

**РАЗДЕЛ 5
АКТУАЛЬНЫЕ НАПРАВЛЕНИЯ ПРИКЛАДНОЙ
ЛИНГВИСТИКИ**

**SECTION 5
TOPICAL ISSUES OF APPLIED LINGUISTICS**

**ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ И НОВЫЕ ТЕОРЕТИЧЕСКИЕ
ПОДХОДЫ В СЕМАНТИЧЕСКИХ ИССЛЕДОВАНИЯХ**

Ю. М. Амелина

*Белорусский государственный университет,
Минск, Беларусь, amelina@bsu.by*

В данной статье анализируются статистические методы и машинное обучение в применении их к большим корпусам текстов на разных языках, чтобы выявить скрытые закономерности в использовании слов и семантические отношения между ними. Также исследуется, то как данные результаты соотносятся с альтернативными теоретическими рамками, такими как теория эмбодированного познания и концепции, основанные на использовании языка, которые поставили под сомнение традиционные представления о кодировании значений.

Ключевые слова: исследования семантики; вычислительные методы; статистический анализ; машинное обучение; альтернативные теоретические подходы; теория эмбодированного познания.

**COMPUTATIONAL METHODS AND NEW THEORETICAL
APPROACHES IN SEMANTIC RESEARCH**

Y. M. Amelina

*Belarusian State University,
Minsk, Belarus, amelina@bsu.by*

This article analyzes statistical methods and machine learning when applied to large corpora of texts in different languages to reveal hidden patterns in the use of words and semantic relationships between them. It also examines how these findings relate to alternative theoretical frameworks, such as the theory of embodied cognition and language-based concepts that have challenged traditional notions of encoding meaning.

Keywords: semantics research; computational methods; statistical analysis; machine learning; alternative theoretical approaches; theory of embedded cognition.

Семантика является одной из самых динамичных и развивающихся областей лингвистики, которая изучает значение слов, выражений и тек-

стов. Знание семантики необходимо для понимания языковой коммуникации, когнитивных процессов, культурных различий и междисциплинарных связей.

В последние десятилетия семантические исследования претерпели значительные изменения под влиянием вычислительных технологий и новых теоретических концепций. Вычислительные методы позволяют анализировать большие объёмы данных, выявлять скрытые закономерности и семантические отношения, а также проверять и развивать различные теории. Новые теоретические подходы, такие как теория эмбодированного познания и концепции, основанные на использовании языка, поставили под сомнение традиционные представления о кодировании значений и предложили новые перспективы на семантику. Однако эти новые направления ставят перед исследователями новые вызовы и вопросы, которые требуют дальнейшего изучения.

Применение сложных статистических методов и моделей машинного обучения ускорило открытия в семантике. Обучая модели на огромных текстовых корпусах, исследователи теперь могут выявлять повторяющиеся закономерности в употреблении слов, раскрывающие скрытые регулярности в значении. Модели семантического распределения могут выводить значимые семантические отношения из статистического совместного появления слов в похожих контекстах. Алгоритмы кластеризации также группируют слова по общим значениям, позволяя автоматически открывать новые семантические категории.

Однако, хотя информация, основанная на данных методах проливает свет на структуру семантики, значение понятий не может быть полностью извлечено только из статистики. Как отмечает компьютерный лингвист Эмили Бендер, "значение возникает из взаимодействия умов друг с другом, объектами и опытом, а не из данных" [1, с. 2]. Интеграция вычислительных подходов с человеческим знанием остается важным условием для надежного понимания семантики.

Существующие теории значения пересматриваются с использованием современных подходов. Парадигма воплощенного познания утверждает, что семантические знания формируются в результате восприятия, двигательного опыта и эмоциональных переживаний, связанных с физическим миром [2, с. 617]. Это противопоставляется амодальным символическим моделям, где значение представлено абстрактно. Теории, основанные на употреблении, также предполагают, что семантическая структура возникает из воздействия языка в контексте и гибко адаптируется в различных сообществах и поколениях [3, с. 181].

В то же время, динамические подходы моделируют потенциал значения как многомерное "семантическое пространство", где слова определяются своей близостью друг к другу, а не фиксированными представлениями. Сочетание новых подходов с установленными методами позволяет углубить наше понимание семантики и расширить возможности ее приложений. Комбинирование вычислительных методов с теоретическими рамками также способствует более полному и точному моделированию сложных явлений, связанных с семантикой [4, с. 149].

Одним из ключевых направлений в современных исследованиях семантики является использование глубокого обучения и нейронных сетей. Глубокие нейронные сети позволяют моделировать сложные семантические взаимосвязи и извлекать более высокоуровневые представления слов и текстовых данных. Например, рекуррентные нейронные сети (RNN) и сверточные нейронные сети (CNN) успешно применяются для задач семантического анализа, включая определение тональности текста, классификацию документов и машинный перевод.

Благодаря применению глубокого обучения были созданы векторные модели слов, которые отражают слова в векторном пространстве, учитывая их смысловое сходство. Так, Word2Vec и GloVe – это известные методы генерации векторных представлений слов, которые позволяют кодировать слова с помощью компактных числовых векторов. Эти векторные модели сохраняют смысловую близость между словами, что дает возможность выполнять операции смысловой аналогии и выявлять смысловые связи. Например, в модели Word2Vec векторное выражение “король - мужчина + женщина” дает вектор, близкий к вектору слова “королева”. А в модели GloVe векторное расстояние между словами “кошка” и “собака” меньше, чем между словами “кошка” и “банан”, что отражает их семантическую близость.

Вместе с тем, новые теоретические рамки, такие как распределенная семантика и семантические сети, предлагают альтернативные способы представления значения. Распределенная семантика основывается на предположении о том, что семантическая информация распределена по всему языку и может быть представлена в виде векторных пространств. Семантические сети, в свою очередь, моделируют семантические отношения между словами в виде графа или сети, где узлы представляют слова, а ребра – семантические связи.

Однако, существуют и другие теоретические подходы к изучению значения, такие как распределенная семантика и семантические сети. Распределенная семантика предполагает, что семантический контент слова определяется его употреблением в языке и может быть выражен в виде вектора в многомерном пространстве. Семантические сети, напротив,

описывают семантические отношения между словами с помощью графовой структуры, где вершины соответствуют словам, а ребра – семантическим связям. Например, семантическая сеть для слова “собака” может состоять из различных типов семантических связей, таких как гипонимия, синонимия, антонимия и т.д.

Применение этих новых вычислительных методов и теоретических рамок к изучению семантики позволяет лучше понимать язык и создавать более точные семантические модели. Однако существуют нерешенные проблемы, такие как интерпретация полученных результатов и адаптация к культурным и контекстуальным особенностям семантики.

Необходимы дополнительные исследования в этих областях, чтобы расширить наши знания о семантике и использовать их в различных практических задачах, связанных с машинным переводом, информационным поиском и автоматической обработкой естественного языка. Например, существует метод машинного перевода, основанный на распределенной семантике и семантических сетях, который показывает лучшее качество перевода по сравнению с традиционными статистическими и нейронными подходами. Также в некоторых исследованиях рассматривается задача информационного поиска, в которой используются векторные модели слов для улучшения релевантности и точности поисковых запросов.

В заключение следует отметить, что, современные исследования семантики продвигаются вперед благодаря развитию вычислительных методов и новых теоретических рамок. Вычислительные подходы, такие как статистический анализ, машинное обучение и глубокое обучение, позволяют раскрыть скрытые семантические закономерности и создать эффективные модели представления значения. Использование эмбедингов слов и нейронных сетей позволяет извлекать семантические отношения и проводить анализ текстовых данных с высокой точностью.

Одновременно с развитием вычислительных методов, новые теоретические рамки, такие как распределенная семантика, семантические сети и воплощенная когнитивность, предлагают альтернативные подходы к представлению значения и объяснению семантических явлений. Эти рамки учитывают контекстуальные и культурные вариации в семантике и подчеркивают роль восприятия, опыта и взаимодействия с физическим миром в формировании смыслового знания.

Взаимодействие вычислительных методов и новых теоретических рамок позволяет углубить наше понимание семантики и расширить границы ее применения. Однако остаются вызовы, такие как объяснимость полученных результатов, учет разнообразия культур и контекстов, а также этические и социальные аспекты применения семантических моделей.

Успешное разрешение этих вызовов требует дальнейших исследований и улучшений в области семантики. Продолжение работы в направлении интеграции вычислительных методов с человеческим знанием и развитие теоретических рамок помогут нам создать более точные и объяснимые модели семантики, способные эффективно применяться в различных практических задачах.

В целом, современные исследования семантики открывают новые горизонты для понимания языка и его семантических аспектов. Прогресс в области вычислительных методов и теоретических рамок позволяет нам лучше понять и использовать семантическую информацию, что имеет важное значение для развития интеллектуальных систем, автоматической обработки естественного языка и межкультурного взаимодействия.

В заключение можно сказать, что современные исследования семантики предоставляют новые возможности для изучения языка и его смысловых характеристик. Благодаря развитию вычислительных методов и теоретических рамок можно лучше анализировать и применять семантическую информацию, что способствует созданию интеллектуальных систем, автоматической обработки естественного языка и межкультурного взаимодействия. Например, важно решение проблемы межкультурного взаимодействия в контексте семантической адаптации текстов к разным культурам и ценностям.

Библиографические ссылки

1. *Bender E. M.* Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics // San Rafael, CA: Morgan & Claypool Publishers. – 2020.
2. *Barsalou L.W.* Grounded Cognition. Annual Review of Psychology // Annual Reviews [Electronic resource] URL: <https://www.annualreviews.org/doi/10.1146/annurev.psych.59.103006.093639> (date of access : 06.01.2024).
3. *Clark A.* Whatever next? Predictive brains, situated agents, and the future of cognitive science // Cambridge University Press. 2013. № 3. P. 181-204.
4. *Куклев В.А.* Модели смешанного обучения // Высшее образование в России. 2018. № 6.