

**В диссертационный совет НИУ ВШЭ  
по филологии и лингвистике  
Отзыв члена Комитета**

**Орехова Бориса Валерьевича, кандидата филологических наук**

на диссертацию Хоменко Анны Юрьевны  
на тему: «Лингвистическое моделирование как инструмент атрибуции  
текста», представленную на соискание ученой степени кандидата  
филологических наук

Сразу хочу определиться с тем, что это хорошая работа, демонстрирующая квалификацию соискателя, соответствующую ученой степени. Как это и требуется, автором диссертации освоен (прочитан и осмыслен) необходимый объем литературы по теме, сделан свой — достаточно любопытный — вклад в научную работу в этом поле, а результаты своих усилий изложены стройно, логично, приличествующим жанру научным языком.

Жанр, в котором я выступаю, требует от меня пересказа содержания диссертации. Я сделаю его по возможности кратким, но при этом буду говорить большей частью о том, что вызывает вопросы и сомнения, нежели о том, что сделано на должном уровне. Так я строю текст не потому что в целом в работе мне видится какая-то недостаточность. Напротив, диссертация, как мне представляется, подготовлена, сконструирована композиционно и написана достойно. То, как я рассказываю о ней, продиктовано исключительно риторическими соображениями, а всё, о чем я не упомяну, следует считать имплицитной высокой оценкой.

В первой главе «Лингвистические модели и их свойства» задается теоретическая и терминологическая рамка диссертации. На широком методологическом материале от Лосева до Хомского формулируются разные подходы к определению модели, дается их классификация и агрегируются основные свойства. Акцентируется внимание на полноте и простоте как важных свойствах моделей, а под самой моделью понимается «абстрактное понятие эталона или образца грамматической, фонологической или иной системы и представление самых общих характеристик к.-л. языкового явления; общая схема описания системы языка или его подсистем» (с. 41). В данном случае не по своей вине соискатель становится заложником научной традиции, которая, с одной стороны, предписывает создать условия полной терминологической прозрачности работы, с другой, не обязывает затем опираться на эту базу в дальнейшем. В сущности, первая глава почти не отзывается в дальнейшем описании исследования, и выглядит во многом избыточной. В то же время член комитета осознает, что если бы такой главы не было, то конструкция диссертации стала бы значительно более уязвимой для формальной критики. Сейчас текст защищен от атак с этой стороны, и позволяет сосредоточиться на содержательных вопросах, поднятых в диссертации. Так что слова об избыточности не следует считать упреком. Однако несмотря на эту самую избыточность, вызывает некоторое недоумение и контекстуальная неполнота освещения вопроса. Действительно, в диссертации

упомянуты теоретические и методологические подходы к тому, что такое модель в лингвистике. Но в последние годы под словами «языковая модель» имеется в виду прежде всего прикладное распределение вероятностей последовательностей слов в корпусе текстов. Поскольку диссертация всё же имеет явную компьютерную ориентацию, не упомянуть об этом выглядит странным решением.

Вторая глава «Проблема атрибуции текста. История вопроса» сосредотачивается на языковых проявлениях индивидуальности говорящего. В частности, идея стиля связывается с понятием языковой личности. В историческом обзоре названы ключевые для темы фамилии Менденхолла, Морозова, Хмелева, Фоменко, Бёрроуза и других. В целом обзор выглядит системным и полным. Мелкой погрешностью следует считать ряд тех, кто занимался «разработкой количественной составляющей атрибуционного анализа в России в середине – конце XX в.» (с. 53), так как в этом ряду оказался норвежский исследователь Гейр Хьетсо, который всё же занимался этими вопросами в Норвегии. Не упомянуты работы И. В. Пешкова в связи со спорными текстами круга М. М. Бахтина, но эта методика и осталась принадлежностью локальной научной проблемы, так что упоминать ее было не обязательно.

На что я бы обратил внимание на месте соискателя, это на то, что в историческом обзоре все методики представлены равными друг другу и как будто бы дополняющими одна другую. В действительности разные авторы при атрибуции одних и тех же текстов могли приходить к противоположным результатам (ср. работы Г. Хьетсо и супругов Фоменко о «Тихом Доне»), что делает историю вопроса более драматичной, чем это описано в диссертации. Иными словами, между разными подходами существует поле напряжения, и оно в тексте проигнорировано.

Отчасти этот недостаток призвана компенсировать третья глава «Оценка качества существующих квалификативных иквантитативных атрибуционных моделей», посвященная вопросам оценки методик. Внимание к этой теме следует счесть сильной стороной работы, поскольку далеко не всегда авторы атрибутивных методов утруждают себя проверками разработанных ими инструментов анализа. В то же время эта глава имеет ту же родовую травму, что и предыдущие две: она практически не имеет семантических связей, «переключек» с остальными частями диссертации. При всей своей имманентной полноте и разумности глава представляет собой отдельную работу, не помогающую в раскрытии тем, заявленных в остальных главах. Недостаток это формальный, не из существенных, и скорее эстетический, но при прочтении бросается в глаза.

Глава 4 «Создание прототипа интегративного полуавтоматического алгоритма идентификации автора письменного текста» наконец подводит нас к главному сюжету работы. Предыдущие главы были лишь исполнением необходимых ритуальных номеров. Речь идет о создании интегративной модели, учитывающей и теоретические представления о языке (то, что автор называет квалификативной моделью), и прикладные свойства речи (то, что можно обратить в числовые категории, автор называет их квантитативной моделью). Обоснование такого подхода и предложенные пути его практической реализации выглядят вполне разумно, и достойны внимания научного сообщества. Любопытно, как в

практическую плоскость оказались переведены очень разные теоретические разработки от понятия семантической валентности до индекса удобочитаемости текста. При этом, как показывает оценка работы модели, ей действительно удается разрешать сложные случаи задач атрибуции. Но даже более значимым для меня является то, что в отличие от других атрибутивных методик вроде Delta Берроуза, предложенная соискателем модель имеет большой интерпретационный потенциал. Исследователю трудно понять, почему такие, а не иные распределения служебных слов в текстах некоторого автора служат репрезентацией его авторского стиля, а вот когда речь идет о том наборе лингвистических категорий, которые вошли в модель соискателя, лингвистическая теория может предложить свои объяснительные концепции. Таким образом, мы получаем не столько прикладной инструмент атрибуции, сколько надежду на лингвистически корректное и количественно проверяемое описание идиостиля.

Диссертация имеет гигантское приложение, на внимательное изучение которого у меня ушло довольно много времени. Однако это и хорошо: с помощью приложения, а также кода, размещенного в открытом репозитории на GitHub, соискатель стремится повысить прозрачность своей работы и сделать более отчетливым свой вклад в тему. Удобно и создание на основе методики автора развернутого ресурса.

Скажу также и еще об одном частном аспекте. Полностью автоматическая атрибуция — это тенденция и необходимость, поэтому о ней и говорят так много на упомянутых в диссертации конференциях. Поэтому создание полуавтоматической методики, которая представлена в диссертации, кажется не вполне своевременным. В то же время работа содержит несомненную теоретическую пользу, и поэтому может быть признана важным вкладом в проблемное поле, а соискателю должно быть присуждена искомая степень.

Орехов Борис Валерьевич

Кандидат филологических наук,

4 октября 2021

Национальный исследовательский университет «Высшая школа экономики»

Доцент Школы лингвистики

borekhov@hse.ru