

ОТЗЫВ

на автореферат диссертации И.В. Смирнова
**"Методы разноуровневого анализа текстов на естественном языке
и их приложения в системах информационного поиска
и психолингвистических исследованиях"**,
представленную на соискание ученой степени
доктора технических наук
по специальности 1.2.1 - Искусственный интеллект и машинное обучение.

Общая цель всего комплекса исследований, основные результаты которых составили диссертационную работу И.В. Смирнова, – создание технологии анализа текстов на естественном языке, которая учитывает их сложную организацию, представленную множеством уровней структур – морфологических, синтаксических, семантических и дискурсивных. Это *комплексное междисциплинарное* исследование, поскольку для его реализации автор использовал методы компьютерной лингвистики, машинного обучения, информационного поиска, математической статистики, программной инженерии, отнесенные к проблематике лингвистики, психолингвистики, психологии.

План исследования, его результаты и текст диссертации построены в соответствии с методологическим представлением о многоуровневом строении текстов (объекта исследования), составляющие которого (морфологические, синтаксические, семантические и дискурсивные), хотя они и имеют собственные описания, но представляют собой лишь аспекты языка. Аспекты, абстрагируемые из сложной знаковой системы языка, не могут быть изолированы, не обладают «самостоятельностью существования». В обзоре литературы И.В. Смирнов замечает, что необходимость интеграции этих аспектов в единое описание остается общей проблемой фундаментальных исследований и актуальной задачей в разработке технологий, оперирующих языком, но трудности в решении этого круга проблем таковы, что в

интеллектуальном анализе текстов на русском языке разноуровневый анализ практически не применяется.

Для построения единой модели, сохраняющей важнейшие свойства выделенных уровней целостной знаковой системы языка, автор использует методологию системного анализа.

В формулировках задач и результатов работы следует выделить устойчивое взаимопроникновение уровней, их интегрированное описание. Так, в построении алгоритмов семантического поиска важнейшее, не «подчиненное» значение придано синтаксической и дискурсивной структуре текста. Именно такой системный анализ дает основания для совмещения семантического и синтаксического анализа, позволяет вычленять из текстов предикатно-аргументные структуры, представленные в морфологически и синтаксически определенных конструкциях высказываний. Таким образом, в работе обосновано и реализовано системное (разноуровневое) описание дискурсивной формы существования естественного языка (именно в дискурсе может быть реализована коммуникативная функция языка).

Особая линия исследования ориентирована на применение психологических процедур и методик. Эти разделы диссертации позволяют дать оценку разработанной модели как инструмента для оценки качества модели. Предполагалось, что морфологические, синтаксические, семантические и дискурсивные характеристики текстов, предложенные в модели, могут соответствовать индивидуально-психологическим характеристикам авторов текстов. Для проверки этой гипотезы проведено два исследования. Первое из них – для разделения лиц с диагнозами эндогенных психических расстройств и аффективных состояний и «нормы» на двух выборках: клинической (22 человека) и популяционной (152 человека); второе – для разделения лиц с

симптомами депрессии и без таких проявлений, также на двух выборках: клинической (93 человека) и популяционной (233 человека).

Установлено, что в текстах лиц, составляющих контрастные выборки (клиника vs. норма), статистически значимо различаются значения лексических, морфологических, стилистических и семантических характеристик созданных ими текстов. Этот результат демонстрирует высокую степень валидности показателей, сконструированных для построения модели. Важно, что значения этих характеристик соответствуют клиническим оценкам состояния участников исследования. Специально следует отметить, что выявленные соответствия характеристик письменной речи и индивидуально-психологических характеристик участников исследования обеспечиваются именно использованным в работе *системным* характером соотношений разноуровневых психолингвистических характеристик организации текстов.

В оценке значимости полученных результатов для лингвистики следует согласиться с автором: построенный инструмент, учитывающий разноуровневое строение текстов, несомненно, имеет значительные перспективы использования для корпусных лингвистических исследований, и, шире, для статистического анализа больших объемов текстового материала.

Фундаментальная новизна диссертационного исследования И.В. Смирнова состоит в *системной формулировке* проблемы описания разноуровневого строения текстового материала. Эта формулировка позволила в значительной степени преодолеть трудности, которые возникают при попытках построения единого описания уровней как их *суммирования* в общую организацию. Новизна алгоритмических решений задач работы, предложенных автором, достаточно убедительно показана в текстах автореферата и диссертации. Принципиальной новизной обладает построенное системное описание разноуровневого строения текста. Следует отметить, что применение психологических и

психолингвистических вариантов проверки валидности построенной модели обеспечивает прикладную значимость результатов исследования.

Работа И.В. Смирнова представляет собой современное в методическом и методологическом отношении исследование и отвечает всем требованиям, предъявляемым к докторским диссертациям. Считаю, что Смирнов Иван Валентинович заслуживает присуждения ученой степени доктора технических наук по специальности 1.2.1 – «Искусственный интеллект и машинное обучение».

Доктор психологических наук,
главный научный сотрудник
лаборатории психофизиологии им. В.Б. Швыркова,
Институт психологии РАН



И.О. Александров

10 апреля 2024 г.



«Личную подпись Александрова И.О.
ЗАВЕРЯЮ: Александр Леонович Р.Р.
должность заместитель отдела кадров