Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное

образовательное учреждение высшего образования

**«Московский государственный лингвистический университет»**

**(ФГБОУ ВПО МГЛУ)**

**Аннотация**

к выпускной квалификационной работе

Г**удкова Никиты Андреевича**

факультет английского языка

Направление подготовки – 45.03.02 -Лингвистика

Направленность (профиль) — Теоретическая и прикладная лингвистика

группа 5-8-32

на тему:

**ОСОБЕННОСТИ ОПРЕДЕЛЕНИЯ ТЕМЫ ТЕКСТА, ВОСПРИНЯТОГО В УСЛОВИЯХ ПОМЕХ**

***Ключевые слова:*** смысловое восприятие, затрудненные условия, определение темы сообщения, кластерный анализ.

Данная выпускная работа состоит из введения, двух глав с выводами после каждой главы, заключения, списка отечественной и иностранной литературы, аннотации и приложения. Первая глава по анализу литературы, касающейся вопросов восприятия речи в затрудненных условиях, состоит из 6 разделов: восприятие как многоплановый процесс; проблемы восприятия речи; влияние шумов и помех на восприятие речи человеком; устно-речевой дискурс; различные подходы к выделению темы текста и методы обработки текста. Глава об эксперименте по определению темы сообщения, воспринятого в шумах, содержит 4 раздела: материалы и методики для дальнейшего эксперимента; дикторы и аудиторы; методика проведения эксперимента и анализ данных и полученные результаты. Материалы работы изложены на 81 странице машинописного текста, иллюстрированы 2 таблицами, 3 диаграммами. Приложение включает также 3 варианта инструкции, 3 таблицы с результатами распределения тем по группам, пример списка распознанных слов, предоставленного аудитором и исходные записи диалогов.

Смысловое восприятие сообщения в затруднённых условиях - это сложный многоуровневый иерархический процесс, изучению которого по настоящие время уделяется большое внимание со стороны специалистов из разных областей науки. До сих пор не существует единой теории и общепринятых моделей процессов восприятия речевого сообщения. Нерешенной проблемой является и зависимость восприятия от формы продуцирования речи, от речевых характеристик говорящего, от условий, в которых происходит общение и восприятие информации человеком, от фоновых знаний говорящего и слушающего, а также от самого типа информации. **Актуальность** данной работы заключается в необходимости решения вышеуказанных задач и разработки рекомендаций по совершенствованию работы в рассматриваемой области.

Множество экстралингвистических факторов влияют на восприятие речи. В данной работе стоит упомянуть об одном из таких факторов, а именно о затруднении восприятия из-за наличия в сигнале различных шумов и помех.

Настоящее исследование посвящено такому немаловажному аспекту распознавания вербального контента, как определению темы сообщения. У понятия темы существует много различных трактовок, которые можно разбить на два основных направления: первое, связанное с актуальным членением предложения (высказывания), и второе, описывающее смысловую структуру текста.

Актуальное членение [В. Матезиус 1967] выявляет способ включения предложения в предметный контекст, на базе которого оно возникает. Основными элементами АЧП являются исходная точка (основа), то есть то, что известно или можно понять в конкретной ситуации высказывания, и ядро высказывания, что несёт в себе сообщение говорящего об исходной точке высказывания. Темой в данном случае является то, о чём говорится в предложении (высказывании).

Во втором значении тема связана с таким понятием, как сложное синтаксическое целое. ССЦ – это сложное структурное единство (смысловой, тематический блок), состоящее более чем из одного предложения, обладающее смысловой целостностью в контексте связной речи [Г.Н. Поспелов 1948].  Тема целого речевого произведения в данном случае понимается как продукт синтеза микротем, каждая из которых представляет собой тему сложного синтаксического целого (ССЦ)[[1]](#footnote-2).

В качестве **объекта** данной работы выступает смысловое восприятие речевого сообщения, а **предметом** изучения служат особенности определения темы сообщения, воспринятого в условиях помех.

**Основной целью** исследования являлось установление принципов выделения темы текста и их зависимости от степени зашумления.

**Главной задачей работы** является применение метода кластерного анализа на материале текстов, воспринятых в условиях шума. В основе данного метода лежит упорядочивание выборки данных в сравнительно однородные группы для последующего анализа внутреннего состава образовавшихся групп и принципов их разбиения.

**Перспективность** исследования связана прежде всего с тем, что рассматриваемый метод кластерного анализа до настоящего момента не применялся при изучении процессов определения темы сообщения, в особенности на основе данных, полученных в результате эксперимента по восприятию речи в затрудненных условиях.

Переходя непосредственно к проведенному исследованию, стоит отметить, что работа проводилось на кафедре прикладной и экспериментальной лингвистики ФГПН МГЛУ в рамках государственного заказа НИР по теме «Теоретико-экспериментальное исследование распознавания вербального контента устной речи в затрудненных для восприятия условиях». Научным руководителем НИР является доктор филологических наук, профессор Потапова Родмонга Кондратьевна.

**Методика проведения эксперимента**

Методика проведения эксперимента заключалась в следующем: сначала был проанализирован экспериментальный материал, полученный ранее в рамках НИР, шифр «Оглавление ПО» (руководитель - доктор филологических наук, профессор Р. К. Потапова). Он состоял из аудиозаписей (диалоги на такие темы, как спорт, музыка, отдых) с разным соотношением сигнал-шум (0 дБ, -10 дБ и -20 дБ) и отекстованных фрагментов или списков распознанных слов (один список на один диалог от каждого аудитора). На основе прослушивания диалогов участники ранее проведенного эксперимента пытались определить тему текста или выписывали возможные варианты определения основной темы и подтем. Выделенные испытуемыми темы были представлены в виде списка. Позже другой группе информантов (студенты 3 курса МГЛУ с направления Теоретической и прикладной лингвистики) было предложено придумать способ разбиения тем внутри каждой группы так, чтобы элементы из разных пунктов не пересекались, то есть, например, чтобы Музыка не оказалась в одном ряду со Спортом. Для каждого выделенного списка тем нужно было сформулировать название. Те темы, которые не удавалось никуда отнести, стоило заносить в категорию Остальное. Таким образом, информанты разбивали группы на категории. Для объективности полученных данных, термин «категория» в ходе эксперимента не употреблялся.

**Методика обработки результатов**

В начале был проведен анализ точности определения тем информантами предыдущего исследования по критериям соответствие/несоответствие, сужение/расширение и опознана/не опознана тема. Также полученный ранее материал был проанализирован с точки зрения формальной организации выделенных тем. Все используемые элементы разбились на три типа: изолированные слова, словосочетания и предложения. Наконец, была установлена принадлежность тем к различным частям речи. Все лексические составляющие относились к следующим частям речи: существительное, глагол, прилагательное, наречие, местоимение, союз, предлог и числительное.

Что касается непосредственно эксперимента с разбиением тем на группы, при обработке полученного материала было выдвинуто 3 вида условий:

1. Сравнить, как испытуемые разбивали темы в разных режимах шума
2. Насколько менялись темы
3. Какими были принципы группировки, а также сходства и отличия данных принципов.

Результаты анализировались следующим образом:

* во-первых, нужно было посчитать количество сформулированных групп каждым испытуемым;
* во-вторых, делались предположения относительно принципов разбиения:
* в-третьих, анализировалось, насколько единообразно выделение;
* в-четвёртых, была сделана попытка объединения некоторых сформулированных испытуемыми групп;
* в-пятых, отслеживалось, что получилось в результате такого объединения.

**Основные результаты работы**

На основе полученных результатов можно сформулировать следующие предварительные выводы:

1. Количество опознанных тем составило при -20 дБ 97%, при -10 дБ – 86%, при 0 дБ – 85%, иными словами оно уменьшается в зависимости от отношения сигнал/шум. Тема соответствует в большинстве (94%, 93% и 94% соответственно) случаев. Почти в половине случаев наблюдается сужение темы (43% при -20 дБ, 43% при -10 дБ и 44% при 0 дБ). Что касается расширения темы, то процент встречаемости данного явления крайне мал (5%, 4% и 4%). Тема выделяется верно в половине случаев (52%, 53% и 52% соответственно).
2. По частям речи наибольшее количество составляют существительные (170, 173 и 180 раз при -20, -10 и 0 дБ соответственно), далее идут предлоги (14, 26 и 27 раз), прилагательные (17, 18, 22 случая) и союзы (11, 15, 17 раз), редко использовались местоимения (3, 4 и 4 случая) и причастия (2, 3, 3 раза) и реже всего употреблялись наречия и глаголы (1-2 раза), а также числительные (всего 1 случай употребления при 0 дБ).
3. Что касается формальной организации формулировок тем, здесь ситуация следующая: чаще всего участники эксперимента прибегали к изолированным словам, примерно в 2 раза реже – к словосочетанием и практически не использовали предложения. Последнее предположительно связано с тем, что составление предложений занимает больше времени и сил, а также с тем, что, по мнению испытуемых, предложения могли показаться избыточными по своему смысловому содержанию и редко употребляемыми в качестве темы сообщения.

Выводы по эксперименту:

1. Испытуемые разбивали темы на группы в соответствии с определенными принципами.
2. Несмотря на разнообразие способов группировки, можно выявить некоторые схожие черты. Большинство информантов сначала определяли основную предметную область, которую после сужали путем указания дополнительных характеристик, таких как место, о котором идёт речь, виды деятельности, этапы процесса или явления, отношение к нему, конкретный аспект внутри данной области, связи с другими областями и т.д. Некоторые участники эксперимента использовали при этом распределение по бинарным признакам или противопоставления типа «ЕГЭ или стандартный экзамен по билетам?».
3. Производя разбиение, каждый испытуемый несомненно руководствовался личным опытом, опирался на свои знания вопроса и свою картину мира, тем самым представляя себе контекст, при котором определялась тема. Поэтому в названиях групп встречаются понятия и характеристики, которые непосредственно не указаны в сформулированных темах, но о которых возможно идёт речь в аудиоматериале.
4. По результатам эксперимента не удалось установить, как наличие разных режимов шума влияет на принципы разбиения тем и влияет ли вообще.
5. Существует определенное количество базовых категорий, которые выделяют большинство испытуемых вне зависимости от режима шума и формулировок тем.

**Заключение**

Таким образом, были выявлены некоторые принципы выделения тем, и было установлено, что метод кластерного анализа применим к данному материалу, а значит, имеет смысл продолжать изучение данного вопроса. Однако в ходе исследования не удалось определить зависимости между режимом шума и принципом разбиения. Это может быть вызвано недостаточным количеством испытуемых, привлеченных для проведения исследования, неэффективностью методики в данном аспекте, либо отсутствием какой-либо зависимости.

В дальнейшем планируется исследование организации других видов экспериментальных текстов. Наложение различных видов шумов на аудиозаписи, использование большого количества вариантов соотношений сигнал/шум и привлечение новых дикторов и аудиторов с разными социальными и лингвистическими характеристиками может существенно увеличить валидность эксперимента и поможет достичь объективных и значимых результатов. На основе полученных данных станет возможным представить структуру выделения темы текста в виде дерева.

Полученные результаты имеют значение для изучения смыслового восприятия и могут быть применены в сфере когнитивных наук. Данный эксперимент является пилотным и выполняется в рамках договора НИР.

Студент: Н. А. ГУДКОВ

Согласовано:

д-р филол. наук М.В. ХИТИНА

1. См. подробнее работы [М.Я. Дымарский 2001], [О.И. Москальская 1981] и др. [↑](#footnote-ref-2)