

Секция «Лингвистика: Текст и дискурс: проблемы анализа и интерпретации»

## К вопросу о создании когнитивной модели электронного гипертекста

Научный руководитель – Шестопалова Валентина Ивановна

*Шульгинов Валерий Александрович*

*Кандидат наук*

Дальневосточный федеральный университет, Восточный институт - Школа  
региональных и международных исследований, Владивосток, Россия

*E-mail: prostovalera@yandex.ru*

В нашем исследовании рассматриваются вопросы разработки и прикладного использования базы данных онтологического типа, включающей в себя структурные единицы электронного гипертекста. Под электронным гипертекстом мы понимаем коммуникативно-познавательную единицу нового типа, которая, с одной стороны, отвечает всем критериям текстуальности (целостность, связность, намеренность, приемлемость, информативность, ситуативность, интертекстуальность), с другой, характеризуется сложной структурой и нелинейными связями между фрагментами. Наличие ссылочного аппарата в структуре электронного гипертекста видоизменяет опыт его прочтения. Он становится «сенсорно-моторным» и даже «тактильным», так как пользователь одним движением курсора может совершить переход к новому текстовому фрагменту.

Рассматривая электронный гипертекст как мотивированную и целенаправленную коммуникативную единицу, отражающую установки и ценностные ориентиры говорящего, необходимо проделать обратный читателю маршрут: первичным окажется целевой текстовый фрагмент, а ссылка на него станет носителем вербализованной реакции автора. Такой подход позволяет осуществить анализ системы личностного сознания как функционального базиса текстовой деятельности автора и, в конечном счёте, построить комплексную когнитивную модель электронного гипертекста.

Построение данной модели включает в себя несколько этапов: сбор корпуса гипертекстов, предобработка текстовых фрагментов, аннотирование полученных данных (исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-312-0010).

На этапе сбора данных перед нами стоит задача формализации связей и отношений в структуре электронного гипертекста. Мы выдвигаем гипотезу о существовании минимальных гипертекстовых единиц («гипертекстем»), представляющих собой трехчастную структуру (исходный текст - ссылка - целевой текст). Вычленение структурных единиц электронного гипертекста позволяет применить корпусный подход и составить корпус гипертекстем [Шульгинов 2016: 233].

Для обеспечения однородности материала на начальном этапе работ мы ограничили целевые тексты по жанровому признаку. В качестве основных источников материала были выбраны новостные статьи, опубликованные в электронных СМИ («РБК», «Lenta.ru», «Meduza», «ТАСС»). Поиск и сбор гипертекстем осуществляется в автоматическом режиме с помощью парсера.

На следующем этапе происходит предобработка текстовых фрагментов для проведения автоматического анализа:

1. Лемматизация целевого текста и организация списка входящих в него лемм;
2. Сортировка лемм по частотности употребления и выделение ключевых для данного текста слов;
3. Лемматизация номинаций ссылок и выявление степени семантической близости в паре «ссылка - ключевое слово».

На этапе аннотирования полученных данных центральной задачей является выявление степени семантической близости между узловыми компонентами гипертекстом. При этом необходимо учитывать влияние лингвистических (лексическая и морфологическая характеристика ссылок, семантика гипертекстового перехода) и экстралингвистических (жанр, тематика текстов, пространственно-временные связи между фрагментами, тональность коммуникации, возраст и пол автора) параметров.

В основе модуля, определяющего степень семантической близости между номинацией ссылкой и содержанием целевого текста, лежит семантическая классификация ссылок Р.К. Потаповой и О.В. Деловой, согласно которой ссылки функционируют в парадигме «референциальные/семантические ссылки» [Дедова 2008: 201].

Для определения количественных показателей семантической близости между номинацией ссылки и ключевыми словами мы задействовали открытый фреймворк «WebVectors», позволяющий строить векторные модели дистрибутивной семантики слов. Благодаря компьютерным технологиям мы можем сопоставить большой объём контекстов и установить степень совпадения дистрибуции отдельных единиц в тексте: «vector models of distributional semantics are well established in the [U+FB01]eld of computational linguistics and have been here for decades. . . However, recently they received substantially growing attention. The main reason for this is a possibility to employ arti[U+FB01]cial neural networks trained on large corpora to learn low-dimensional distributional vectors for words [Kutuzov 2017: 155].

В результате автоматической обработки мы выявляем показатель семантического сходства номинации ссылки с каждым ключевым словом, а также усреднённый показатель по всему тексту. Диапазон значений составляет от 0 до 1, где 0 - это абсолютное отсутствие семантических пересечений, а 1 - полное совпадение контекстов, в которых используются слова (а значит, и их значений).

Электронный гипертекст обладает рядом признаков, позволяющих соотнести его с мышлением человека: он обладает многообразием, множественностью, вариативностью, фрагментарностью, открытостью, а также ассоциативностью с структуры. Такая природа электронного гипертекста определила технологические пути его развития: в настоящее время центральным процессом в эволюции электронного гипертекста становится семантизация сети. Трансформация глобальной паутины в семантическую позволяет представить сеть в качестве корпуса, с помощью которого возможно решение самых разных лингвистических, когнитивных и методических задач.

### Источники и литература

- 1) 1. Дедова О.В. Теория гипертекста и гипертекстовые практики в Рунете. М.: МАКС Пресс, 2008. – 284 с.
- 2) 2. Шульгинов В.А. Когнитивная модель электронного гипертекста. Вестник Кемеровского государственного университета, № 4, 2016. – 233-238 с.
- 3) 3. Kutuzov, A., Kuzmenko, E. (2017) WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, vol 661. Springer, Cham