

## **Предварительная научная программа Международной конференции «Корпусная лингвистика - 2017»**

### **27 ИЮНЯ 2017 ГОДА, ВТОРНИК**

#### **День Школ-семинаров**

10:00 – Регистрация

11:00 – Школа-семинар OpenCorpora. Ведущий: В. В. Бочаров.

12:45 – Школа-семинар Visual Analytics, Text Mining or Shiny Visualization for Corpus Linguistics. Ведущий: О. Scrivner.

14-30 – Школа-семинар «Система автоматической обработки тестов “Хурма”». Ведущий: А. М. Попов, Е. В. Еникеева.

### **28 ИЮНЯ 2017 ГОДА, СРЕДА**

#### **Пленарное заседание**

9:00 – Регистрация

9:45 – Открытие конференции:

Приветственное слово Виктора Павловича Захарова, доцента кафедры математической лингвистики СПбГУ, ведущий научный сотрудник Института лингвистических исследований РАН.

Пленарные доклады

- [ 10:00 – Annotation-based Digital Text Corpora Analysis within the TXM Platform. Докладчик: S. Heiden.
- [ 11:00 – Collostructions across text types: variability in lexicon-grammar associations. Докладчик: A. Stefanowitsch.
- [ 12:00 – Архипелаг корпусов: создание инструментов для работы в сети диалектных и диахронических корпусов. Докладчик: R. von Waldenfels.

13:00 – кофе-брейк

#### **Заседание секции «Исторические корпусы текстов»**

14:00 – Корпусное исследование рукописной традиции славянского евангелия. Докладчики: И. В. Азарова, Е. Л. Алексеева, Д. М. Миронова.

14:25 – Грамматический словарь церковнославянского языка (по материалам корпуса). Докладчик: А. Е. Поляков.

14:50 – Корпус русских франкоязычных дневников XIX в. как материал для исследования взаимодействия языков и культур. Докладчики: М. Дебрэнн, А. М. Лаврентьев.

15:15 – Создание открытого морфологического словаря и лемматизация старофранцузских текстов с использованием платформы TXM. Докладчик: А. М. Лаврентьев.

15:40 – кофе-брейк

15:55 – Возможности использования Корпуса текстов XIX века. Докладчик: Е. В. Рахилина, В. П. Фесенко.

16:20 – Дискурсивная «нормализация» текстов северорусских житий. Докладчики: И. В. Азарова, Е. Л. Алексеева, Е. А. Рогозина.

#### **Заседание секции «Речевые и мультимедийные корпусы»**

14:00 – Побудительная реплика в диалогическом окружении: пары реплик типа

«императив + вербальная реакция» и способы их разметки в речевом корпусе.

Докладчик: О. В. Блинова.

14:25 – Устная спонтанная речь: судьба некоторых грамматических единиц.

Докладчик: Н. В. Богданова-Бегларян.

14:50 – Наложение коммуникативных функций: изучение на мультимодальном корпусе «REC» и перенос на работа «Ф-2». Докладчики: А. А. Зинина, А. А. Котов, Н. А. Аринкин, Л. Я. Зайдельман.

15:15 – Стратегия передачи чужой речи в устном дискурсе в сравнении с письменным: опыт корпусного исследования. Докладчик: В. И. Подлесская.

*15:40 – кофе-брейк*

15:55 – Устная публичная речь в мультимедийном модуле НКРЯ. Докладчик: С. О. Савчук.

16:20 – Подходы к тематическому аннотированию звукозаписей повседневного бытового общения в корпусе «Один речевой день». Докладчик: Т. Ю. Шерстинова.

16:45 – «Конduit»: корпус устных детских текстов. Докладчик: П. М. Эйсмонт.

17:10 – Адаптация CHILDES для казахской детской речи. Докладчик: А. Шаяхмет.

15:55 – Корпус русской устной речи: текущие исследования и перспективы. Докладчик: Е. И. Риехакайнен.

## **29 ИЮНЯ 2017 ГОДА, ЧЕТВЕРГ**

### **Заседание секции «Создание и разметка корпусов текстов»**

9:00 – Introducing RusDraCor, a TEI-encoded Russian drama corpus for the digital literary studies. Докладчики: Frank Fischer, Daniil Skorinkin, Natasha Tyshkevich.

9:25 – Text mining toolkit for digital corpora. Докладчики: O. Scrivner, I. Trapido, J. Lee.

9:50 – Proximity as Entailment: its probability, causality, intuitive opacity and empiricism. Докладчики: V. Louw, M. Milojkovic.

10:15 – Автоматизация сегментации и частеречной разметки тибетского текста. Докладчики: П. Л. Гроховский, М. О. Михайлова.

10:40 – Открытый корпус вепсского и карельского языков (VepKar): предварительный отбор материалов и словарная часть системы. Докладчики: Н. Г. Зайцева, А. А. Крижановский, Н. Б. Крижановская, Н. А. Пеллинен, А. П. Родионова.

*11:05 – кофе-брейк*

### **Заседание секции «Обучающие корпуса»**

9:00 – Teaching corpus linguistics with Aranea web corpora. Докладчики: V. Benko, A. Butasova.

9:25 – Features on the interface between language teaching methodology and morphosyntax for classification of Russian study texts. Докладчик: S. Birzer et al.

9:50 – Information density in a corpus of university student writing. Докладчик: H. Nesi.

10:15 – Метод поиска групп слов в текстах с неснятой омонимией. Докладчики: Э. С. Клышинский, Д. В. Королев, А. А. Власова.

10:40 – Автоматическая разработка учебных тестов по английскому языку на основе корпуса. Докладчик: М. Кустова.

11:05 – Корпусная прагматика: о возможности создания учебного корпуса. Докладчик: М. К. Тимофеева.

*11:30 – кофе-брейк*

## **Стендовая секция**

11:45 – 12:25 – Представление докладов

12:25 – 13:15 – Обсуждение докладов

- [ The Corpus-based Electronic Historical Latvian Dictionary: from the idea of corpus to the dictionary and back to corpus. Докладчик: E. Andronova.
- [ The reconstruction of the context of culture through corpus stylistics. Докладчик: M. Milojkovic.
- [ Двухязычный корпус в сопоставительном анализе глагольных форм. Докладчик: Е. Г. Андреева.
- [ Устойчивость коллокаций: диахронический подход. Докладчики: В. В. Бочкарев, В. Д. Соловьев, А. В. Шевлякова.
- [ Verbless kakoj-exclamatives in Russian: Corpus Study. Докладчик: А. Вишенкова.
- [ Поиск кластеров в сети текстуальных связей на материале газеты «Гардиан». Докладчик: С. А. Зуева.
- [ Визуализация тематических моделей с помощью iPython. Докладчик: С. Н. Карпович.
- [ Диалектный текстовый корпус: проблемы репрезентативности, сбалансированности, единиц хранения и выдачи. Докладчик: О. Ю. Крючкова, В. Е. Гольдин.
- [ Посессивные конструкции с «у»-локализатором и их эквиваленты в арабском языке (на материале параллельного корпуса) . Докладчик: В. В. Шеремет.

## **Заседание секции «Создание корпусов»**

14:00 – Новый китайско-русский параллельный корпус с дискурсивно-структурной разметкой. Докладчик: И. Ян.

14:25 – Анализ переводческого соответствия лингвистических терминов на основе параллельного корпуса русского и китайского языков. Докладчики: Тао Yuan, В. П. Захаров.

14:50 – О разработке диахронического исторического корпуса румынского языка. Докладчики: A. Colesnicov, L. Malahov.

15:15 – Статистические исследования во фразеологии: проблема фразеологичности корпусов текстов. Докладчики: А. И. Баранов, М. М. Вознесенская, Д. О. Добровольский, А. Д. Козеренко, К. Л. Киселева.

15:40 – Automatic detection of neologisms in russian newspaper corpora. Докладчик: T. Iakovleva.

## **Заседание секции «Морфология и синтаксис в корпусах»**

14:00 – К разработке системы аннотирования аналитических конструкций для татарского национального корпуса. Докладчики: А. Р. Гатиатуллин, А. М. Галиева, Д. Ш. Сулейманов.

14:25 – Корпус-менеджер для морфосинтаксической разметки: опыт разработки корпуса тибетских грамматических сочинений. Докладчики: П.Л. Гроховский, А. В. Добров, А. Е. Доброва, Н. Л. Сомс.

14:50 – Микросинтаксическая разметка в корпусе русских текстов. Докладчик: Л. Л. Иомдин.

15:15 – К методике создания корпусов для машинного обучения: «Тайга», синтаксический корпус и парсер. Докладчики: Т. Шаврина, О. Шаповалова.

15:40 – Квантитативный подход к оценке употребительности синтаксических молекул на основе грамматически размеченного корпуса. Докладчик: С. А. Крылов.

**30 ИЮНЯ 2017 ГОДА, ПЯТНИЦА**

### **Заседание секции «Семантика в корпусах. Лексикография и корпуса»**

9:00 – Синонимия общественно-политических терминов в татарском языке (на корпусных данных). Докладчики: А. М. Галиева, О. А. Невзорова.

9:25 – Эмпирическое корреляционное отношения как метод разыскания границы между ядром и периферией в частотных словарях. Докладчик: Г. Я. Мартыненко.

9:50 – Параллельный корпус текстов на основе двуязычных глоссариев: проблемы слияния и конвертации. Докладчик: Л. Н. Беляева.

10:15 – Средства машинного перевода при обработке параллельных текстов. Докладчики: Г. Е. Кедрова, С. Б. Потемкин

10:40 – Об использовании данных Национального корпуса русского языка для иллюстрирования статей компьютерного семантического словаря. Докладчик: С. Ю. Семенова.

11:05 – К вопросу об аппроксимации зависимости объема словаря от объема выборки. Докладчик: А. О. Гребенников.

*11:30 – кофе-брейк*

### **Заседание секции «Создание и разметка корпусов текстов»**

9:00 – Текстовые базы данных по русским говорам. Докладчики: С. А. Крылов, А. В. Тер-Аванесова.

9:25 – О корпусе официально-деловых текстов русского языка#. Докладчики: С. А. Крылов, О. Е. Фролова.

9:50 – Разметка языковых ошибок в корпусе REALEC. Докладчики: О. Н. Ляшевская, Е. А. Кузьменко, М. А. Грабовская, О. И. Виноградова.

10:15 – Корпус текстов на ижорском языке как основа лингвистической экспертной системы. Докладчик: И. С. Николаев

10:40 – Разметка бриджинг-анафоры на материале русскоязычных новостных текстов. Докладчики: А. Ройтберг, К. Христосова.

11:05 – К методике исследования утраты ударения неоднозначными словами на материале корпуса. Докладчик: Д. В. Сичинава.

*11:30 – кофе-брейк*

### **Заседание секции «Коллокации»**

11:45 – Parallel keyword analysis. Докладчики: Jim Clarke.

12:10 – Multiword Expressions in Czech: Between Lexicon and Grammar. Докладчики: M. Hnatkova, V. Petkevic, H. Skoumalova

12:35 – Автоматическое выделение ключевых слов и словосочетаний из русскоязычных корпусов текстов с помощью алгоритма RAKE (Rapid Automatic Keyword Extraction). Докладчик: О. А. Митрофанова и др.

13:00 – Автоматическое распознавание терминов предметной области «Вирусология». Докладчики: Ю. И. Морозова, Е. Б. Козеренко.

13:25 – К вопросу о полноте и точности автоматического выделения словосочетаний в русском языке. Докладчик: М. В. Хохлова.

13:50 – Автоматическое составления словаря коллокаций на основе корпуса. Докладчики: Е. О. Каллас, К. Р. Коппель, Р. Г. Каллас.

### **Заседание «Статистические исследования на основе корпусов текстов»**

15:00 – Средний интервал в дистрибутивно-статистическом анализе текстов. Докладчик: А. Я. Шайкевич.

15:25 – Диахроническое исследование частотного поведения некоторых фиксированных словосочетаний политической лексики XX века на основе корпуса

Google Books. Докладчики: В. П. Захаров, А. Ц. Масевич.

15:50 – Вероятностная модель для оценки объема лексикона по данным корпуса Google Books Ngram. Докладчики: Ю. С. Масленникова, В. В. Бочкарев, В. Д. Соловьев.

16:15 – Сравнение корпусов мерой ???: Символы, слова, леммы или частеречные пометы? Докладчик: А. Ч. Пиперски.

16:40 – Динамика частот употребления семантически близких слов. Докладчики: В. Д. Соловьев, В. В. Бочкарев, Л. Янда.

**17:15 – круглый стол и закрытие конференции**