

ПРИКЛАДНАЯ ЛИНГВИСТИКА

Н. А. Богданова

РАЗГРАНИЧЕНИЕ МИКРОТЕМ ПРИ ПОСТРОЕНИИ ЛОГИКО- СЕМАНТИЧЕСКОЙ ФОРМУЛЫ ТЕКСТА

Тема текста является его первейшим и основным параметром. Будучи сверткой, она всегда может быть развернута в полный связный текст. Большинство исследователей правомерно отмечает, что тема текста не является чем-то аморфным, нерасчлененным. Она включает составные части, называемые субтемами, подтемами либо микротемами, текстовыми блоками, топиками, партитекстами. Можно сказать, что тема текста представляет собой иерархическую структуру, состоящую из более мелких единиц, которые соединяются между собой не хаотично, а по определенным правилам, обеспечивая ту или иную форму тематического развертывания.

Для решения задач моделирования содержания текста необходимо использование денотативного подхода, который опирается на концептуальные структуры. С позиций этого подхода основной единицей содержания является денотат, представляющий собой образ фрагмента действительности. Содержание структурно, так как в нем можно выделить дискретные элементы и отношения между ними. Денотатная структура текста предполагает выделение ключевых, наиболее важных элементов текста (слов, предложений), по которым можно судить о его содержании. Одна из первых методик компрессии текста с использованием ключевых слов была разработана Л. В. Сахарным. Суть ее состоит в том, что к каждому развернутому тексту в процессе его компрессии могут быть получены наборы ключевых слов различного объема, а в предельном случае компрессия может привести к выделению одного основного ключевого слова. Иерархия «развернутый текст – набор ключевых слов – основное ключевое слово» отражает смысловую структуру текста. Таким образом, содержание текста – это эксплицитно выраженная при помощи ключевых слов информация текста, являющаяся основой его темы.

Схожие композиционные принципы построения текстов, принадлежащих одному речевому жанру, можно обобщить в виде логико- семантической формулы текста (ЛСФТ). Метаязык построения ЛСФТ опирается на:

- ключевые слова текста;
- речевые регистры текстовых фрагментов;
- рематические доминанты текстовых фрагментов.

Выделение ключевых слов происходит автоматически на основе алфавитно-частотного словаря текста. Слова распределяются по категориям: главные опорные слова, второстепенные опорные слова и прочие слова с учетом их частотности и покрытия текста. Речевые регистры и рематические

доминанты текстовых фрагментов выявляются в процессе анализа текста исследователем. Базовое правило выделения микротем опирается на следующие критерии:

- состав главных / второстепенных / прочих опорных слов остается неизменным;
- остается неизменной принадлежность фрагмента речевому регистру;
- рематическая доминанта текстового фрагмента не меняется.

Таким образом, смена микротемы происходит в случае смены состава действующих лиц, смены способа подачи информации, а также смены типа ремы фразы. Рассмотрим несколько примеров, взятых из немецкого короткого рассказа Альфреда Андерша «Ein Vormittag am Meer».

«Er streifte seine Hose ab, zog die Badehose an, dann die Hose wieder darüber». В данном случае речь идет о действиях главного героя, пришедшего на пляж. Единство микротемы обусловлено однородностью акциональной рематической доминанты, принадлежностью к повествовательному подрегистру изобразительного регистра и отнесенностью действий к главному действующему лицу. Такая микротема получает уникальный код и используется во всех случаях, где встречается схожий текстовый фрагмент.

Следующий случай иллюстрирует смену регистра повествования: *«Ja, aber, hatte sie gesagt, du zahlst an Zinsen insgesamt mehr als den Betrag der Hypothek, das Haus kostet dich also mindestens das Doppelte von dem, was sein Bau wirklich gekostet hat. Seine Frau war vor ihrer Heirat Sekretärin gewesen».* Диалог главных героев прерывается информацией автора о профессии главной героини. Данная информация относится к информативному регистру, так как в отличие от фрагментов описательного регистра не участвует в непосредственном сюжетодвижении. Информативный регистр предлагает сообщения о фактах, событиях, свойствах, поднимающихся над наблюдаемым в данный момент, отвлеченных от конкретной длительности единичного процесса. Такой же случай представлен и в следующем фрагменте: *«Entschlossen ging er auf sie zu; er wollte heute unbedingt schwimmen. Wenn der gelbe Wimpel flatterte, war es eigentlich verboten zu schwimmen».* Информация о действиях и желаниях главного героя прерывается пояснением автора об обстоятельствах отдыха на море, причем не в данном конкретном случае, а вообще, поэтому вместе со сменой регистра происходит смена микротемы. Здесь наблюдается явная ретардация повествования, т.е. замедление динамики сюжетодвижения. Данное замедление может объясняться как логикой самого повествования, так и особенностями авторского стиля.

В следующем примере микротема меняется в рамках одного предложения, так как происходит смена состава действующих лиц: *«Die beiden kleinen Mädchen sprangen erfreut auf, als er kam, und er ging mit ihnen eine halbe Stunde lang am Strand spazieren, half ihnen, Muscheln zu finden».* В данном сложном предложении вначале речь идет о действиях дочерей главного героя, а затем о действиях самого главного героя. Несмотря на единство регистра и рематических доминант происходит смена действующего лица, поэтому меняется и микротема.

Не всегда удастся точно определить границы микротема. Проблемные случаи разграничения микротема можно классифицировать на основании вызывающих их причин. С одной стороны, это причины, связанные с автоматической обработкой текста компьютером. Выделение главных и второстепенных слов текста происходит на основании коэффициента важности слова, который вычисляется для каждого слова на базе алфавитно-частотного словаря текста. Этот словарь, в свою очередь, также строится автоматически, однако для более точного отображения результатов требуется предварительная обработка текста и постредактирования словаря. От того, как именно будет происходить предобработка и постредактирование словаря, зависит список гласных и второстепенных опорных слов. С другой стороны, есть причины, связанные с лексической, синтаксической и графической организацией текста, когда возникают вопросы отнесения фрагмента к рематической доминанте, смене регистра внутри одного предложения или выделения отдельных слов в предложении.

Решение проблемных случаев в целом видится двояким и зависит от подхода исследователя. Первое решение, основанное на информационном подходе к тексту, будет исходить из того, что текст является носителем информации и эта информация представлена так или иначе в опорных словах текста, выявляемых статистически, поэтому все единицы, содержащие схожую информацию, должны объединяться в одну группу и частотность упоминания каждой единицы должна суммироваться со схожими. Другое решение будет опираться на текст как отражение авторской интенции, авторского видения мира. В данном случае необходимо будет привести частотные показатели токенов в соответствие с авторским видением картины мира и выделить разные аспекты упоминания одних и тех же референтов в разные токены с разной частотностью. Тогда может оказаться например, что упоминание мужа и жены как пары обладает большей частотностью, чем их упоминание по отдельности, что является существенной особенностью как для интерпретации текста, так и для описания идиостиля автора.