

КЛАССИФИКАЦИЯ КАК МЕТОД ИЗУЧЕНИЯ ВАРИАТИВНОСТИ ВОСПРИЯТИЯ ТЕКСТА

Ирина Г. Овчинникова

Пермский государственный университет, Россия

The paper is dedicated to the discussion of classification and cluster analysis as methods of text perception studies in linguistics and psycholinguistics. It emphasizes in particular that naïve classifications of speech material make clear the criteria for class formation. In cognitive psychology two types of classifications were determined: hierarchical and situational. The last one is preferred by subjects in the experiment of free classification of texts and slogans. Thanks to our experiment we found out that the main criteria for class formation is the topic of a text. The errors and mistakes do not play significant role in text perception in mass media discourse.

*"...животные делятся на
а) принадлежащих Императору,
б) набальзамированных,
в) прирученных,
г) сосунков,
д) сирен,
е) сказочных,
ж) отдельных собак,
з) включенных в эту классификацию,
и) бегающих как сумасшедшие,
к) бесчисленных,
л) нарисованных тончайшей кистью из верблюжьей шерсти,
м) прочих,
н) разбивших цветочную вазу,
о) похожих издали на мух".*
Хорхе Луис Борхес

Классификацию считают одним из основных способов интеллектуальной переработки поступающей информации: «процессы классификации являются конкретизацией интуитивного познания и основополагающей частью мышления» (Малиновский 1986: 155). Исходным для классификации оказывается понятие класса. При объединении объектов в класс интуитивное познание ориентируется на «разумные способы группировки», в которых отражена внутренняя структура ситуации (Вертгеймер 1987: 160). Объединение в классы и подклассы (группы и подгруппы) предполагает сопоставление объектов, их сравнение или различение. В свою очередь, сопоставление может осуществляться как на основе совпадения некоторых частей или свойств объектов, так и на основе целостного представления об объектах. Объединение в класс на основе совпадения частей или свойств объектов приводит к созданию иерархических группировок. В таком случае понятие класса соотносимо с понятием категории. Классификация позволяет установить отношения «категория – член категории» (Фрумкина 2001: 67). Иначе говоря, категория – ментальное образование, формирующееся в результате классификации объектов. Онтологически более ранней классификацией является ситуативная группировка, при которой в один класс включаются относящиеся к одной теме объекты; ситуативная группировка основана на целостном представлении об объекте (Брунер 1977: 328). Ситуативная группировка более опосредованно связана с категоризацией. Она предполагает объединение объектов в классы на основе их функции или смысла в деятельности индивида, т.е. отражает операцию концептуализации. В познавательной деятельности человека категоризация и концептуализация переплетаются, поскольку изменение ситуации или целей деятельности

субъекта предполагает возможность трансформации ситуативных группировок в иерархические. Предпочтение иерархической или ситуативной группировки обусловлено культурной традицией. Любая классификация имеет право на существование, позволяя решать различные задачи. В частности, ситуативная группировка лексики оказывается релевантной при обучении языку. Это обусловлено как особенностями коммуникации, так и спецификой психологической группировки языкового материала (Фрумкина 1984).

В сложившейся психолингвистической традиции при помощи классификационных методов изучают спонтанные речевые стратегии и психологические группировки языкового материала (Алгоритмы обработки экспериментальных данных 1986; Налимов, Дрогалина 1995; Петренко 1988; Тулдава 1987; Фрумкина 1984 и др.). Использование классификационных методов позволяет выявить явно не представленные в коммуникации способы категоризации и концептуализации языкового материала. В статистике классификация наиболее явно репрезентируется в методах деревьев классификации и кластерного анализа. Оба метода представлены в стандартном программном обеспечении (SPSS, Maple). Деревья классификации – это метод, позволяющий предсказывать принадлежность наблюдений или объектов к тому или иному классу (категории) в зависимости от соответствующих значений одной или нескольких переменных. Деревья классификации представляют собой один из методов непараметрической статистики. Построение деревьев классификации – один из наиболее важных методов, используемых при систематизации данных, поскольку позволяет выявить скрытую структуру анализируемого материала (см.: http://proxy/home/textbook/glossary/gloss_d.html). Термин «кластерный анализ» в действительности включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т.е. развернуть таксономии (<http://www.statsoft.ru/home/textbook/modules/stduan.html>; http://www.statsoft.ru/home/textbook/glossary/gloss_a.html). Кластерный анализ позволяет оценить принадлежность объекта к двумерной или многомерной совокупности, что дает возможность при сравнении различных классификаций одних и тех же объектов установить структуру наблюдаемых данных. Основная функция кластерного анализа – выявление структуры исследуемого материала. В лингвистике и психолингвистике кластер-анализ используют при обработке результатов эксперимента «свободная классификация».

Экспериментальная методика «свободная классификация» предполагает достаточно жесткую процедуру. Испытуемые работают индивидуально, каждому предлагается для классификации исходный языковой материал, время выполнения задания ограничивать не принято. Стандартная формулировка задания подробна, но не категорична. Обычно используют предложение разложить материал на «кучки», избегая более терминологических слов «класс» или «группа», что может натолкнуть участников эксперимента на построение иерархической классификации как единственного способа решения поставленной задачи. Количество кучек и элементов в каждой кучке обычно не ограничивают. Полученную классификацию фиксируют на бланке (записывают в отдельный файл в случае компьютерной версии эксперимента).

Экспериментальное исследование по методике «свободная классификация» позволяет получить так называемые наивные классификации, отражающие ощущения носителей языка о сходстве или несходстве значений слов и высказываний (Фрумкина 1984). Наивные классификации позволяют уяснить скрытую структуру языкового материала и выявить алгоритмы его разбиения на классы, установить критерии определения сходства и различия в исходной для классификации совокупности объектов. В свою очередь, знание критериев, по которым определяется сходство, дает возможность выяснить значимость признаков сравниваемых объектов для решения различных задач в речевой и мыслительной деятельности человека.

Покажем на примере возможности решения прикладной задачи классификационным методом. Наше исследование нацелено на выявление критериев определения сходства

текстов носителями языка и установление значимости ошибок в текстах для эффективности коммуникации. Основным методом исследования – эксперимент «свободная классификация». Исследование выполнено на материале рекламных текстов, опубликованных региональными СМИ. Примем в качестве исходной гипотезы: ошибки существенно затрудняют коммуникацию и влияют на группировку текстов в рамках дискурса СМИ.

Особую важность отбора стимульного материала для классификации подчеркивает Р.М. Фрумкина. Исследователь считает, что в некоторых экспериментах «материал был так заранее классифицирован авторами, что достаточно сильная суггестия о принципах классификации заключалась именно в нем. Во всех этих экспериментах экспериментатор заранее представлял себе, каким должно быть “идеальное” поведение испытуемых и тем самым как *следует* классифицировать материал. Иными словами, хотя экспериментатор не включал в инструкцию указаний о том, чем следует руководствоваться в процессе классификации, чтобы получить правильный результат, но сам материал был структурирован так, что некий “правильный” результат существовал» (Фрумкина 1984: 87). В нашем случае таким «правильным» (или ожидаемым) результатом классификации может считаться группировка текстов в зависимости от наличия или отсутствия в них ошибок, в зависимости от типа ошибок. Снизить суггестию материала можно за счет включения в исходный для классификации набор выборки текстов одного размера, жанра, близких по стилю, функционирующих в одной сфере коммуникации и принадлежащих одному типу дискурса, но разнообразных по тематике.

Следует учитывать и объем предъявляемого материала. Как бы то ни было, текст – это особое языковое образование. Восприятие, осмысление и оценка текста, даже самого короткого, характеризуются не простой трехуровневой моделью (сенсорный – перцептивный – смысловой уровни восприятия), но и весьма сложной структурой последнего, смыслового, уровня (см. подробнее: Штерн 1992: 5). Задание на свободную классификацию текстов по степени сходства предполагает осмысление каждого из предъявленных текстов, на основе чего испытуемые и определяют – осознанно или неосознанно – признаки, по которым возможно установить сходство. В результате свободная классификация текстов оказывается довольно трудной интеллектуальной задачей, если текстов много, если они длинные и сложные. Это накладывает еще одно ограничение на выбор стимульного материала – количественное, т.е. текстов не должно быть много, они не должны быть длинными и сложными.

В результате мы остановились на рекламных текстах-слоганах. Рекламный слоган характеризуется достаточно жестким фреймом, четким и недвусмысленным обозначением темы (предмета рекламы), краткостью и стереотипными приемами (Шенерт 2001). Благодаря этому выбор рекламных текстов в качестве стимульного материала позволяет достаточно легко и естественно ограничить не только жанр, стиль и объем, но и «художественность» (точнее – «квазихудожественность») текстов. Стимульный материал включал 20 карточек, на каждой из которых был напечатан один рекламный текст. Длина текстов не превышала 12 знаменательных слов. В некоторых из предложенных для классификации рекламных текстах встречались лексические (*Стандартный размер спреда – 3-7 пинсов*), семантические (*5 лет живи без бед!*), синтаксические (*Средство для ухода за ногами из Англии; Ваше строительство обеспечит трест «ДОМ» бетонными изделиями*) и стилистические (*Продукция «Хлебокомбината №1» – счастье в доме*) ошибки.

Испытуемым предлагалась стандартная инструкция: «Разложите предложенные вам 20 карточек с текстами по кучкам таким образом, чтобы в одной кучке оказались похожие друг на друга тексты. Число кучек может быть любым, число текстов в кучке – также. Каждый текст может помещаться одновременно только в одну кучку. Дайте условные названия кучкам. Запишите получившуюся классификацию на бланке, указывая название кучки и номера входящих в нее текстов».

В нашем эксперименте приняли участие 83 студента Пермского государственного университета. Результаты эксперимента обрабатывались методом кластер-анализа. Была

выбрана иерархическая процедура кластеризации агломеративного типа. «Принцип работы агломеративных алгоритмов состоит в последовательном объединении в кластер сначала самых близких, а затем и все более отдаленных друг от друга объектов» (Тулдава 1987: 181). В основу кластеризации положен признак субъективного сходства текстов для носителей языка. Этот признак совпадает с основанием для классификации текстов, заданным в эксперименте. Напомним: в экспериментальном задании ни в коей мере не уточнялось основание для установления сходства (т.е. критерий сходства), оно выбиралось самими испытуемыми, осознанно или неосознанно. Таким образом, мы используем (а отчасти и преодолеваем) субъективность кластер-анализа: благодаря кластеризации измеряется сходство текстов по произвольно выбранным носителями языка критериям, а не устанавливаются «объективные» критерии сходства (см. подробнее: Овчинникова 2002).

Обсудим на примере полученное объединение текстов. «Ближайшим соседом» слогана *Почувствуй себя богиней* является слоган *Ведь ты этого достойна*. Расстояние между ними равно 0,41: это означает, что почти 60% участников эксперимента поместили тексты в один класс на основе тематической общности и общности художественных приемов. В свою очередь, «ближайшим соседом» слогана *Ведь ты этого достойна* оказывается текст *Ваши ноги будут Вам благодарны*; расстояние между ними заметно больше, чем в предыдущем случае: 0,74 (треть участников эксперимента объединили тексты в класс). Между тем последний слоган оказывается более тесно связан с текстом *Средство для ухода за ногами из Англии*: тематическое сходство между текстами оценила половина испытуемых. Заметим, что речевая ошибка в последнем тексте не повлияла на его группировку. Таким образом, мы обнаружили относительно автономный кластер текстов, связанных с остальными за счет последнего слогана, содержащего речевую ошибку. Испытуемые практически единогласно назвали эту группу текстов *Товары для женщин*, вариации незначительны: *Обращенные к женщине*, *Женщина и красота* и т.п.

В результате обработки 83 классификаций мы установили, что наличие ошибок в тексте все же, пусть незначительно, но влияет на установление его сходства с другими текстами. Тексты с грубыми ошибками тем или иным образом выделяют в отдельный класс не более 20% испытуемых, обозначая его как *Реклама ни о чем*, *Нелепость* и т.п. Например, один из участников эксперимента объединил слоганы *Наши цены дешевле всех*, *Продукция Хлебокомбината №1 – счастье в доме*, *Средство для ухода за ногами из Англии*, обозначив группу как *Малограмотные*.

В названии групп отражен критерий выделения текстов: испытуемые считают основанием для объединения в одну группу содержательную характеристику, а не наличие ошибок. Как правило, тексты с ошибками присоединяются к кластерам на сравнительно большем расстоянии, чем тексты без ошибок. Большее расстояние означает, что немногие испытуемые склонны присоединять их к уже выделенным тематическим группам. Это можно интерпретировать как осознание их несходства с другими текстами близкой тематики. Большее расстояние свидетельствует о том, что тексты с ошибками разные испытуемые классифицируют по-разному, руководствуясь различными критериями.

В целом классификация рекламных слоганов оказалась довольно диффузной; различные участники эксперимента один и тот же текст помещали в разные кучки, в соответствии с различными критериями. Например, слоган *Почувствуй себя богиней* 15% испытуемых объединили с текстами *Изменим жизнь к лучшему*, *Бери от жизни все* в класс, названный *Побуждение*; в этом случае тексты объединены на основе типа речевого акта.

Типичным основанием классификации оказывается тематический принцип: в одну группу попадают тексты, рекламирующие сходные товары. Вторым по популярности основанием классификации являются речевые приемы (лексические повторы, рифмы и т.п.). Тем не менее мы вынуждены констатировать, что исходная гипотеза не подтвердилась: ошибки не оказывают существенного влияния на смысловую переработку речевого сообщения в рекламном дискурсе и эффективность коммуникации. Решающее значение для восприятия и оценки рекламного сообщения имеет субъективная значимость

рекламируемого товара или услуги, т.е. реклама оценивается на основе экстралингвистического критерия, имеющего косвенное отношение к коммуникативному акту. Классификация в таком случае основана на операции концептуализации, т.е. на осмыслении значимости текста в деятельности индивида. Естественно, полученные результаты не могут быть распространены на иные курсы.

Тематическое объединение основано на выделении существенного признака текста – выраженной ключевым словом темой или предметом рекламы. Установление сходства текстов на основе общности речевых приемов также, на наш взгляд, отражает аналитический подход к тексту: в качестве существенного признака испытуемые выдвигают рифму, использование местоимений и т.п. Напротив, объединение текстов на основе типа речевого акта предполагает в качестве критерия целостность классифицируемого объекта, поскольку тип характеризует речевое высказывание как целостное образование. Заметим, что это довольно редко используемый критерий для установления сходства текстов. Можно сказать, что основным критерием для объединения текстов является тематическая общность, дополнительным – речевые и коммуникативные особенности слогана. Выявленные критерии не образуют иерархию; в нашем материале нет ни одной классификации, в которой в рамках класса *Реклама товаров для женщин* выделялся бы подкласс *Рифмованная реклама товаров для женщин*.

Насколько мы можем судить, наш эксперимент спровоцировал ситуативные классификации. Причем ситуативными оказываются даже классификации, основанные на признаках текста, таких, как ключевые слова. На первый взгляд, это противоречит разграничению иерархических и ситуативных классификаций. Однако мы не склонны усматривать здесь противоречие. Ключевое слово, выделяемое из текста и номинирующее его тему, представляет один из способов обозначения цельности текста (Сахарный, Штерн 2006). Ключевое слово таким образом «заменяет» целый текст, является результатом его смысловой компрессии. Причем в результате компрессии и обозначения цельности ключевым словом или набором ключевых слов реципиент использует не только встретившиеся в тексте слова, но и их синонимы, что обусловлено соотносительностью синонимов с одним концептом, принадлежностью десигнатов синонимичных слов одной категории (Сахарный 1994). В результате использование совпадения ключевых слов и стоящих за ними ключевых для текста концептов в качестве критерия объединения текстов в класс приводит к тому, что в классификации текст отражен целостно, а не определенным своим признаком. Ситуативность классификации отражается и в названиях выделенных классов, как правило, содержащих ключевые слова.

Ситуативность классификаций, на наш взгляд, обусловлена принципиальной ситуативностью коммуникативного акта. Речевая деятельность – деятельность по планированию и порождению речевого высказывания, его восприятию и пониманию – вписывается в социальное взаимодействие, контекстом которого выступает коммуникативная ситуация. При планировании речевого сообщения в равной степени значимы две ситуации: ситуация общения и ситуация-тема высказывания (Леонтьев 1997: 114). Ситуация общения оказалась центральным объектом лингвистической прагматики и основанием для построения типологии речевых актов. Ситуацию-тему в лингвистике называют референтной ситуацией, относят ее к области референции высказывания.

В нашем эксперименте «свободная классификация» испытуемые воспринимали и интерпретировали тексты в рамках экспериментальной ситуации; естественная коммуникативная ситуация не моделируется в такого рода экспериментах. В таком случае приоритетной для восприятия и переработки речевого сообщения оказалась ситуация-тема, отраженная ключевыми словами текста. С ситуацией общения, как нам кажется, соотносятся такие признаки воспринимаемого сообщения, как тип речевого акта, отражающий интенцию говорящего (пишущего). В экспериментальных условиях, когда нет необходимости реагировать на сообщение, когда коммуникативное взаимодействие сведено к

взаимодействию «экспериментатор – участник эксперимента» (см. подробнее Доценко 2006), параметры коммуникативной ситуации менее значимы, чем ситуация-тема.

В свою очередь для ситуации-темы малозначимы формальные параметры текста, к которым мы склонны относить и речевые ошибки. Ошибки в тексте в целом мало влияют на его смысловую обработку. Ошибки попадают в поле зрения реципиента в том случае, когда затрудняют доступ к ситуации-теме; в этом видится причина появления в классификациях пятой части испытуемых класса текстов с грубыми ошибками, озаглавленного *Реклама ни о чем, Нелепость* и т.п.

Эксперимент «свободная классификация» и кластер-анализ как метод обработки полученных данных являются, на наш взгляд, надежным и достоверным методом выявления вариативности восприятия и оценки текста. Рассмотренные методы позволяют установить значимые для носителей языка основания группировки языкового материала, отражающие его ценность в речевом и когнитивном опыте индивида. Знаменитая пародия Борхеса на классификацию животных из якобы китайского источника обнажает природу наивных классификаций, необходимых для конкретизации интуитивного познания.

Библиографический список

- Алгоритмы обработки экспериментальных данных / М., 1986.
- Брунер Дж. Психология познания / Дж. Брунер. М., 1977.
- Вертгеймер М. Продуктивное мышление / М. Вертгеймер. М., 1987.
- Доценко Т.И. Ассоциативный эксперимент и речевое поведение информанта в экспериментальной ситуации / Т.И. Доценко // ...Слово отзовется: памяти Аллы Соломоновны Штерн и Леонида Владимировича Сахарного. Пермь, 2006. С. 149-158.
- Леонтьев А.А. Основы психолингвистики / А.А. Леонтьев. М., 1997.
- Малиновский Л.Г. Процессы классификации – основа построения наук о действительности / Л.Г. Малиновский // Алгоритмы обработки экспериментальных данных. М., 1986. С. 156-168.
- Налимов В.В. и др. Реальность нереального. Вероятностная модель бессознательного / В.В. Налимов, Ж.А. Дрогалина. М., 1995.
- Овчинникова И.Г. Ассоциативный механизм в речемыслительной деятельности: Автореф. дис. ... докт. филол. наук / И.Г. Овчинникова. Спб., 2002.
- Петренко В.Ф. Психосемантика сознания / В.Ф. Петренко. М., 1988.
- Сахарный Л.В. Человек и текст: две грамматики / Л.В. Сахарный // Человек. Текст. Культура. Екатеринбург, 1994. С. 7-59.
- Сахарный Л.В. и др. Набор ключевых слов как тип текста / Л.В. Сахарный, А.С. Штерн // ...Слово отзовется: памяти Аллы Соломоновны Штерн и Леонида Владимировича Сахарного. Пермь, 2006. С. 50-59.
- Тулдава Ю. Проблемы и методы количественного исследования лексики / Ю. Тулдава. Таллинн, 1987.
- Фрумкина Р.М. Цвет, смысл, сходство / Р.М. Фрумкина. М., 1984.
- Фрумкина Р.М. Психолингвистика / Р.М. Фрумкина. М., 2001.
- Шенерт В. Грядущая реклама / В. Шенерт. М., 2001.
- Штерн А.С. Перцептивный аспект речевой деятельности / А.С. Штерн. СПб., 1992.