться этим временем по своему усмотрению; в основном представление доклада занимало 20 минут и 10 – ответы на вопросы. Каждый участник имел на руках расписание работы секций с указанием точного времени каждого выступления, что позволяло слушателям выбирать то, что их интересует.

27 мая состоялось первое пленарное заседание «Типология. Интернет-лингвистика», которое открыл А.Е. Кибрик с докладом «К проблеме вариативности языка: метод многофакторного исчисления 2-го порядка». Исходным материалом послужили сочинительные конструкции и в частности нулевая анафора в 23 родственных дагестанских языках и диалектах. В исследуемой выборке не обнаружилось ни одной пары идиомов с идентичными сочинительными конструкциями. Примечательно, что языковой материал был собран 30 лет назад и тщательно перепроверялся, но в то время уровень типологии не позволял дать объяснения этой кажущейся хаотичности в выборе синтаксических средств. А.Е. Кибрик предложил многофакторное исчисление 2-го порядка, т.е. не на уровне самих конструкций, а на уровне отдельных параметров, общий набор и иерархическая организация которых различны в исследуемых языках. Доклад Анны А. Зализняк «О понятии семантического перехода» представлял проект «Каталог семантических переходов в языках мира», который ведется на протяжении семи лет коллективом авторов. Под семантическим переходом понимается факт совмещения в пределах одного слова двух разных значений – в синхронии или в диахронии. Живой интерес вызвал доклад Д.О. Добровольского и И.Б. Левонтиной «Русское *нет*, немецкое *nein*, английское *no*: сопоставительное исследование семантики на базе параллельных корпусов». Авторы сосредоточились на таких контекстах, где наблюдаются расхождения в употреблении данных единиц; особенно примечательны случаи, когда на месте русского *нет* в соответствующих текстах на других языках отсутствует какая бы то ни было единица. Далее позвучал доклад М.А. Кронгауза «Народ против Яндекса», в котором содержалась критика решений вопросов орфографии в поисковой системе

Яндекса (на следующий день во время круглого стола «Поисковые системы и коррекция грамотности» представители Яндекса отвечали на замечания и сообщения участников конференции). Завершил заседание доклад В.И. Беликова и М.В. Ахметовой «Статистическая оценка функциональных свойств лексики по материалам Интернета». Блоги были рассмотрены как лингвистический источник, позволяющий оценить синхронное состояние лексики, её устаревание, частотность, стилистические характеристики, территориальную принадлежность, распределение по возрастным группам, причём полученная картина отличается от той, которая представлена в языке СМИ, официальных и художественных текстах. К сожалению, с некоторых пор Яндекс отказался от предоставления возможности поиска по возрасту, полу, регионам.

28 мая ряд докладов утреннего пленарного заседания был посвящен семантике прилагательных: доклады М.В. Яворской и И.В. Азаровой (Санкт-Петербург) «Структура атрибутивных значений в тезаурусе Russnet (на материале перцептивных прилагательных)», Б.Х. Парти (Амхерст, США) «Динамический аспект семантики прилагательных», Е.В. Рахилиной, О.С. Карповой и Т.И. Резниковой «Модели семантической деривации многозначных качественных прилагательных: метафора, метонимия и их взаимодействие». Р.И. Розина в докладе «*Так называемый*: семантика вводных метаязыковых оборотов» рассмотрела эволюцию оборотов *так называемый*, *короче говоря*, *проще говоря*, *грубо говоря* и т.п. на протяжении XIX – XX вв.; в докладе Е.В. Падучевой «Посессивы и имена способа действия» рассматривались конструкции типа *его исполнение Шопена* в сопоставлении со структурами типа *исполнение им Шопена*; В.Ю. Апресян в докладе «Семантические источники уступительности» рассмотрела уступительные лексемы *только* и *правда*. Б.Л. Иомдин выступил с ярким докладом «Терминология быта. Поиски нормы», в котором рассматривались названия бытовых артефактов *рюмка – бокал – фужер, бумажник – кошелек – портмоне и свитер – джемпер – пуловер – кофта – водолазка*. Толкования данных лексем в различных словарях практически несводимы и при этом расходятся с официальной номенклатурой, представленной в нормативных документах, и данные обоих типов источников не совпадают с реальным узусом, который был исследован на основе текстового и мультимедийного поиска в Интернете и тоже обнаруживает полный разноречивый; также использовались опросы носителей и различные эксперименты. Докладчик предложил свой проект словаря-тезауруса бытовой лексики, который должен включать и предпочтения узуса, и региональные различия, и особенности социолектов, а также данные нормативных документов, если они существуют.

Дневное пленарное заседание было посвящено лингвистическим ресурсам. И.А. Большаков выступил с докладом «КроссЛексика – большой электронный словарь сочетаний и смысловых связей в русском языке». Словарь отражает не только подчинительные, но и устойчивые сочинительные связи (*автобусы и троллейбусы, быть или не быть*). Доклад С.В. Кодзасова, А.В. Архипова, Л.М. Захарова и О.Ф. Кривновой был посвящен базе данных «Интонация русских повествовательных текстов». В докладе В.Г. Диконова и И.М. Богуславского «Универсальный словарь концептов» был представлен совместный российско-испанско-французский проект семантического языка-посредника. Концепт по-

нимается как абстрактная семантическая единица, совпадающая с отдельными значениями слов, выделяемыми толковыми словарями. О.Н. Ляшевская и Ю.Л. Кузнецова в докладе «Русский фреймнет: к задаче создания корпусного словаря конструкций» представили свой оригинальный проект, не повторяющий английский FrameNet. Завершил пленарное заседание обзорный доклад Н.В. Лукашевич «WordNet: критика, развитие, приложения». В числе отмеченных недостатков была упомянута слишком большая детализация значений; другой недостаток, свойственный вообще многим словарям – существительные и соответствующие глаголы описываются совершенно по-разному (как правило, разными людьми), и поэтому получают отдельные синсеты для одной сущности. Наконец, существует так называемая «теннисная проблема»: невозможно собрать всю лексику, относящуюся, например, к теннису, так как теннисный инвентарь классифицирован как «Артефакты», корты как «Местоположения», удары как «Действия» и т.д. Слушатели (несомненно, имеющие опыт работы с этой и другими системами) в целом высказывались о WordNet'e весьма критически. Теме лингвистических ресурсов был также посвящен круглый стол следующего дня.

29 мая заседания проходили по секциям. Работу секции «Модели общения» открыл доклад С.И. Переверзевой «Невербальный коммуникативный акт утешения: материал к построению словаря невербальных коммуникативных актов». Словарь должен содержать также и речевые акты с учетом голосовых и просодических характеристик и зоны жестового сопровождения – обязательного или высоковероятного. В докладе был представлен как языковой материал, так и видеофайлы, иллюстрирующие коммуникативный акт утешения – произведения живописи, фрагменты мультфильмов. В ходе обсуждения была отмечена необходимость учитывать этнокультурную специфику воплощения коммуникативных актов. Г.Е. Крейдлин в докладе «Невербальное поведение людей разных культур в диалоге: финская и русская жестовые системы» сосредоточился на финских прерванных жестах (например, когда человек протягивает руку для приветствия, но вдруг внезапно её убирает, прежде чем собеседник протянет свою) и на специфическом финском жесте *katseluasento* (букв. «увидел море»), который сигнализирует о сосредоточенности на том, что человек увидел своим внутренним взором. Е.А. Гришина в докладе «К вопросу о соотношении слова и жеста (вокальный жест *O* в устной речи)» на материале Мультимедийного русского корпуса (МУРКО), содержащего не только расшифровку устных текстов, но и звуковую дорожку, а для фильмов и видеоряд, отметила три типа употребления вокального жеста *O*: дейктическое, эмоциональное и физиологическое (*O* как возглас) – и указала для каждого типа особенности произношения и список значений. П.М. Котта Рамузино (Милан, Италия) в докладе «Непосредственный институциональный диалог. Опыт прямой линии с президентом В.В. Путиным. Дискурсивные стратегии» обратилась к материалу самых первых прямых линий 2002 и 2003 гг., когда, по мнению докладчицы, несмотря на явную асимметричность участников, ещё оставалась некоторая свобода и для «наивных интервьюеров», которые, однако, прибегали к различным стратегиям смягчения: осуждение под видом вопроса, перенос ответственности на других политиков, критика государственных органов, но не президента, модализация высказывания (*похо-*

же, кажется, на наш взгляд), что в какой-то степени снимает ответственность с адресата. В последующие годы преобладающей стала стратегия на создание эмпатии с собеседником. А.А. Котов в докладе «Паттерны эмоциональных коммуникативных реакций: проблема создания корпуса и перенос на компьютерных агентов» привлек материал видеозаписи поведения студентов во время передачи экзаменов в РГГУ (запись проводилась с согласия студентов). Разумеется, материал был воспринят слушателями (и зрителями) с большим оживлением. Завершил утреннее заседание доклад Е.Я. Шмелевой и А.Д. Шмелева «Вариативность, продолжение и серийность анекдотов: проблемы построения базы данных».

Секция «Просодия» открылась докладом М.Л. Палько «Просодия обращений в немецком языке в сопоставлении с русским». Доклад Т.Е. Янко «Русские обращения: словарная информация и вокативные конструкции» касался способности слов употребляться в качестве обращений. В.И. Подлеская и А.А. Кибрик в докладе «Дискурсивные маркеры в структуре устного рассказа; опыт корпусного исследования» на материале устного корпуса «Рассказы о сновидениях» рассмотрели дискурсивные маркеры *ну* и *вот* с учетом их фонетических характеристик и роли в организации дискурса. С.Б. Степанова и А.И. Рыко (Санкт-Петербург) в докладе «Стратегия членения спонтанной речи на синтаксические единицы» проанализировали результаты эксперимента по членению спонтанных монологов на предложения. Монологи были взяты из Звукового корпуса русского языка, а разделение их на предложения было предложено шести лингвистам, имеющим опыт работы со спонтанной речью. Эксперимент, однако, показал, что профессиональные лингвисты, сталкиваясь при анализе звучащей речи с неполной согласованностью семантических, синтаксических и интонационных характеристик отрезков звучания, прибегают к разным стратегиям вычленения синтаксических единиц, причем ни один испытуемый не был последователен на протяжении всего эксперимента. В итоге расхождений оказывается больше, чем совпадений. По мнению авторов, анализ «уверенных» границ, т.е. проведенных одинаково большинством или даже всеми испытуемыми, позволяет выявить некоторые устойчивые критерии, что должно в будущем помочь в расшифровке устных текстов. Вместе с тем можно утверждать, что такая единица, как предложение, чужда спонтанной речи. Л.Л. Иомдин и Б.М. Лобанов (Минск) в докладе «Синтаксические корреляты просодически маркированных элементов предложения и их роль в задачах синтеза речи по тексту» представили попытку на основании синтаксической структуры высказывания восстановить автоматическим способом просодические акценты для синтеза речи по письменному тексту и продемонстрировали фрагмент текста, озвученный синтезатором.

Две секции были посвящены представлению знаний, онтологии и компьютерному анализу документов.

30 мая в работе секции, посвященной разработке и использованию корпусов, приняли участие Р.О. Муталов (Махачкала) с докладом «Опыт создания корпусов дагестанских языков», В.Г. Хуршудян, Д.В. Левонян, А.Е. Поляков и С.В. Рубаков, представившие Восточноармянский национальный корпус www.eanc.net, А.Е. Поляков, М.Б. Бергельсон, И.А. Пильщиков «Конкорданс к тек-

стам Ломоносова – концепция и реализация», С.А. Крылов и А.В. Тер-Аванесова «Использование лексико-грамматических баз данных в русской диалектной лексикографии», Д.В. Ландэ и В.В. Жигало (Киев) «Подход к созданию многоязычных параллельных корпусов веб-публикаций» и др.

На секции «Лексикография, фразеология, семантика» В.Е. Гольдин, А.О. Мартьянов и А.П. Сдобнова (Саратов) представили доклад «Электронный русский ассоциативный словарь школьников», который впоследствии был особо отмечен организаторами конференции. Были представлены результаты эксперимента, проводимого в течение 10 лет среди учащихся всех классов в 136 школах Саратова и области. В словаре зафиксированы 62,5 тысяч различных реакций; содержатся прямые и обратные статьи «стимул – реакция» с характеристикой по возрастному, гендерному, региональному признакам, времени проведения и типу школы. Обычно считается, что для ассоциативного словаря достаточно 500 реакций на один стимул, но так как в данном словаре идет дальнейшая фильтрация, объем существенно больше. Многие из полученных данных весьма неожиданны: так, на стимул *тяжелый* третья по частотности реакция (после *легкий* и *груз*) – слово *камень*, т.е. камень воспринимается как прототипически тяжелый предмет, что никак не отражено ни в одном толковом словаре. Замечательной и ценной особенностью словаря является то, что анкеты представлены целиком. Как отметил докладчик, реакция определяется не только стимулом, но и общей стратегией; так, реакции учеников младших классов часто связаны со стимулом по созвучию, рифме, или даются формы тех же слов, обычно – множественное число. Существует также тип анкет – и это одна пятая анкет малышей – в которой стимул не вызывает никакой реакции или реакция никак не связана со стимулом и выдается, как правило, то, что перед глазами. Обычно такие анкеты отбраковываются, но в данном эксперименте они учитываются, и они, на фоне множества отказов, содержат интересные, яркие ответы (так, второклассница на стимул *страшное* дает реакцию *зубастое*). Словарь позволяет проследить динамику языкового сознания: ядро реакций – около 60 % – константно и с возрастом всё больше сближается с ядром взрослого, причем какие-то слова с возрастом упрощаются в сознании, а другие – усложняются.

Интересное обсуждение вызвал также доклад М. Петренко «Онтологическая семантика и абдукция: обработка эллипсиса». Была предложена модель компьютерного анализа естественных текстов с эллиптическими сегментами. Если человек, решая подобные задачи, опирается на всю совокупность своих знаний о мире, то решение компьютера основывается на заложенных в него сведениях о лексических связях и о фреймах, т.е. типовых ситуациях и стандартных наборах их участников; таким образом, восстанавливаются стандартные и наиболее вероятные ситуации (что, разумеется, верно не для всех контекстов).

В тот же день состоялась стендовая секция, работа которой проходила в два этапа (организаторы считают это своим ноу-хау): на первом этапе участникам давалось 5 минут на краткую презентацию доклада, на втором шло обсуждение у стендов с демонстрацией компьютерных программ и аудио- и видеоматериалов. Стоит отметить доклад А.Л. Воскресенского «Словарь RusLED как инструмент семантических исследований», связанный с созданием системы автоматизиро-

ванного сурдоперевода; работа, первоначально направленная на решение чисто прикладных задач, высветила ряд более общих проблем, в том числе отличия когнитивных процессов у глухих от рождения, у потерявших слух в достаточно позднем возрасте и у слышащих.

31 мая секция «Речевые технологии» была посвящена синтезу речи по тексту. На параллельной секции «Синтаксис и семантика» были представлены как прикладные, так и теоретические доклады. В докладе Е.Л. Григорьян (Ростов-на-Дону) «О характере синтаксической полисемии» было отмечено, что большинство структур могут использоваться для выражения различных смыслов и что конструкции, противопоставляемые в одних контекстах, могут не нести никаких семантических различий в других и различаться только на уровне прагматики. А.В. Циммерлинг в докладе «Синтаксические нули в теории грамматики», обращаясь к синтаксической типологии, указал 4 различных по функции синтаксических нуля и рассмотрел вопрос о целесообразности использования этой категории применительно к конкретным случаям. В.В. Окатьев, Т.Н. Ерехинская и Д.С. Скатов (Нижний Новгород) в докладе «Модели и методы учета пунктуации при синтаксическом анализе предложения русского языка» продемонстрировали алгоритм для определения границ фрагмента, позволяющий программе извлекать цитаты из текста (это актуально: например, в Налоговом кодексе есть статья об НДС, состоящая из одного предложения, которое занимает 8 страниц!). В докладе А.Ю. Недолужко (Прага) «Разметка кореференции на синтаксически аннотированном корпусе чешских текстов» был представлен проект, разрабатываемый на кафедре прикладной лингвистики физико-математического факультета Карлова университета – программа, предназначенная для обучения компьютера распознаванию анафоры разных видов, в том числе ассоциативной анафоры.

Надо отметить исключительный уровень конференции: практически не было слабых или проходных докладов. Замечательна и сама атмосфера «Диалога», которую, к сожалению, передать невозможно.

Материалы конференции опубликованы в сборнике «Компьютерная лингвистика и интеллектуальные технологии. Вып. 8 (15). По материалам международной конференции "Диалог 2009"». М., 2009 и в Интернете на сайте www.dialog-21.ru, где также представлены некоторые материалы, не вошедшие в сборник. В данном обзоре мы попытались обратить внимание на некоторые моменты, которые не нашли отражения в опубликованных материалах.

Е.Л. Григорьян

**III МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ
«ЯЗЫК, ЧЕЛОВЕК И ОБЩЕСТВО В СОВРЕМЕННОМ МИРЕ»
(Болгария, 5 – 8 сентября 2009 г.)**

5 – 8 сентября 2009 г. в Болгарии состоялась ставшая уже традиционной Международная научная конференция «Язык, человек и общество в современном мире». Организаторами конференции стали Болгарская Академия наук, Союз ученых Болгарии, фонд «Наука и образование» и факультет лингвистики и словесности Педагогического института Южного федерального университета.