

Принципы организации психолингвистического эксперимента и обработки его результатов

К.С. Карданова, А.Г. Сонин

В силу того что в российской психолингвистике до настоящего времени не сложилась единая система норм и правил организации и проведения эксперимента, актуальной представляется проблема систематизации знаний о психолингвистическом эксперименте и разработки рекомендаций по его проведению с описанием всех основных этапов работы: от принятия решения о целесообразности использования эксперимента до интерпретации его результатов. В статье рассмотрены универсальные принципы планирования и проведения психолингвистического эксперимента, которые в наименьшей степени допускают вариативность и не зависят от конкретных задач, решаемых исследователем. Во второй части статьи описаны некоторые возможности анализа полученных результатов с помощью программ статистической обработки данных. Многие ученые настороженно относятся к статистическим пакетам, считая, что их багаж знаний в области математики недостаточен для успешного применения этих программ в работе. Именно поэтому систематизация знаний о процедурах статистического анализа результатов и обоснование необходимости их применения также представляет актуальную задачу.

Ключевые слова

Психолингвистический эксперимент, методология и технология психолингвистического эксперимента, статистическая обработка данных

The Framework of Experiment in Psycholinguistics and Statistical Analysis of the Data

Since psycholinguistics in Russia can't boast a comprehensive system of norms and principles of planning and carrying out an experiment, it is essential to devise a framework of a psycholinguistic experiment which will steer a researcher through every phase of the process starting with the decision to prefer experimental approach over other empirical methods to processing and interpreting the findings. The paper outlines the foundations of a psycholinguistic experiment, which are independent of the objectives of a particular research and which are to be considered whenever relying on this method. The paper further argues that statistical analysis of the data obtained is indispensable. It propounds an algorithm of processing the findings relying on specialized software.

Key words

Psycholinguistic experiment, framework of a psycholinguistic experiment, foundations of a psycholinguistic experiment, statistical analysis of the data

Как известно, психолингвистика начала оформляться в качестве самостоятельного направления науки о языке в 50-е годы XX века в связи с растущим интересом языковедов к изучению человека как носителя языковой способности и как субъекта речевой деятельности. Первые попытки построения моделей, раскрывающих принципы хранения языка в памяти человека и отображающих психические процессы, лежащие в основе порождения и понимания речи, продемонстрировали необходимость обращения к новым методам.

Прежние методы, разработанные представителями структуралистского подхода, были направлены на моделирование виртуальных отношений между единицами языка и не позволяли включать в модель его носителя.

Поиск метода, адекватного новому теоретическому объекту, вывел ученых за пределы лингвистической науки в область психологии, в то время еще находившейся под влиянием бихевиоризма и имевшей богатый опыт экспериментального изучения поведения человека, в частности речевого. Освоение техник проведения психологического эксперимента стало важным шагом в становлении нового направления. Достаточно было изменить характер проверяемых гипотез, перенеся внимание с речевых действий на организацию языковой способности, и эксперимент становился психолингвистическим. Впрочем, сама психология вскоре пошла по этому пути под влиянием идей, пришедших вместе с когнитивной революцией. Исследования языка в когнитивной психологии стали строиться усилиями международного сообщества на строгом отборе методов экспериментального исследования языка / языковой способности / речевой деятельности и на тщательной алгоритмизации всех этапов используемых экспериментов. Отечественная лингвистика не включилась до конца в этот процесс. И сегодня методологические и методические проблемы использования экспериментальных процедур в российской лингвистике стали препятствием для полноценной ее интеграции в мировую языковедческую науку (речь идет не только о собственно языкознании, но и о множестве междисциплинарных областей изучения языка).

В свете изложенного актуальной представляется проблема систематизации знаний о психолингвистическом эксперименте и разработки рекомендаций по его проведению с описанием всех основных этапов работы: от принятия решения о целесообразности использования эксперимента (а не другого эмпирического метода) до интерпретации его результатов. Именно эту задачу ставили перед

собой авторы статьи. Естественно, при организации конкретного исследования ученый руководствуется собственной исследовательской логикой, творческим отношением к решению поставленных задач, научной интуицией. Однако ряд универсальных принципов, суммированных в статье, может представлять собой стержень любого экспериментального исследования вне зависимости от нюансов.

1. Методология и технология психолингвистического эксперимента

1.1. Условия выбора эксперимента в качестве инструмента исследования

Сегодня многие специалисты считают необходимым включать в исследование эксперимент, чтобы маркировать свою принадлежность к сообществу психолингвистов, и тем самым превращают эксперимент в своего рода знак отличия психолингвистических исследований. Это непродуктивная позиция. Эксперимент – лишь инструмент, пусть и с широкими, но все же весьма ограниченными возможностями применения. В связи с этим первое, что требует осмысления при планировании эксперимента, это его адекватность решаемой проблеме.

Эксперимент (в самом общем виде) – это наблюдение в искусственной ситуации. Искусственность эксперимента – существенное препятствие к его использованию для проведения исследования в гуманитарных науках. Если физик ничем не рискует, создавая эту искусственность, потому что поведение молекулы зависит исключительно от внешних условий, то в случае, когда испытуемыми являются существа с эмоциями, волей и разумом, дело обстоит иначе: в искусственной ситуации мотив деятельности изменяется, потенциально влияя на действия испытуемого. По этой причине любой вывод, сделанный на основании эксперимента, всегда можно поставить под сомнение, так как зафиксированные действия совершались в неестественной ситуации, и оснований полагать, что

аналогичные действия будут производиться в естественной ситуации, нет.

Помещение испытуемых в искусственные условия, нацеленное на снижение риска контаминации результатов неизвестными и неконтролируемыми факторами, приводит к получению искусственных же результатов, что превращает эксперимент в крайнюю меру, к которой следует прибегать лишь тогда, когда другого выхода нет. В противном случае уместно искать другие эмпирические методы. Если гипотеза касается наблюдаемых явлений, то бóльшую экологическую валидность продемонстрирует хорошо организованное наблюдение, а если гипотеза относится к области ненаблюдаемого, то, возможно, следует применить интроспекцию, поскольку эксперимент не способен что-либо рассказать, например, о переживаемом индивидом в процессе порождения или понимания речи, он не дает знание о живом смысле. В число гипотез, требующих экспериментальной верификации, прежде всего входят гипотезы о специфической операции, оторванной от общего контекста деятельности, а именно об отношении между неким фактором и отдельной операцией: например, формат предъявления материала – скорость ответа, скорость показа – количество усвоенных слов и др.

1.2. Соотнесение эксперимента с теоретической платформой исследования

Выше отмечалось, что эксперимент – это инструмент для решения соразмерной ему проблемы. Однако использование эксперимента не только «объективно» зависит от решаемой проблемы, но и субъективно определяется теоретической моделью, с позиции которой предлагается решение. С этой точки зрения эксперимент – это способ проверить вытекающие из теоретической модели предположения о решении проблемы. Поэтому «объективность» проблемы относительна, так как ее постановка в науке обязательно связана с противоречиями, возникающими либо (1) внутри самой теории, либо (2) при

сопоставлении базовых теоретических положений с эмпирическим материалом, либо (3) при столкновении двух конкурирующих способов объяснения некоего явления. Без соотнесения исследования с теоретической платформой, на которую опирается представляемая ученым школа, приступать к эксперименту бессмысленно.

Чтобы обосновать эту мысль, рассмотрим пример, связанный с изучением специфических поликодовых текстов (комиксов) одним из авторов статьи (Сонин 1999). Довольно долго изучение материала носило хаотический характер, пока не произошло знакомство с представителями Барнаульской психолингвистической школы, занимавшимися проблемой влияния смысловой доминанты на речемыслительные процессы. Вхождение в школу и принятие теории позволили выработать оригинальный взгляд на исследуемый объект: был проведен анализ взаимодействия гетерогенных составляющих комиксового текста с позиции выражаемой автором комикса смысловой доминанты. Принятая теория позволила сформулировать гипотезы о функционировании объекта исследования и провести эксперимент для проверки этих гипотез (Сонин 1999). Обратной стороной процесса стало обогащение и укрепление теории смысловой доминанты новыми эмпирическими данными из особой предметной области. Таким образом, залогом плодотворного экспериментального исследования выступает его опора на базовые теоретические постулаты той школы, к которой примыкает исследователь. В противном случае автор оказывается за границами научного познания. Если в теории той школы, к которой примыкает автор, нет постулатов, выступающих основой для формулирования гипотез о некоем объекте, автору следует либо поменять объект, либо опереться на другую теорию.

1.3. Работа над экспериментальными гипотезами

В современной науке сложился ряд принципов выдвижения гипотез. Эти принципы призваны гарантировать надежность

полученного результата и возможность его дальнейшего использования в рациональных построениях. На этапе выдвижения экспериментальных гипотез необходимо придерживаться следующих постулатов.

1. Экспериментальная гипотеза должна описывать отношение. Например, отношение типа «если А, то Б», или «если меняется А, то меняется Б», или «как меняется А, так меняется Б».
2. Экспериментальная гипотеза должна быть проверяема, то есть соотносима с определенными экспериментальными действиями, в которых фиксируется реакция испытуемого на некое событие.
3. Экспериментальная гипотеза должна предполагать возможность ее опровержения, поэтому формулировки типа «Возможны случаи, при которых...» недопустимы.
4. Экспериментальная гипотеза логически выводится из результатов предшествующих исследований ученого или его школы.
5. Поскольку в эксперименте создаются группы, отличающиеся неким фактором, экспериментальная гипотеза – это позитивная гипотеза о различиях. Выдвижение гипотезы об отсутствии различий создает методологические трудности (наука нацелена на выявление нового), а также проблемы статистической интерпретации результатов.

Кроме общих принципов выдвижения гипотез, существует порядок их дедуктивного встраивания в эксперимент. На первом этапе формулируется **общая рабочая гипотеза**, которая касается отношения между двумя классами фактов. Будучи предположением общего характера, такая гипотеза не может проверяться в эксперименте непосредственно, однако предполагает различные способы опосредованной проверки. Когда эти способы найдены, для каждого из них может быть выдвинута специфическая **операциональная** гипотеза. В ней рабочая гипотеза детализируется с учетом выбранного экспериментального построения (в частности, с учетом эмпирического материала). Операциональной такая гипотеза называется потому, что ее содержание зависит от планируемых

операций, направленных на проверку рабочей гипотезы. Наконец, в ходе проведения эксперимента могут возникать **технические гипотезы** о соотносимости групп и сопоставимости используемого материала, а также **статистические гипотезы**, необходимые для обработки полученных данных.

В качестве примера может быть рассмотрен эксперимент, проведенный в рамках программы исследования механизма воздействия на реципиента поликодовых текстов (в этом случае материалом послужили поликодовые рекламные тексты) (Сонин 2005). Согласно построенной автором теоретической модели,

базовая репрезентация содержания поликодового вербально-изобразительного текста (его первичное понимание) строится преимущественно на основе обработки изобразительной составляющей, а влияние вербальной составляющей рассматривается как модулирующее.

Эта отвлеченная идея, выдвигаемая на основе современных теоретических представлений о когнитивных процессах, лежащих в основе понимания текста, не может быть проверена в принципе. Первым шагом к подготовке эксперимента может стать выбор экспериментального материала. Допустим, что мы остановили свой выбор на такой разновидности поликодовых текстов, как тексты газетной рекламы. Поскольку они представляют собой разновидность поликодовых текстов, к ним должны быть приложимы утверждения, относящиеся ко всем поликодовым текстам. Отсюда

базовая репрезентация содержания поликодового вербально-изобразительного текста газетной рекламы (его первичное понимание) строится преимущественно на основе обработки изобразительной составляющей, а влияние вербальной составляющей рассматривается как модулирующее.

Однако и в этом случае гипотеза не может быть проверена, так как касается процессов, не поддающихся непосредственному наблюдению. Для формулировки операциональной гипотезы

необходимо определиться с операциями, по которым мы сможем судить о предполагаемой иерархии составляющих и связанных с их пониманием процессов. Чтобы найти такие операции, переформулируем гипотезу:

поскольку основой понимания поликодового вербально-изобразительного текста газетной рекламы выступает его изобразительная составляющая, при проверке понимания его отдельных взятых частей (изобразительной и вербальной) реакция на изобразительную составляющую будет ближе к реакции на цельный текст, чем реакция на вербальную составляющую.

Эта формулировка приближает нас к выдвижению операциональной гипотезы, так как переводит предположения из области ненаблюдаемых когнитивных процессов в область поведенческих реакций. Для дальнейшего продвижения к цели необходимо определиться со способом оценки реакций испытуемых. С учетом технических факторов, возможностей дальнейшей статистической интерпретации, а также временных и материальных ресурсов для оценки был выбран метод шкалирования характеристик предъявляемых материалов (поликодового текста и его отдельных составляющих). После этого была сформулирована операциональная гипотеза:

оценки изобразительной составляющей по комплексу предложенных испытуемым шкал будут в большей степени коррелироваться оценками интегрального текста, чем оценки вербальной составляющей.

Эту гипотезу, выведенную дедуктивно из теоретической модели, можно проверить эмпирически.

1.4. Подготовка к проверке гипотез

Переход к завершающей части подготовки эксперимента предполагает уточнение обстоятельств, в которых будет проверена гипотеза. Для этого проводится анализ выдвинутой операциональной гипотезы с точки зрения формализации ключевых **переменных** –

параметров, вариации которых могут быть заданы или измерены и подвергнуты категоризации. Одни переменные связаны с особенностями условий, в которых будет проходить испытание, с характеристиками испытуемых, с нюансами предлагаемых им заданий и другими моментами, которые могут повлиять на результат. В этом случае речь идет о заранее спланированных параметрах эксперимента, которые исследователь может варьировать или просто иметь в виду. Такие параметры называются факторами или независимыми переменными. Независимые переменные могут вводиться исследователем (например, тип текста, скорость предъявления материала) или могут только контролироваться (например, возраст, пол и т.д.). Как видно из приведенных примеров, факторы не зависят от действий испытуемых, но подчиняются выбору экспериментатора.

Другая группа фактов, раскрываемых в эксперименте, связана с реакциями испытуемых на проводимую манипуляцию. Это зависимые переменные, или данные. Последние могут использоваться в последующем анализе непосредственно, однако нередко первичные данные подвергаются переработке: кодификации, классификации и т.д. В выборе средств стандартизации ограничивать экспериментатора может только его фантазия.

Продолжим использовать в качестве примера эксперимент, проведенный в рамках исследования специфики поликодовых рекламных текстов (Сонин 2005). Для сбора оценок предъявляемых испытуемым текстов необходимо было определить категории, по которым будут производиться оценки. Для установления релевантных категорий было проведено интервьюирование 26 информантов (13 мужчин и 13 женщин в возрасте от 18 до 30 лет), не принимавших участия в основном эксперименте. Интервьюирование проводилось на основе 18 рекламных текстов, взятых из современных печатных СМИ. Рекламные тексты предъявлялись каждому испытуемому индивидуально. В задании предлагалось охарактеризовать каждый их

предъявленных текстов. Выбор языковых средств, используемых для характеристики, никак не ограничивался. Все использованные характеристики протоколировались, после чего на основе анализа ответов устанавливалась частотность использования в описании рекламы разных вербальных средств. Из 232 языковых единиц, зафиксированных в ответах опрошенных, для названия шкал (категорий оценки), при помощи которых изучалось восприятие рекламных текстов на следующем этапе эксперимента, были отобраны наиболее частотные (использованные в ответах 15% информантов и более). Отобранные категории представляли самые разные сферы субъективной оценки, позволяя охарактеризовать рекламные тексты одновременно с нескольких сторон:

- с точки зрения производимого на реципиента эмоционального воздействия («навязчивый», «привлекательный», «приятный»),
- в связи с анализом его рациональных построений («логичный», «понятный»),
- на основе выявления его эстетической ценности («оригинальный», «стильный», «удачный»),
- в связи с особенностями восприятия его формы («яркий»),
- по его воздействию на чувственную сферу личности реципиента («сексуальный»).

Отдельно решался вопрос о формализации ответов испытуемых. С учетом ряда технических моментов для этой цели была выбрана семибалльная шкала:

- 7 – это качество выражено максимально
- 6 – это качество выражено в значительной степени
- 5 – это качество выражено в некоторой степени
- 4 – это качество не выражено
- 3 – скорее, в некоторой степени выражено противоположное качество
- 2 – противоположное качество выражено в значительной степени
- 1 – противоположное качество выражено в максимальной степени

Таким образом, эксперимент, предназначенный для проверки гипотезы о различиях в оценке семиотически гетерогенных составляющих поликодового рекламного текста, был построен на манипуляции тремя факторами. Демонстрация рекламного текста и отдельная демонстрация каждой из его составляющих были направлены на операционализацию фактора «модальности демонстрируемого текста» (М). Контроль проводился на основе сравнения результатов трех групп (участники, оценивавшие вербальную составляющую, vs. участники, оценивавшие изобразительную составляющую, vs. участники, оценивавшие целый рекламный текст). Таким образом, фактор включал три модальности. Второй фактор – «шкала» (Ш) был представлен 25 модальностями (по числу категорий оценки). Третий фактор (он лишь контролировался, но не использовался для интерпретации результатов) связан с различиями между пятью использованными в экспериментах текстами. Таким образом, план эксперимента может быть сформулирован так: $S_n < M_3 > * Ш_{25} * T_5$.

С учетом последних соображений можно переформулировать суть эксперимента следующим образом: эксперимент – это средство установления отношений между независимыми и зависимыми переменными. Когда все необходимые средства для проверки этих отношений собраны, можно приступать к сбору данных.

2. Статистическая обработка экспериментальных данных

Грамотное планирование и успешное проведение эксперимента – важные шаги на пути к получению достоверного результата. Однако усилия экспериментатора могут оказаться пустыми из-за неправильной обработки полученных данных. Согласно принятым нормам, обработка включает два этапа: статистический анализ результатов эксперимента и их интерпретацию.

Как показывает анализ диссертационных исследований, проведенных за несколько последних лет, в ряде случаев

статистический анализ подменяется простейшими арифметическими действиями по подсчету количества полученных ответов и вычислениями процентного соотношения разных вариантов ответа.

Настороженное отношение психолингвистов к сложному математическому аппарату, необходимому для статистической обработки результатов эксперимента, вполне объяснимо. Они с полным на то основанием полагают, что не обладают достаточным багажом математических знаний для проведения сложных расчетов. Когда гуманитариям предлагают высчитать «вручную» коэффициент надежности или стандартную ошибку измерения, не стоит ждать от него большого энтузиазма. Однако эта проблема уже давно осталась в прошлом. Сегодня в арсенале современного исследователя присутствует по крайней мере два десятка программ статистической обработки данных. Среди них особой популярностью пользуются пакеты SPSS, SAS, SYSTAT, Minitab, Statistica, STADIA и др. Такие программы позволяют быстро и эффективно анализировать полученные экспериментальные данные и осуществлять контроль качества эксперимента (с их помощью исследователь может, например, оценить валидность полученных результатов, репрезентативность выборки испытуемых и пр.).

2.1. Почему психолингвист не может обойтись без статистической обработки данных?

Применение методов статистической обработки данных позволяет решать разные задачи. Некоторые из них факультативны, они открывают для экспериментатора более широкие возможности при интерпретации полученных результатов. Другие выступают обязательным компонентом любого экспериментального исследования. К ним в первую очередь относятся методы оценки надежности измерений, методы оценки выборки испытуемых и методы преодоления экспертных ошибок. Остановимся более подробно на этих трех группах проблем и на связанных с их решением методах.

А) важность установления корреляций между тестовым баллом и оценками измеряемой латентной переменной (истинным баллом и истинным значением характеристики испытуемого)

При оценке действий испытуемого (выставлении балла за выполнение задания) экспериментатор имеет дело с так называемым **тестовым баллом** (*test score, observed score*), т.е. наблюдаемым баллом респондента по данному заданию. Однако если исследователь принимает этот балл как действительное отображение изучаемого признака, он допускает грубую ошибку. Дело в том, что тестовый балл в значительной степени зависит от выбора экспериментальных заданий, от условий проведения эксперимента и пр. факторов, а следовательно, в нем *a priori* заложено искажение. Для того чтобы его ликвидировать специалист проводит корректировку тестового балла с учетом ошибки измерения, получая в результате этой процедуры доверительный интервал для **истинного балла** (*true score, unobserved score*).

Второй оценкой латентной переменной является **истинное значение характеристики испытуемого**, измерение степени проявления которой составляет цель эксперимента (например, толерантность (Голикова 2005) или вежливость (Степыкин 2011) и др.). Степень проявления характеристики может меняться с течением времени, но на момент оценивания она является константой для каждого отдельно взятого респондента. Методы оценки истинных значений характеристики испытуемого разрабатываются в рамках современной теории тестирования IRT (*Item Response Theory*).

Рассмотрим, как соотносятся между собой эти величины. Экспериментальный тест всегда представляет собой набор заданий (выборку) из генеральной совокупности заданий, измеряющих степень проявления некоторой характеристики. Поэтому тестовые баллы испытуемых и соответствующие им истинные баллы всегда зависят от того, какие задания были включены в эксперимент. Если отобраны трудные задания, наблюдаемый и истинный баллы оказываются более

низкими, а если отобраны легкие задания, баллы являются более высокими. Однако при этом степень проявления характеристики остается постоянной величиной для каждого испытуемого, т.е. не зависит от того конкретного экспериментального теста, который используется для измерения данного конструкта. Поэтому задача, стоящая перед экспериментатором, заключается в том, чтобы соотнести наблюдаемые переменные (тестовые баллы) с ненаблюдаемыми переменными (истинные баллы участников эксперимента, степень проявления характеристики у испытуемого).

Б) ограничения, накладываемые выборкой испытуемых

Как известно, возможности привлечения испытуемых всегда ограничены. Крайне редко встречаются эксперименты с выборкой в десятки тысяч человек. В большинстве случаев выводы, которые делает исследователь, основываются на ответах 50-500 участников. Эти выводы далее экстраполируются на всю генеральную совокупность, т.е. на каждого представителя исследуемой группы.

Так, если мы исследуем структурно-содержательные особенности фрагмента языкового сознания русских и англичан, на этапе подготовки эксперимента мы подбираем репрезентативную экспериментальную группу, стараясь максимально сбалансировать ее в соответствии со всеми релевантными для эксперимента параметрами (возрастными, гендерными, социальными, образовательными и пр.). Например:

Национальность	Русские				Англичане			
	<25		>25		<25		>25	
Возраст								
Пол	М	Ж	М	Ж	М	Ж	М	Ж
Количество испытуемых	70	76	67	80	65	75	80	73

Таблица 1. Анализ релевантных характеристик экспериментальной выборки

Однако, как бы мы ни старались сделать выборку репрезентативной, мы не можем учесть все параметры, а следовательно, задействованная в эксперименте группа испытуемых будет лишь с определенной степенью приближения отражать

характеристики генеральной совокупности. Данное искажение, несомненно, следует учитывать при формулировании выводов, однако высчитать погрешность можно только опираясь на методы математической статистики.

В особенно сложной ситуации оказывается экспериментатор, работающий с неравными выборками испытуемых. Так, в диссертационных исследованиях часто встречается формулировка: «Меньшее количество американских испытуемых связано с техническими сложностями организации эксперимента. Тем не менее, то, что в эксперименте приняло участие достаточное количество человек из разных возрастных групп делает полученные результаты валидными». В таких случаях процентное соотношение, которое многие исследователи применяют при анализе полученных данных, не отражает действительное распределение признака.

В качестве примера приведем данные, полученные в ходе ассоциативного эксперимента с носителями русского и английского языков (98 носителей РЯ и 26 носителей АЯ). Цель исследования заключалась в моделировании структуры и содержания ядра языкового сознания. Среди реакций, полученных при предъявлении испытуемым ряда слов-стимулов, были реакции ДОМ, Я, РАБОТА. Исследователь высчитал в процентах соотношение между количеством данных реакции и общим числом реакции, полученных на данное слово-стимул, и представил данные в виде следующей таблицы:

Реакция на слово-стимул	Количество от общего числа реакций	
	Носители РЯ (98 человек)	Носители АЯ (26 человек)
ДОМ	15%	13%
Я	7%	15%
РАБОТА	2%	10%

Таблица 2. Количественный анализ реакций ДОМ, Я и РАБОТА

В результате несложного сравнительного анализа был сделан предварительный вывод о том, что ядро языкового сознания носителей русского и английского языков по-разному структурируется, поскольку

процентные соотношения реакций на предъявленные в эксперименте слова-стимулы не равны.

Однако в результате последующей статистической обработки данных было выявлено, что экспериментальные данные не обнаруживают значимых расхождений в отношении реакции ДОМ, маргинальные различия наблюдаются в отношении реакции Я, и только различие в количестве реакций РАБОТА является значимым.

В) искажения, вызванные экспертными ошибками (rater effects)

Выводы, которые основываются не на собственных наблюдениях или на интроспекции исследователя, а на данных, полученных экспериментально, часто представляются как более надежные и валидные. Однако испытуемые (респонденты, эксперты) – это тоже люди, их суждения и решения могут быть ошибочными. На оценку могут оказать влияние внешние, незначительные факторы, вызывая искажения, которые существенно ухудшают качество исследования. Некоторые искажения наблюдаются в оценках отдельных испытуемых (т.е. единичны), однако есть и такие искажения, которые наблюдаются регулярно (например, склонность не использовать крайние значения на шкале оценки, тенденция позволять общему впечатлению о человеке повлиять на оценку конкретного вида деятельности).

Более 75 лет специалисты в области психометрии изучают экспертные ошибки, систематизируют их и разрабатывают статистические модели, которые позволяют высчитать вызванные экспертными ошибками искажения. Первоначально проблема качества оценивания рассматривалась в связи с введением тестов в системе образования, что обусловило появление термина «экспертная ошибка», обозначающего искажение, которое происходит при оценивании экзаменуемого экзаменатором (экспертом). Аналогичным образом психолингвист, обращаясь к носителям языка, видит их как своего рода экспертов, которых он просит оценить некоторое явление или процесс (пусть и не другого человека). Поэтому экспертные

ошибки регулярно обнаруживаются и при анализе результатов психоинвестиционного эксперимента. Рассмотрим наиболее известные из них (см. Myford, Worfe 2003; 2004):

1) Эффект строгости / снисходительности (*leniency / severity effect (generosity error)*) проявляется в склонности испытуемого занижать / завышать баллы в каждой ситуации оценки (т.е. его баллы будут, в среднем, более низкими / высокими, чем ожидалось для этой экспериментальной группы). Так, в приведенном выше эксперименте с поликодовыми (рекламными) текстами эффект строгости проявляется тогда, когда испытуемый занижает баллы при выполнении экспериментального задания, направленного на оценку воздействия полного вербально-изобразительного рекламного текста и препарированного рекламного текста (содержащего только вербальную или только изобразительную составляющую).

2) Эффект гало (*halo effect*) заключается в склонности испытуемого использовать близкие оценки вне зависимости от того, какой параметр ему предлагается оценить. Иными словами, такой испытуемый плохо дифференцирует задания и шкалы. Так, в описанном ранее эксперименте (см. параграф 1.4) было задействовано 25 шкал: «логичный», «навязчивый», «оригинальный», «понятный», «привлекательный», «приятный», «сексуальный», «стильный», «таинственный», «удачный», «яркий», «возвышенный», «выразительный», «глубокий», «добрый», «доступный», «интересный», «красивый», «поэтический», «простой», «радостный», «светлый», «серьезный», «тяжелый», «эмоциональный» (Сонин 2005). Если испытуемый, которому было предложено разместить свои оценки ряда параметров экспериментального текста на этих шкалах, демонстрирует эффект гало, он размещает на каждой шкале близкие оценки, не учитывая того, что шкалы направлены на измерение различных параметров текста.

3) Эффект центральной тенденции (*central tendency effect*) проявляется в «стремлении избегать крайних (положительных или

отрицательных) оценок или в предпочтении баллов, близких к середине шкалы» (Landy, Farr 1983: 71). Одна из причин появления этого эффекта связана с тем, что испытуемый опасается ставить заниженные или завышенные баллы, и поэтому старается выбирать значения, близкие к средним (Guilford 1954: 278). Это ведет к снижению надежности теста и валидности полученных результатов, так как значительно сокращается спектр используемых оценок. Чтобы определить, присутствует ли эффект срединной тенденции в оценках группы испытуемых, достаточно посмотреть на то, насколько близко расположен средний балл к центру шкалы (Landy, Farr 1980; 1983).

4) Эффект ограничения спектра (*restriction-of-range effect*) заключается в том, что все баллы сосредоточены на одном участке шкалы. Это может быть либо завышенная (снисходительная) оценка, либо заниженная (строгая) оценка, либо оценка, близкая к центру шкалы (т.е. эффект центральной тенденции) (Saal et al. 1980: 418). Это искажение проявляется существенным отклонением кривой оценок от кривой нормального распределения.

5) Эффект случайности (*inaccuracy effect*) демонстрирует склонность одного из испытуемых применять шкалы оценки (одну или несколько) иначе, чем другие участники эксперимента. Принимая решение, этот респондент может учитывать нерелевантные для эксперимента факторы или руководствоваться личными мотивами. При этом его оценки носят случайный характер. Выявить такого испытуемого легко, поскольку его баллы существенно отклоняются от средних значений по всей экспериментальной выборке. Анкету, в которой присутствует эффект произвольности, следует отбраковать на этапе первичной обработки, поскольку она *a priori* содержит искажение и может привести к неточностям в ходе дальнейшей обработки данных.

Помимо этих экспертных ошибок в психометрии также изучаются эффекты логической ошибки, контрастной ошибки, влияния предубеждений, убеждений, отношения и личных характеристик

испытуемого, влияния предшествующего опыта испытуемого (*logical error, contrast error, influences of rater biases, beliefs, attitudes and personality characteristics, influences of rater / rate background characteristics*) и пр. (Myford, Wolfe 2003; 2004).

Следует отметить, что устранение экспертных ошибок – это сложная задача не только для исследователя-гуманитария, но и для специалиста-статистика, поскольку расчеты производятся с опорой на современную теорию тестирования IRT (*Item Response Theory*), например, на многофасетное Раш измерение (см. Myford, Wolfe 2003; 2004). Работа с математическими моделями современной теории тестирования требует специальных знаний и программного обеспечения, поэтому чтобы измерить степень влияние экспертной ошибки, психолингвист должен привлечь статистика, работающего с моделями IRT, которых в отечественной науке, к сожалению, очень немного (Карданова 2008).

Вместо этого, экспертные ошибки можно предупредить на этапе подготовки эксперимента. Для этого достаточно разработать шкалу с таким количеством делений, которые будут дифференцироваться испытуемыми (по мнению специалистов, оптимальный вариант – 5 или 7 делений); включить в эксперимент шкалы, оценивающие степень проявления разных характеристик; провести тщательную первичную обработку анкет и устранить те из них, в которых были обнаружены экспертные ошибки.

Таким образом, в силу того, что эксперимент представляет собой наблюдение одного человека (экспериментатора) за другим человеком (испытуемым) в искусственных условиях, в нем неизбежно накапливаются искажения, которые приводят к неточностям в интерпретации полученных результатов. Современная наука предлагает арсенал инструментов диагностики этих искажений и корректировки экспериментальных данных, владение которыми позволяет ученому избежать ошибочных интерпретаций и неверных выводов. В связи с этим статистический анализ ответов испытуемых

представляет собой необходимый этап любого экспериментального исследования.

Однако было бы неверно считать, что к статистической обработке данных исследователь прибегает только тогда, когда было аккумулировано достаточное количество задействованных в эксперименте анкет. Выбор конкретного метода математической обработки должен делаться еще на этапе планирования эксперимента, поскольку тип эксперимента и формулировка экспериментальных заданий в первую очередь определяют выбранным статистическим методом. Например, хорошо известный всем дисперсионный анализ оказывается неприменим к номинативным шкалам (поэтому результаты ассоциативного эксперимента не могут быть обработаны с опорой на этот метод). Рассмотрим, каким образом ученый делает выбор в пользу того или иного метода статистической обработки.

2.2. Какой метод статистической обработки данных применим к результатам эксперимента?

Прежде всего, исследователь решает, какая шкала измерений применима в задуманном им эксперименте. Пожалуй, наиболее известной классификацией шкал измерений является классификация, разработанная С. Стивенсом (Стивенс 1960: 60). Она включает четыре типа шкал:

А) Номинативная шкала предполагает классификацию объектов / субъектов путем распределения их по ячейкам классификации. Эта шкала позволяет различить классифицируемые объекты / классифицируемых субъектов, однако она не позволяет измерить это различие количественно. Так, если мы включим в эксперимент задания, в которых испытуемым предлагается отнести исследуемую единицу к одной из предложенных исследователем категорий, мы, вероятно, планируем использовать номинативную шкалу.

В самом простом случае номинативная шкала содержит всего две ячейки (является дихотомической), при этом признак, который

измеряется с помощью этой шкалы, называется альтернативным. Ср.: «русский – англичанин», «единственный ребенок в семье – имеет братьев и сестер», «проголосовал за – проголосовал против» и пр.

В более сложном варианте номинативная шкала содержит больше двух ячеек: «в возрастной группе до 25 / от 25 до 40 / «старше 40»; «имеющий лингвистическое / историческое / искусствоведческое образование» и пр.

Таким образом, номинативная шкала позволяет высчитать частоты встречаемости разных значений признака без установления между ними количественных соответствий.

Б) Порядковая шкала представляет собой классификацию по принципу «больше-меньше», однако точное количественное соотношение (интервал) между ячейками остается неизвестным. Иными словами, в данном случае экспериментатор ранжирует ячейки от ячейки с наименьшим значением до ячейки с наибольшим значением. Так, если исследователь предлагает испытуемым расположить три текста в последовательность в зависимости от степени проявления признака связности, он предполагает использовать порядковую шкалу. При обработке полученных результатов ученый не сможет определить, насколько один текст более/менее связный, чем другие, однако он сможет определить, в каком тексте этот признак выражен в наибольшей степени, а в каком – в наименьшей.

Аналогичным образом, экспериментатор может предложить испытуемым расклассифицировать лексемы в несколько групп: «положительно коннотированная – нейтральная – отрицательно коннотированная» или распределить переводы текстов в группы: «адекватный – скорее адекватный, чем неадекватный – скорее неадекватный, чем адекватный – неадекватный».

В) Интервальная шкала представляет собой классификацию объектов / субъектов на классы с учетом количественного соотношения (интервала) между ячейками. Иными словами, на этой

шкале определен шаг, или расстояние между значениями признака. Так, в психолингвистических экспериментах, проводимых в США и Европе, часто измеряется время, за которое испытуемый принимает решение. Например, при предъявлении сочетания букв участнику эксперимента предлагается написать полное слово, при появлении слова на экране испытуемый должен отнести его в ту или иную категорию, при предъявлении слова испытуемого просят предложить ассоциацию и пр. При этом измеряется время, которое требуется респонденту для выполнения задания: участник №1 справился с заданием за 4 секунды, участник №2 – за 2 секунды, а участник №3 – за 5 секунд. В данном случае применяется интервальная шкала, поскольку между временем ответа каждого из участников можно высчитать интервал (шаг). Так, различие между первыми двумя участниками составляет 2 секунды, между 2 и 3 участниками – 3 секунды, а между 1 и 3 участниками – 1 секунда.

Г) Шкала равных отношений представляет собой классификацию объектов / субъектов пропорционально степени выраженности измеряемого свойства. Однако такая шкала практически не встречается в психолингвистике: ср. испытуемый №1 в два раза более толерантный, чем испытуемый №2, рекламный текст №1 в четыре раза более структурированный, чем рекламный текст №2 и пр.

Исследователь, планирующий проведение эксперимента, как правило, заранее решает, какую шкалу измерений целесообразно применить в эксперименте. А это позволяет ему уже на этапе подготовки экспериментального исследования решить, какие методы статистической обработки он может задействовать для обработки данных.

Как известно, выбор шкалы измерений зависит от того, какую задачу планируется решить с помощью эксперимента. Так, исследователь, возможно, намеревается сопоставить несколько выборок испытуемых по какому-либо признаку или оценить степень эффективности воздействия, или установить степень согласованности

изменений двух признаков в одной и той же выборке испытуемых и т.д. В каждом из этих случаев применимы свои методы статистической обработки, описанные в соответствующей литературе (например, в Гласс, Стенли 1976). Рассмотрим несколько примеров.

При планировании лонгитюдного исследования на одной выборке испытуемых, в ходе которого предполагается высчитать частотность жаргонных слов в речи подростков (замеры будут сделаны в 2011, 2012 и 2013 г.), ученый ставит задачу оценки сдвига значений исследуемого признака. Как указывает Е.В. Сидоренко, обработка данных, полученных при проведении подобного эксперимента, может осуществляться следующими методами: T – критерий Вилкоксона, G – критерий знаков или Угловое преобразование Фишера (Сидоренко 2000: 34).

Если исследование направлено на выявление различий уровня толерантности мужчин и женщин, то задача, которую ставит исследователь, заключается в выявлении различий в уровне исследуемого признака. Следовательно, для обработки данных можно применять следующие методы: Q – критерий Розенблума, U – критерий Манна-Уитни, Угловое преобразование Фишера (там же).

Таким образом, на этапе планирования эксперимента возможно предусмотреть пути анализа полученных данных методами математической статистики.

Когда исследователь определил, какой статистический метод применим к анализу аккумулярованных им в ходе проведения эксперимента результатам, он обращается к программам статистической обработки данных, в которых заложены необходимые математические операции. На этой стадии экспериментатору требуется только внести в программу данные и применить соответствующий статистический метод. Рассмотрим, каким образом можно осуществить обработку данных с помощью программы SPSS. В качестве примера были выбраны результаты наиболее регулярно используемого психолингвистами ассоциативного эксперимента (Степыкин 2011).

2.3. Алгоритм применения статистических методов на этапе обработки результатов ассоциативного эксперимента (данные Н.И. Степыкина)

Реакции, полученные при предъявлении стимулов *politeness*, *courtesy* и *civility*, были распределены в группы: понятийные ассоциации, эмотивно-оценочные реакции, представления, формальные ассоциации, носитель качества и операциональные реакции. Общее количество ассоциатов на стимул *politeness* составило 127 реакций, на стимул *courtesy* – 136 реакций, на стимул *civility* – 131 реакция. В приведенном ниже фрагменте таблицы представлены понятийные реакции на предъявленные в эксперименте слова-стимулы и общее количество реакций:

Компонент концепта	Ассоциаты на стимул		
	"politeness"	"courtesy"	"civility"
понятийные ассоциации	manners 7, please 6, courteous 5, courtesy 4, respect 4, rule 3, formal 2, friendly 2, quiet 2, thank you 2, upbringing 2, acknowledge, attention, attentive, behaviour, bland, calm, careful, classy, considerate, consideration, decent, etiquette, excuse me, good morning, good evening, hello, hospitality, manner, not at all, sensibility, servility, smooth, social, social situations, thanks, urbane, well dressed, well mannered – 67 ассоциатов (48 %)	manners 10, respect 7, polite 5, please 4, common 3, generosity 2, well mannered 2, etiquette 2, helpful 2, I'm sorry 2, thank you 2, being civil, civil, decent, excuse me, friendly, generous, grace, hello, honesty, polite greeting, politeness, respectful, showing respect, society, well dressed – 56 ассоциата (40 %)	manners 7, respect 6, polite 4, calm 3, civilian 3, formal 3, good manners 3, politeness 3, social 3, able to intuit, autocratic, brisk, brusque, caring, classy, compromise, consideration, courtesy, decent, decency, disingenuous, distant, etiquette, forced, formalities, friendly, helpful, honour, humble, mistrustful, noble, officialdom, open-minded, reserved, respectful, tolerant – 62 ассоциата (44,5 %)
Общее количество реакций	127 реакций	136 реакций	131 реакция

Таблица 3. Понятийные ассоциации на стимулы *politeness*, *courtesy* и *civility*

Поскольку в данном случае решается задача выявления различий в распределении исследуемого признака, для статистического анализа полученных данных, в частности, применим критерий Пирсона χ^2 . Следует отметить, что этот метод направлен на оценку однородности выборки, т.е. он позволяет проверить, относятся ли выборки к одной и той же генеральной совокупности. Следовательно, нулевая гипотеза

(H_0) заключается в том, что выборки однородны, а альтернативная гипотеза (H_1), напротив, – что выборки не однородны.

Проверим, значимы ли различия между количеством реакций, относящихся к понятийной составляющей концепта, на каждое из предъявленных слов-стимулов: *courtesy*, *politeness*, *civility*. Программы статистической обработки данных оперируют числами, поэтому все реакции надо перевести в числовую форму. Присваиваем всем понятийным реакциям значение «1», а всем прочим реакциям – значение «2». Общее количество понятийных реакций в первом столбце составляет 67 (из 127), во втором столбце – 56 (из 136), в третьем столбце – 62 (из 131). Значит, в первом столбце таблицы будет 67 клеток, содержащих «1», и 60 клеток, содержащих «2»; во втором столбце – 56 клеток с «1» и 80 клеток с «2»; в третьем столбце – 62 клетки с «1» и 69 клеток с «2».

Проводим χ^2 анализ (выбираем в меню Analyze -> Descriptive Statistics -> Crosstabs). В графе "Rows" задаем независимую переменную (в нашем случае – 3 колонки), в графе "Column" задаем зависимую переменную (в нашем случае – ответы испытуемых). Дополнительно ставим галочку в графе "Statistics" – χ^2 .

Программа SPSS автоматически составляет таблицу, в которой критерий χ^2 отражен в первой строке. Все, что нам осталось сделать, это посмотреть уровень значимости (в социальных науках низким уровнем статистической значимости принято считать 5 %-ный уровень): при $\text{Sig.} > 0,05$ нет оснований отвергнуть H_0 , поэтому она принимается (выборки однородны), в противном случае H_0 отвергается (и, возможно, при этом верна H_1 , т.е. выборки неоднородны).

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1,661 ^a	2	,436
Likelihood Ratio	1,666	2	,435
Linear-by-Linear Association	,010	1	,919
N of Valid Cases	401		

Таблица 4. χ^2 – анализ реакций на слова-стимулы *politeness, courtesy* и *civility* (соответственно, 1, 2 и 3 столбцы)

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,346 ^b	1	,246		
Continuity Correction ^a	1,079	1	,299		
Likelihood Ratio	1,348	1	,246		
Fisher's Exact Test				,275	,149
Linear-by-Linear Association	1,341	1	,247		
N of Valid Cases	273				

Таблица 5. χ^2 – анализ реакций на слова-стимулы *politeness* и *courtesy* (соответственно, 1 и 2 столбцы)

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,134 ^b	1	,287		
Continuity Correction ^a	,886	1	,347		
Likelihood Ratio	1,135	1	,287		
Fisher's Exact Test				,323	,173
Linear-by-Linear Association	1,130	1	,288		
N of Valid Cases	264				

Таблица 6. χ^2 – анализ реакций на слова-стимулы *courtesy* и *civility* (соответственно, 2 и 3 столбцы)

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,006 ^b	1	,939		
Continuity Correction ^a	,000	1	1,000		
Likelihood Ratio	,006	1	,939		
Fisher's Exact Test				1,000	,519
Linear-by-Linear Association	,006	1	,939		
N of Valid Cases	265				

Таблица 7. χ^2 – анализ реакций на слова-стимулы *politeness* и *civility* (соответственно, 1 и 3 столбцы)

Сравнение трех колонок показывает, что выборки однородны (Sig. = 0,436). Сравнение колонок попарно также подтверждает H_0 (1 и 2 колонки: Sig. = 0,246; 2 и 3 колонки: Sig. = 0,287; 1 и 3 колонки: Sig. = 0,939). Никаких значимых различий не обнаружено, что препятствует выводу, который мог быть сделан с опорой на количество реакций без статистической оценки.

Литература

1. **Гласс, Стэнли 1976:** Гласс Дж., Стэнли Дж. *Статистические методы в педагогике и психологии*. Москва: "Прогресс".
2. **Голикова 2005:** Голикова Т.А. *Этнопсихолингвистическое исследование языкового сознания (на материале алтайско-русского ассоциативного эксперимента)*. Москва: Институт языкознания РАН.
3. **Дружинин 2003:** Дружинин В.Н. *Экспериментальная психология*. Санкт-Петербург: "Питер".
4. **Карданова 2008:** Карданова Е.Ю. *Моделирование и параметризация тестов: основы теории и приложения*. Москва: Федеральный центр тестирования.
5. **Медик, Токмачев 2007:** Медик В.А., Токмачев М.С. *Математическая статистика в медицине: учебное пособие*. Москва: "Финансы и статистика".
6. **Сидоренко 2000:** Сидоренко Е.В. *Методы математической обработки в психологии*. Санкт-Петербург: "Речь".
7. **Сонин 1999:** Сонин А.Г. *Комикс: психолингвистический анализ: Монография*. / Под ред. В.А. Пищальниковой. Барнаул: изд-во Алтайского ун-та.
8. **Сонин 2005:** Сонин А.Г. *Понимание поликодовых текстов: когнитивный аспект: Монография*. Москва: Институт языкознания РАН.
9. **Степыкин 2011:** Степыкин Н.И. *Способы структурно-содержательного моделирования лингвокультурного концепта: дисс. к.ф.н.* Москва.
10. **Стивенс 1960:** Стивенс С. *Экспериментальная психология*. Т. 1. Москва: изд-во иностранной литературы.
11. **DeCottis 1977:** DeCottis T.A. An Analysis of the External Validity and Applied Relevance of Three Rating Formats // *Organizational Behavior and Human Performance*. 19: 247-266.
12. **Guilford 1954:** Guilford J.P. *Psychometric Methods*. New York: "McGraw-Hill, Inc."
13. **Landy, Farr 1980:** Landy F.L., Farr J.L. Performance rating // *Psychological Bulletin*. 87(1): 72-107.
14. **Landy, Farr 1983:** Landy F.L., Farr J.L. *The measurement of work performance: Methods, theory and applications*. New York: "Academic Press".
15. **Myford, Worfe 2003:** Myford, C.M., Wolfe, E.W. Detecting and measuring rater effects using many-facet Rasch measurement: Part I // *Journal of Applied Measurement*. 4: 386-422.
16. **Myford, Worfe 2004:** Myford, C.M., Wolfe, E.W. Detecting and measuring rater effects using many-facet Rasch measurement: Part II // *Journal of Applied Measurement*. 5: 189-227.
17. **Saal et al. 1980:** Saal, F.E., Downey, R.G., Lahey, M.A. Rating the Ratings: Assessing the Psychometric Quality of Rating Data // *Psychological Bulletin*. 88(2): 413-428.