

## **Первое информационное письмо**

Уважаемые коллеги!

Лаборатория автоматизированных лексикографических систем Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова совместно с Институтом русского языка имени В.В. Виноградова РАН **28–29 сентября 2018 года** проводит Международную конференцию

**«От языковых машинных фондов к лингвистическим корпусам: памяти В.М. Андриященко».**

Необходимость построения языковых машинных фондов – компьютерных хранилищ лингвистических данных – была осознана в конце 70-х – начале 80-х гг. прошлого века. Теоретическим проблемам «машинного фонда данных для автоматизированной системы лексикографических исследований» были посвящены три всесоюзные конференции, проводившиеся в 1983, 1987 и 1989 гг. В 1985 г. по инициативе академика А.П. Ершова начались работы по созданию машинного фонда русского языка. В работе над Машинным фондом принимало участие более 40 организаций-соисполнителей; ведущей организацией стал Институт русского языка им. В.В. Виноградова, а Лаборатория вычислительной лингвистики (ныне Лаборатория автоматизированных лексикографических систем) Научно-исследовательского вычислительного центра МГУ под руководством В.М. Андриященко превратилась в одну из основных площадок проекта: именно здесь вводом в компьютер и лемматизацией Саратовского корпуса было положено начало русского корпуса устной речи, именно здесь был оцифрован Грамматический словарь А.А. Зализняка, без которого сейчас немыслима никакая автоматическая обработка русского текста. Тогда же параллельно с работой над компонентами машинного фонда русского языка началась работа над машинными фондами языков народов СССР.

На заре создания машинных фондов вряд ли кто-то ожидал, что компьютерная техника, на которую в то время многие смотрели с пиететом и с некоторым недоверием, будет развиваться столь стремительно, что разнообразные ухищрения с целью экономии компьютерной памяти или кодирования нестандартной графики станут неактуальными, электронные текстовые корпуса достигнут невообразимых объемов, а их автоматическая обработка – немыслимой скорости, и только ручное индексирование останется узким местом на пути к светлому лингвистическому будущему. Сегодня вместо текстового модуля Машинного фонда русского языка у нас есть Национальный корпус русского языка, вместо компьютерных словарей – электронные словари онлайн. Интернет,

еще одна новая технология, появившаяся с тех пор, сделала мир безграничным, а корпуса и словари общедоступными.

Задача конференции – обсудить развитие тех идей, которые легли в основу языковых машинных фондов, того, в какой степени эти идеи реализованы в современных корпусах, лексических базах и лингвистических программных средствах, что было приобретено и что, возможно, потеряно с развитием технических средств, и в какой степени современные корпуса и словарные базы могут считаться наследниками первых модулей машинных фондов.

**Тематика конференции** включает следующие направления (но не ограничивается ими):

- История развития компьютерной и корпусной лингвистики
- Лингвистические корпуса и их применение
- Квантитативные методы в лингвистических исследованиях
- Компьютерные словари и словарные базы данных
- Документация исчезающих языков, создание мультимедийных архивов и корпусов
- Компьютерные лингвистические ресурсы: программное обеспечение, специализированные сайты и т.п.
- Применение компьютерных лингвистических ресурсов в литературоведении и фольклористике
- Компьютерный анализ текста
- Применение языковых компьютерных ресурсов в образовании

В рамках конференции предполагается проведение мемориального заседания, посвященного В.М. Андриященко (20.02.1934–18.07.2017).

Желающим принять участие в конференции необходимо не позднее **10 апреля 2018 г.** прислать на адрес [confer.lcl@gmail.com](mailto:confer.lcl@gmail.com) неанонимные тезисы доклада в формате .doc, .docx или .pdf объемом не более 500 слов. В письме надо указать фамилию, имя, отчество, название доклада, место работы, электронный адрес. Оргкомитет оставляет за собой право отбора тезисов на основе анонимного рецензирования.

Решение оргкомитета будет сообщено авторам не позднее 15 мая 2018 г. Предполагается электронная публикация тезисов.

Регламент докладов: 30 минут (20 мин. доклад + 10 мин. обсуждение). Рабочие языки конференции: русский и английский.

Предполагаемый оргвзнос 1000 рублей. К сожалению, оргкомитет не сможет возместить расходы на дорогу и проживание.

#### Оргкомитет конференции

О.А. Казакевич (ЛАЛС НИВЦ МГУ), председатель

А.Я. Шайкевич (ИРЯ РАН), сопредседатель

М.И. Воронцова (ЛАЛС НИВЦ МГУ)

С.А. Крылов (ИВ РАН)

Н.В. Перцов (ИРЯ РАН)

В.А. Плунгян (ИРЯ РАН)

А.В. Рафаева (ЛАЛС НИВЦ МГУ), зам. председателя

И.В. Шумарина (ЛАЛС НИВЦ МГУ)

#### Важные даты:

- **10 апреля 2018 года** – окончание приема тезисов
- **15 мая 2018 года** – извещение о решении программного комитета
- **28–29 сентября 2018 года** – проведение конференции (НИВЦ МГУ, Москва)