

Комментарии жюри к системе оценивания работ 2 этапа

Основная сложность, которая стояла перед участниками, заключалась в том, что им предлагалось сравнить частотность местоимений в корпусах разных объёмов: основной подкорпус Национального корпуса русского языка (НКРЯ), включающий в себя прозаические тексты, содержит около 230 млн словоформ, а поэтический — около 11 млн словоформ. Поэтому тот факт, что слово ты в основном подкорпусе встретилось около 800 тыс. раз, а в поэтическом подкорпусе — около 110 тыс. раз, ни о чём не говорит сам по себе. Необходимо нормировать эти данные с учётом размера подкорпусов, разделив количество вхождений искомым слов на объём подкорпуса в словах (смотреть на количество документов не следует, потому что в среднем документы в поэтическом подкорпусе гораздо короче, чем в основном). После этого для удобства можно домножить результат на 100, тем самым выразив его в процентах, или на любое другое большое число — например, на 1 000 000, как это обычно делают в корпусной лингвистике.

Кроме того, при поиске следует обращать внимание не только на статистику, но и просматривать результаты: это позволяет обнаружить несоответствие выдачи желанием исследователя. Так, если в НКРЯ искать в текстах с неснятой омонимией местоимение оно во всех формах, большинство найденных форм на самом деле будет относиться к местоимениям он и они (его / него, ему / нему, им / ним, нём) и частотность местоимения оно окажется завышенной в несколько раз. Есть и другие сложности, которые следовало преодолеть при работе с НКРЯ: например, в поэтическом подкорпусе недоступен поиск по семантическим признакам; притяжательные местоимения не имеют грамматических помет лица и т. д.

Предполагалось также, что участники конкурса не только представят статистику, но и дадут своим результатам содержательную интерпретацию, связав их с тематическими и метрическими особенностями поэтических текстов, отличающимися их от прозы.

Школьники, которые сумели получить нормированные результаты, обобщить их и описать проблемы, с которыми они столкнулись при поиске, стали победителями конкурса. Призёрами стали те школьники, которые не выполнили операцию нормирования или выполнили её некорректно, тем самым не получив возможности сравнивать между собой прозу и поэзию, однако сделали ценные наблюдения на материале отдельных подкорпусов.