

В диссертационный совет НИУ ВШЭ
по филологии и лингвистике

Отзыв члена Комитета

Орехова Бориса Валерьевича, канд. филол. наук

на диссертацию Алексеевского Даниила Андреевича
на тему: «Методы автоматического выделения тезаурусных отношений на
основе словарных толкований», представленную на соискание ученой
степени кандидата филологических наук НИУ ВШЭ

В диссертации Д. А. Алексеевского я прежде всего нахожу попытку эффективно соединить фундаментальное лингвистическое (точнее говоря, лексикографическое) знание с практическими подходами, используемыми в настоящее время в секторе ИТ, имеющем дело с задачами обработки естественного языка. Эта принципиальная интенция фактически предопределила архитектуру работы, в которой существенное место занимает обзор лексикографических классификаций, опытов и достижений. Несмотря на то, что задача диссертантом ставилась сугубо практическая, а компьютерная лингвистика, в рамках которой сделано то, что изложено в работе, область, скорее, инженерная, нежели исследовательская, серьезное внимание к обзору фундаментальных трудов по смежным темам вызывает уважительное отношение к тексту Д. А. Алексеевского.

Так, в первой главе много и по делу говорится о представлении семантики в словарной статье, появляются имена В. В. Морковкина, Ю. Н. Караулова и других ученых, строго говоря, в сугубо инженерной работе не обязательные, но хорошо вписывающиеся именно в научный текст, жанрово требующий не просто решения практической задачи, но и автопозиционирования относительно существующей системы знания.

Конечно, эти исторические и теоретические экскурсы могли бы быть детальнее. В частности, кажется, несправедливым, что диссертант прошел мимо классического идеографического словаря Р. Халлига и В. фон Вартбурга, оказавшего сильное влияние и на западную, и на отечественную лексикографическую традицию.

Но ясно, что для тщательного обзора именно лексикографической традиции пришлось бы писать отдельную работу, а у диссертанта был другой целевой вектор. Однако само по себе внимание к теоретической базе

исследуемой темы кажется мне важным классифицирующим признаком представленной к защите работы как научной и серьезной.

К слову, надо сказать ещё об одном теоретически значимом измерении диссертации, в самой работе не отмеченном. Благодаря подробному разбору текстов толкований в словаре, предпринятому Д. А. Алексеевским, мы можем приблизиться к пониманию границы между естественным языком и пресловутым полуискусственным *метаязыком* лингвистики (или, если угодно, лексикографии).

В то же время упомянутая практическая направленность диссертации прямо коррелирует с практической актуальностью решаемой задачи. Действительно, многие и крупные, и мелкие задачи компьютерной лингвистики, связанные с семантикой текста и его единиц, традиционно опираются на онтологию типа WordNet, которой для русского языка по-прежнему нет. Это сильно затрудняет транспонирование готовых решений на русский языковой материал, то есть, в конечном счете, замедляет прогресс. А предлагаемые Д. А. Алексеевским инструменты позволяют с оптимизмом смотреть в будущее и надеяться, что задача создания русского WordNet'a всё же принципиально реализуема.

Наряду с теоретическим базисом и практическим ядром в диссертации присутствует и ещё один, на мой взгляд, важный для научного исследования компонент: нетривиальный поворот темы. Д. А. Алексеевский сумел сложить в одной работе, с одной стороны, методы и инструменты, широко применяемые в решении задач обработки естественного языка, и, с другой стороны, коллекцию текстов специфического жанра (с определенными оговорками можно их назвать текстами на не вполне естественном языке), с которыми такими методами и с помощью таких инструментов, работать было не принято. Обнаружение этого пути само по себе ценная находка, свидетельствующая об остроумии как самой работы, так и образа мысли написавшего ее автора.

Отдельно хочется отметить, на первый взгляд, формальные, но на деле, как представляется, чрезвычайно идеологически важные достоинства представленного к защите текста. Речь идет о том, что результаты предпринятых экспериментов опубликованы автором, что специально отмечено в тексте и в нынешних условиях может служить критерием качества этой части исследования. И ещё более важно то, что в свободном доступе размещен код, создававшийся в процессе написания работы. Это служит как абстрактному прогрессу (заинтересованные исследователи и разработчики смогут воспользоваться готовыми решениями), так и совершенно конкретной воспроизводимости результатов диссертации.

Ход работы над задачей изложен очень последовательно и транспарентно. В большинстве случаев не возникает сомнений в том, что именно диссертант имеет в виду и как следует трактовать его действия и

слова. Отдельные частные несоответствия этому принципу единичны. В частности, в § 3.4.7 говорится о том, что данные оказываются распределены согласно закону Ципфа, и не совсем понятно, идет ли речь именно о законе (эмпирической закономерности) Ципфа, описывающем частотность слов в корпусе естественного происхождения, или всё же о приложимом к произвольным количествам (не исключительно словам и корпусам) степенном распределении.

Сказанное должно быть резюмировано как положительная оценка диссертации Д. А. Алексеевского. Считаю, что перед нами достойное, зрелое и научно ценное сочинение. Но отдельные стороны работы вызывают и чувство неполного удовлетворения.

Во-первых, от диссертационной работы хотелось бы получить твердый и надежный метод решения поставленной проблемы. Кажется, что все же это удалось диссертанту не вполне. Хотелось бы узнать оценку, какая доработка и в каком объеме будет нужна для внедрения исследованных решений в дело для получения общественного блага.

Во-вторых, именно в научном (не практическом) смысле не хватает более глубокой интерпретации результатов кластеризации, предпринятой в главе «Выделение отношений». Возможно, типы словарных толкований соотносимы не только с типами тезаурусных отношений, но и объективными фактами языка, определенными лексико-семантическими классами и т. д.

Тем не менее, в целом текст вызывает позитивные эмоции, и по обоснованности сформулированных в нем научных положений, и по их достоверности и новизне соответствует всем требованиям, а сам Д. А. Алексеевский заслуживает присвоения ученой степени кандидата филологических наук НИУ ВШЭ.

Орехов Борис Валерьевич
Кандидат филологических наук

Национальный исследовательский университет «Высшая школа экономики»
Должность: доцент Школы лингвистики
Адрес электронной почты: nevmenandr@gmail.com

Подпись
8.11.2018