

### Отчет о командировке М. В. Хохловой

С 16 по 24 июля 2018 г. я находилась в командировке в г. Любляне (Словения) с целью участия в XVIII международном лексикографическом конгрессе Euralex. В рамках научного мероприятия я выступила с двумя докладами. В докладе “Building a Gold Standard for a Russian Collocations Database” был описан процесс создания золотого стандарта сочетаемости для русского языка, разрабатываемого нами на основе имеющихся лексикографических источников. Была произведена обработка Словаря русского языка (Евгеньева 1981–1984) и были выделены около 13 тыс. устойчивых словосочетаний, приведенных в заромбовой части словарных статей. Данная информация сравнивалась с теми биграммami, которые были автоматически извлечены на основе русскоязычного корпуса текстов Aranea Russicum Maximum (объемом свыше 15 млрд. слов). В докладе “Contrastive Collocation Analysis — a Comparison of Association Measures across three Different Languages using Dependency-Parsed Corpora” (соавторы Stefan Evert, Thomas Proisl, Peter Uhrig) обсуждались эксперименты по сопоставлению статистических мер для автоматического выделения коллокаций (модели прилагательное-существительное) на материале трех языков (английского, немецкого и русского). Мера log-likelihood продемонстрировала наиболее успешные результаты относительно параметра AP50 для английского и немецкого языков, для русского языка лучшие результаты были показаны мерами mutual information и Dice. Наибольшая полнота (процент словосочетаний, совпадающих с золотым стандартом) для англоязычного корпуса составляет 78,5 %, для немецкоязычного корпуса — 95,5 % и для русскоязычного — 90,2 %. Материалы конгресса опубликованы на сайте конференции (<http://euralex2018.cjvt.si/publication/>) и будут проиндексированы в системе Scopus. В обоих докладах были отражены результаты работы по гранту Президента Российской Федерации для государственной поддержки молодых российских ученых «Исследование методов автоматического извлечения лексических конструкций на основе машинного обучения» (МК-2513.2018.6), а также совместного исследования, проводимого с немецкими коллегами из университета Нюрнберга-Эрлангена (Германия) в рамках программы DAAD «Дмитрий Менделеев».

Нами также было налажено сотрудничество с зарубежными исследователями. С коллегами из Люблянского университета (проф. С. Крек) и Института эстонского языка

(Е. Каллас и М. Лангеметс) обсуждалась возможность подачи совместной заявки на грант Erasmus+ Key Action 2 — Capacity Building Projects. В ходе конгресса была достигнута договоренность с фирмой Lexical Computing (М. Якубичек) о дальнейшей работе по созданию новой версии системы для изучающих русский язык ruSkELL.



25.07.2018

М. В. Хохлова