

Коломийцева Екатерина Леонидовна, Россия, г.Мурманск

Kolomiitseva Ekaterina Leonidovna, Russia, Murmansk

Федеральное государственное бюджетное образовательное учреждение
высшего образования "Мурманский арктический государственный
университет" (ФГБОУ ВО «МАГУ»)

Federal State Budget Educational Institution of Higher Education "Murmansk
Arctic State University" (FSBEI HE "MASU")

e-mail: eoukaterina@mail.ru

Проблемы идентификация текста по его авторской принадлежности

Статья описывает попытки разработок формально-количественной модели идентификации текста на лексическом уровне. Текст (как персонотекст – носитель образа его автора) содержит проявление свойств языковой личности – языковые характеристики, выраженные, в том числе, количественными показателями. Формально-количественное описание лексических единиц в той или иной степени способно отразить авторскую принадлежность текста.

Ключевые слова: идентификационная лингвистика; лингвоперсонология; персонотекст; формально-количественная модель, идентификация текста; квантитативная лингвистика; юрислингвистика.

Problems identifying text according to its author

The article describes attempts to develop a formal-quantitative model for identifying text at the lexical level. The text (as a person text - the carrier of the image of its author) contains a manifestation of the properties of a linguistic personality - language characteristics, expressed, among other things, by quantitative indicators. The formal-quantitative description of lexical units in one degree or another can reflect the author of the text.

Keywords: identification linguistics; linguopersonology; person text; formal quantitative model, text identification; quantitative linguistics; jurisprudence.

Идентификация текста по воплощенным в нем языковым особенностям автора является одной из лингвистических проблем. Речь и ее продукты - одни из основных средств идентификации личности, так как с их помощью можно составить представление о внутреннем мире человека. Способность языковой личности говорить и воспринимать речь других людей, помимо всего прочего, получает свою реализацию за счет устойчивых речевых

«образцов», которые являются бессознательными маркерами личности. В работе мы рассматриваем характеристики лексических единиц, составляющих текст, по их частоте.

Феномен идентификации изучается в разных научных дисциплинах: логике, философии, криминалистике. Идентификация стала объектом для исследования и в лингвистике, и научную основу её составляют знания и законы и закономерностях речевого поведения индивида, что, в свою очередь, определяет характеристики письменной речи человека. В лингвистике может осуществляться идентификация разных объектов – например, идентификация языковой личности (разновидность портретирования языковой личности), где учитываются индивидуальные языковые предпочтения. Так как «речевой портрет – это воплощенная в речи языковая личность» [4, с.28], то в речевом портрете отражаются разнообразные личностные характеристики (социальные, возрастные, этнокультурные, возрастные, гендерные и др.) [3, с.116].

Диагностические (не только идентификационные) задачи ставятся при проведении автороведческой экспертизы. При определении речевого портрета также решаются задачи диагностического плана (пол, возраст, профессия, социальный статус, моральные ценности, установки и др.). При идентификации персонотекста объектом становится текст в аспекте проявления в нем свойств языковой личности. «Язык устроен так, а не иначе еще и потому, что так устроена персонологическая сфера» [5, с. 24]. Любой человек для выражения своих коммуникативных намерений из разнообразия средств и способов выбирает то, что находится «в соответствии с его внутренними личностными потребностями и способностями» [5, с. 12]. Персонотекст является разновидностью речевого произведения и отражает языковой потенциал личности, реализуемый автором в тексте. Выбор из бесконечного множества формирует индивидуальные предпочтения языковой личности - её идиостиль. Задача эксперта-лингвиста – увидеть в тексте и описать варианты способностей языковой личности, что поможет сравнивать персонотексты и осуществлять их идентификацию.

Исследователями проблемы идентификации текстов описаны различные методы определения авторства на различных уровнях (орфографическом, пунктуационном, синтаксическом, стилистическом, лексико-фразеологическом). Напреенко Г.В., например, предлагает вариант методики идентификации персонотекста на лексическом уровне – с точки зрения частоты употребления лексем, составляющих персонотекст (формально-количественная модель). Частотность здесь выступает как особая характеристика слова. Качественная интерпретация количественных

характеристик лексем дает определенные данные по распределению лексем в речевом произведении. Сопоставление словариков по каждому тексту в данном ракурсе отражает разность авторских предпочтений языковой личности. Функциональные зависимости отражают не только строй языка, они проявляются в тексте индивидуально: в них отражаются способности языковой личности.

Трудность заключается в том, что языковые характеристики текста должны быть настолько формализованы, чтобы быть готовыми к обработке компьютерными программами, то есть необходим перевод знаков из одной системы в другой. В данном случае это перевод вербальных знаков в невербальные (количественные характеристики). Эти количественные данные «заменяют» лексические единицы и выступают в роли идентификаторов текста, что будет способствовать автоматизации экспертизы (исследования).

Большое количество исследований направлено на поиск и создание универсальной методики обработки текстов для усовершенствования лингвистической экспертизы, однако исследователи видят следующие актуальные проблемы лингвистической идентификации, требующие профессионального решения: субъективность текстов, проблема объективизации методик, трудность создания универсальной методики, применение матанализа параметров текста, определение критериев отбора параметров и их количества [6, с.20].

Например, алгоритм, предложенный Хоменко А.Ю., совмещает несколько методологий (в частности, стилометрию), но оправдывает свою достоверность лишь на 50% и подходит лишь для текстов определённого объёма и стиля. Родионова Е.С. делала попытку применять методы статистики на синтаксическом уровне. Мещеряков Р.В. даёт совокупное представление о существующих параметрах, которые могут выполнять идентифицирующую функцию. Мещеряков параметром идентификации считает особенности словаря автора текста и подчёркивает, что выявление отличительных черт авторского лексикона во многом носит субъективный характер, зависит от личности исследователя, и, кроме того, данный признак проявляет значительную зависимость от смены темы и жанра коммуникации [там же, с.21]. Таким образом, методы идентификации на данный момент субъективны, универсальной методики не существует.

Методика Напреенко Г.В. предлагает формально-количественную модель идентификации персонотекста, реализуя такую логику: универсальность – формализация – автоматизация – объективность.

Перспектива исследования – выявление особенностей текста, обусловленных особенностями языковой способности автора. Выдвигается гипотеза: текст содержит индивидуальные лексические характеристики, которые, выраженные контрастными по отношению друг к другу количественными данными, могут являться идентификаторами персонотекста. Наиболее частотные единицы русского языка содержатся в любом тексте и являются высокочастотными для любого текста. Выпадение их в частотном словаре конкретного текста из этого статуса или, напротив, появление в списке высокочастотных единиц дает основания для утверждения о том, что данный текст нетиповой. Подобная специфичность, если она повторяется в разных персонотекстах, даёт основания для их идентификации [6, с.22].

Базовый метод данной методики – статистический анализ. Формализация идентификации персонотекста способствует повышению уровня объективности исследования и даёт возможность использования в судебной лингвистической экспертизе.

ЛИТЕРАТУРА

1. Голев Н.Д. Лингвистическое сравнительное и автороведческое исследование трех текстов // Юрислингвистика-10. Лингвоконфликтология и юриспруденция. Кемерово ; Барнаул : Изд-во Алт. ун-та, 2010. С. 422–431.
2. Голев Н.Д. От редактора: Актуальные проблемы юрислингвистической экспертизы // Юрислингвистика-3. Проблемы юрислингвистической экспертизы : межвуз. сб. науч. тр. Барнаул : Изд-во Алт. ун-та, 2002. С. 5–13.
3. Кожина Н.Г. Аксиологический подход к исследованию идентификации современной личности // Историческая и социальнообразовательная мысль. № 1. 2012. URL: <http://cyberleninka.ru/article/n/aksiologicheskaya-identifikatsiya-lichnosti> (дата обращения: 10.05.2019).
4. Леорда С.В. Речевой портрет современного студента : автореф. дис. ... канд. филол. наук. Саратов, 2006. 20 с. URL: <http://www.dissforever.ru/order.php?id=714> (дата обращения: 26.05.2019).
5. Лингвоперсонология и личностно-ориентированное обучение языку : учеб. пособие. Кемерово : КемГУ, 2009. 384 с.
6. Напреенко Г.В. Идентификация текста по его авторской принадлежности на лексическом уровне (формально-количественная модель) // Вестн. Том. гос. ун-та. Филология. 2014. № 379. С. 17-23.
7. Резанова З.И., Романов А.С., Мещеряков Р.В. О выборе признаков текста, релевантных в автороведческой экспертной деятельности // Вестн. Том. гос. ун-та. Филология. 2013. № 6 (26). С. 38–52.