

Глава 3.

Проблематика и методики экспериментального синтаксиса

3.1. Введение

Необходимым предварительным условием для моделирования языкового феномена является сбор достаточного количества эмпирических данных, которые давали бы полное представление о процессах, происходящих в языке. В частности, для исследования конструкций с низкой частотностью нередко используются методы экспериментальной лингвистики. Ввиду своей специфичности некоторая конструкция может быть плохо представлена в корпусе, поэтому удобным источником данных о ее использовании носителями языка становится контролируемый опрос респондентов.

Цель текущей главы — обосновать использование экспериментальных методов в синтаксисе. Ниже мы определим место экспериментального синтаксиса относительно других лингвистических дисциплин. Будут рассмотрены основные методологические проблемы, которые возникают при использовании экспериментальных методов для синтаксического моделирования. Кроме того, в главе представлен обзор основных методологических проблем, которые связаны с использованием лингвистических экспериментов.

3.2. Место экспериментального синтаксиса в науке о языке

Экспериментальный синтаксис в широком смысле — это область знаний, в которой синтаксические структуры исследуются посредством экспериментальных методов [Eddington 2009]. Для того чтобы обосновать использование экспериментальных методов в синтаксисе, обсудим, что такое эксперимент и какова его роль в получении новых знаний о мире.

Эксперимент — исследование, проводимое в специально созданных, управляемых условиях в целях проверки экспериментальной гипотезы о причинно-следственной связи [Дружинин 1999]. Исследователь меняет внешние условия так, чтобы повлиять на изучаемый объект. При этом внешнее воздействие на объект считается причиной, а изменение состояния (поведения) объекта — следствием. Экспериментальные методики пришли в лингвистику через область психолингвистики, которая, в свою очередь, ориентировалась на применение эксперимента в психологии. Наконец, поведенческий психологический эксперимент является результатом переноса естественнонаучного эксперимента на почву психологии. Как в случае естественнонаучного эксперимента, так и при поведенческом экспериментальном исследовании важно учитывать предположения «идеального исследования», которые состоят в следующем: время непрерывно, законы инвариантны относительно времени, пространство изотропно, а научное знание интерсубъективно [Дружинин 1999]. Тем не менее, достижение идеальных условий в реальности оказывается невозможным, развитие мира необратимо, разные моменты времени не идентичны, нет двух идентичных объектов, которые можно было бы включить в класс эквивалентности. Чтобы учитывать отклонения реального исследования от концепта идеального исследования, используются понятия внешней валидности и внутренней валидности, которые определяются как мера соответствия исследования изучаемой объективной реальности и мера соответствия реального исследования идеальному соответственно.

3.2.1. Экспериментальные методы в синтаксисе

Важным представляется вопрос обоснованности использования экспериментальных методик при проведении синтаксических исследований. Экспериментальные методики могут понадобиться в синтаксисе для проверки гипотез о языковом материале. При этом возможные вопросы, которые ставит перед собой исследователь, могут касаться разных аспектов лингвистического анализа. Так, эксперимент может ответить на вопросы о том, соблюдаются ли определенные ограничения на языковом материале, возможны ли конструкции, которые предсказываются определенной теорией, но не содержатся в корпусах. При этом важно понимать, что результаты лингвистического эксперимента не дают ответ на вопрос о том, как наиболее изящно и стройно описать рассматриваемое явление

в рамках теории. Однако лингвистический эксперимент позволяет получить как можно более полное описание языковой действительности.¹³

В психолингвистике выделяется два типа экспериментальных методик: опосредованные (*offline*), при которых исследователь изучает синтаксическое представление предложения уже после того, как его анализ завершился, и непосредственные (*online*), при которых исследователь изучает действие синтаксических механизмов в режиме реального времени [Федорова 2008]. Хотя часть авторов считают, что экспериментальный синтаксис — это в первую очередь эксперименты на оценку грамматичности или приемлемости, и именно такое определение было впервые дано этому термину в [Coward 1997], методики второго типа также могут быть востребованы при теоретических построениях. Рассмотрим два аргумента в пользу такой позиции.

Методика извлечения суждений, которую У. Коварт назвал основным инструментом экспериментального синтаксиса, заключается в следующем: суждения относительно приемлемости грамматической структуры выносятся репрезентативной группой испытуемых, которые оценивают языковые стимулы с различным лексическим наполнением по некоторой шкале (по [Федорова 2013; Schütze 1996]). Строгая экспериментальная процедура должна применяться даже в ситуации извлечения суждений, потому что необходимо обеспечивать контроль за условиями эксперимента. Если условия не будут строго контролироваться, как в случае неформальных опросов носителей языка, экспериментальные манипуляции, которые проводятся с целью понять, что вызвало изменение в оценке грамматичности, потеряют всякий смысл: исследователь не сможет утверждать, что к изменению оценки привела его манипуляция, а не какой-то случайный фактор. Тем не менее, даже при обеспечении максимально возможного контроля за условиями вынесения оценок не вполне ясно, как устроен сам процесс вынесения оценки: как человек сопоставляет свою реакцию на стимул со шкалой, как он интерпретирует свою реакцию и пр. Помимо этого, возникает вопрос, определяется ли суждение о предложении одной лишь грамматикой. Ведь на оценку приемлемости могут повлиять общие когнитивные способности человека, в том числе, например, объем рабочей памяти. В связи с невозможностью дать однозначный

¹³ Важно понимать, что в науке о языке, как и в других областях исследования, могут быть гипотезы, которые невозможно ни подтвердить, ни опровергнуть. В частности, это касается гипотез о происхождении языка, а также гипотез о грамматических изменениях, для которых необходимо длительное накопление достаточного языкового материала [Eddington 2009].

ответ на два поставленных вопроса, наряду с традиционным подходом, в котором оценки респондентов связываются с ограничениями грамматики, выделяется так называемый редукционистский подход (*reductionist approach*) [Phillips 2013]. Согласно этому подходу, источником неприемлемости синтаксической структуры могут служить как ограничения, накладываемые грамматикой, так и ограниченные ресурсы когнитивных механизмов, участвующих в обработке речи. Соответственно, методологический принцип редукционизма состоит в том, что сложные явления могут быть полностью объяснены с помощью законов, свойственных явлениям более простым. Поскольку редукционизм предполагает, что взаимодействие грамматики и некоторого когнитивного механизма обработки речи (как при порождении, так и при понимании) идет постоянно, то и результат этого будет замечен не только при вынесении оценки предложению, но и при других типах экспериментов. При этом установить влияние других когнитивных механизмов, помимо грамматики, на конечный экспериментальный результат возможно только при применении непосредственных (*online*) экспериментальных методик, в которых учитывается время реакции респондента на лингвистический стимул.

Второй аргумент в пользу использования временной характеристики заключается в том, что фактор времени может быть использован в качестве дополнительного контроля за поведением испытуемого. Если респондент предоставлен самому себе, то ничто не мешает прочитать одно стимульное предложение несколько раз и поменять свое первоначальное суждение. Та же проблема существует и при порождении: респондент может записать свой ответ, а затем, перечитав его несколько раз, изменить его (например, руководствуясь тем, что ему кажется более приемлемым с точки зрения правил школьной программы по языку).

3.2.2. Экспериментальный синтаксис и психолингвистика

При обсуждении проблематики экспериментального синтаксиса важно установить, в чем заключается отличие данного направления от психолингвистики. Несомненно, экспериментальная лингвистика в части методов пересекается с психолингвистикой. Основное отличие между двумя областями заключается в целеполагании. Цель экспериментального синтаксиса заключается в построении таких теорий о языковых структурах, которые основаны на количественных данных. Причем гипотеза проверяется с использованием статистических методов, а для проверки привлекается случайная выборка респондентов. Цель психолингвистики — описание основных когнитивных механизмов, задействованных в обработке

языковых единиц. Таким образом, мы имеем дело с противопоставлением грамматики и механизмов, задействованных в процессах понимания и порождения, которые определяются более общими познавательными способностями человека.

Особого внимания заслуживает обсуждение, предпринятое в работе [Lewis, Phillips 2015], касающееся того, насколько в действительности отличаются явления, которыми занимаются две названные области. В частности, авторы ставят вопрос о соотношении грамматических теорий и моделей обработки языка (т.н. *language processing models*). Другими словами, авторы задаются вопросами: задействована ли грамматика только в процессах порождения и понимания речи или может служить какой-то иной цели; имеют ли дело механизмы, задействованные в обработке языковых представлений, только с репрезентациями, построенными по правилам грамматики, или с какими-то еще. К. Филлипс и Ш. Льюис формулируют эту теоретическую проблему, используя достаточно абстрактный термин *когнитивная система*, под которым они понимают совокупность когнитивных механизмов, которые определенным образом оперируют языковыми представлениями.

Исторически первой является гипотеза о том, что грамматические теории и модели обработки языка представляют собой две различные когнитивные системы. В частности, грамматика является фиксированной системой знаний о языке (и соответствует результатам применения опосредованных экспериментов), а система обработки языковых единиц представляет собой совокупность механизмов, отвечающих за процессы понимания и порождения речи (а следовательно, проявляется при применении непосредственных экспериментальных методик). Основная аргументация в пользу такой позиции связана с идеей о том, что в основе языковой способности человека лежит умение рекурсивно комбинировать концепты. Поэтому вполне вероятно, что умение оперировать сложными представлениями в процессе эволюции складывалось независимо от умения использовать эти сложные представления в коммуникации [Berwick et al. 2013; Jacob 1977]. Проблема данного подхода состоит в том, что необходимо постулировать механизм взаимодействия двух автономных систем. Кроме того, если системы автономны, неясно, почему результаты офлайн и онлайн экспериментов могут не различаться. Оказываются недостаточно аргументированы основания для установления однозначного соответствия между системами и типами экспериментальных данных: критика может касаться того, что мы не можем наверняка сказать, в какой момент времени завершается обработка языкового выражения и действи-

тельно ли непосредственные методики всегда имеют дело с процессом обработки языка, а опосредованные — с созданным представлением.

Альтернативная гипотеза состоит в том, что грамматические теории и модели обработки языка представляют собой единую когнитивную систему. При этом грамматика есть некая абстрактная совокупность лексикона и правил построения языковых единиц, а система обработки языковых единиц объединяет механизмы, которые определяются общими особенностями когнитивных функций человека (объем рабочей памяти, когнитивный контроль и пр.) и отвечают за экстерниоризацию структур, которые лицензируются грамматикой. При этом гипотеза о единой системе допускает различия между онлайн и офлайн данными, но только в том случае, если ошибки механизмов обработки могут быть систематически предсказаны теорией. Можно было бы считать «ошибками» всякое несоответствие, но в таком случае языковая теория была бы слаба и не обладала бы никакой предсказательной силой. Проблема, с которой сталкивается данный подход, заключается в том, что под влиянием процессов обработки языковой информации иногда порождаются или успешно понимаются структуры, которые не лицензируются грамматикой. Например, это происходит при т.н. эффектах притяжения согласовательных признаков (*agreement attraction effects*), когда человек не замечает ошибку в согласовании глагола с именной группой из-за того, что в предложении есть другая именная группа с теми же признаками, которые демонстрирует глагол ([Bock, Miller 1991] и др.).

Тот факт, что мы не можем однозначно принять ту или иную гипотезу в качестве верной, связан с более общей проблемой наук о человеке, а именно с *проблемой черного ящика*. Основным источником данных для синтаксической теории — суждения о грамматичности предложений. Это связано с тем, что идеальная теория должна уметь порождать грамматичные предложения, но не допускать при этом порождения неграмматичных структур. Однако у исследователя нет прибора, который бы мог измерять свойства когнитивных объектов, а именно грамматичности. Оценка грамматичности или приемлемости, которую озвучивает респондент, совсем не обязательно является истинным уровнем грамматичности в рамках когнитивной языковой системы. Из-за проблемы черного ящика мы не можем ни отделить механизмы, связанные с обработкой речи, от собственно знания правил построения языковых выражений, ни показать, что механизмы порождения/понимания и языковая способность суть одно и то же¹⁴.

¹⁴ Подробнее о взгляде на синтаксический эксперимент в узком смысле с точки зрения психолингвистики см. в обзорной статье [Федорова 2013].

3.2.3. Достоверность экспериментальных результатов

В связи с обсуждением использования экспериментальных данных возникает проблема достоверности эмпирических оснований синтаксической теории: вопрос в том, насколько надежны оценочные данные, действительно ли оценка передает грамматичность языковой структуры и в какой степени формальный сбор суждений может усовершенствовать эмпирическую базу синтаксиса по сравнению с неформальными и интроспективными методами.

3.2.3.1. Надежность суждений о приемлемости

В работе [Gibson, Fedorenko 2010] была предпринята критика традиционных методов вынесения суждений. В частности, были рассмотрены три синтаксических явления, которые в свое время широко обсуждались в литературе по генеративному синтаксису¹⁵, но суждения по которым, по утверждениям авторов, не соответствуют действительности. Э. Гибсон и Э. Федоренко провели формальное исследование с извлечением суждений и показали, что ни в одном из экспериментов не обнаружилось статистически значимого различия между двумя условиями ни для одного из явлений. Из этого авторы сделали вывод о том, что данные, на которых строятся синтаксические теории, ненадежны.

¹⁵ В качестве материала [Gibson, Fedorenko 2010] использовали три явления, которые образуют минимальные пары. Во-первых, использовался факт большей приемлемости предложений типа (ib), в которых относительная клауза занимает крайнюю правую позицию, в сравнении с предложениями типа (ia), в которых относительная клауза вложена в центр матричной клаузы. Во-вторых, было использовано противопоставление, представленное в (ii): предложение с тремя вопросительными *wh*-словами оказывается более приемлемым в результате эффектов превосходства (*superiority effect*). В-третьих, использовалось противопоставление конструкций вида (iii), в котором приемлемость (iiia) также определяется эффектами превосходства.

- (i) a. **The man [that the woman [that the dog bit] likes] eats fish.*
b. ?*I saw the man [that the woman [that the dog bit] likes].*
- (ii) a. **I'd like to know where who hid it.*
b. ?*I'd like to know where who hid what.*
- (iii) a. *What do you wonder who saw?*
b. **I wonder what who saw.*

Тем не менее, аргументация, представленная Э. Гибсоном и Э. Федоренко, уязвима для критики. Согласно работе [Sprouse, Almeida 2013a], данные, которые использовали авторы исследования, выбраны не беспристрастно: три явления проанализированы потому, что оценки по ним не удалось воспроизвести. Чтобы ответить на поставленные вопросы о надежности оценочного материала, надо проанализировать либо все данные, которые используются при построении синтаксической теории, либо случайную выборку данных. Только проделав такое исследование, можно будет сказать, какой процент ненадежных данных используется в синтаксисе и как этот показатель соотносится с аналогичными засвидетельствованными показателями в других науках (например, экспериментальной психологии).

Подобное исследование предприняли Дж. Спрауз и Д. Алмеида [Sprouse, Almeida 2012]. Были экспериментально исследованы языковые данные, использованные в учебнике синтаксиса [Adger 2003], в частности, 469 примеров из американского английского. Было привлечено 440 респондентов, которые оценивали предложения с использованием одной из двух методик: оценки величины стимула и категориальной оценки¹⁶. В результате оказалось, что как минимум 98% данных из учебника воспроизводимы с помощью формальных экспериментов.

В [Gibson, Fedorenko 2010] также утверждается, что если традиционные и формальные методы не совпадают в результатах, лучше предпочесть формальные методы. В связи с этим утверждением возникает вопрос, является ли эксперимент более эффективным методом, чем традиционные методы сбора данных. Чтобы исследовать способность экспериментальных методов обнаружить различие в грамматичности двух предложений, было проведено исследование [Sprouse et al. 2013]. В нем использовалась случайная выборка из языковых данных, приведенных в научном журнале *Linguistic Inquiry* за 2001–2010 годы. Дж. Спрауз и его соавторы исходили из того, что в журналах ситуация может отличаться от учебника и уровень невоспроизводимых суждений окажется выше. Тем не менее, 95% суждений оказались воспроизводимы с помощью формальных экспериментов.

В другой работе [Sprouse, Almeida 2013b] было проведено исследование статистической мощности формального и традиционного методов сбора материала. Теперь в качестве материала было выбрано 95 явлений из учебника [Adger 2003] и журнала *Linguistic Inquiry* за 2001–2010 годы. 140 респондентов выносили свои суждения членам минимальных пар при

¹⁶ Различные методики вынесения оценки о приемлемости будут подробнее рассмотрены в разделе 3.3.2.2.

помощи методик оценки величины стимула (*magnitude estimation*) и выбора между альтернативами (*forced-choice*), которые имитировали формальный контролируемый эксперимент и интроспективные суждения исследователей соответственно. Затем Дж. Спрауз и Д. Алмеида провели последовательную перегруппировку выборок объемом от 5 до 100 респондентов. Процесс симулирования выборки объемом 5 респондентов выглядел следующим образом: случайно выбирались 5 респондентов, затем выборка подвергалась статистической обработке, а потом данные два шага повторяли 1000 раз для симуляции 1000 экспериментов с выборкой объемом 5 респондентов. Затем был подсчитан процент симуляций эксперимента, в которых удалось установить различие между условиями. В результате подсчетов получена эмпирическая оценка статистической мощности для выборки объемом 5 респондентов. Подобная симуляция выборок позволяет сравнить две методики, а именно установить, для какой методики требуется меньше респондентов, чтобы достичь определенной мощности¹⁷. В результате Дж. Спрауз и Д. Алмеида установили, что традиционные методы, представленные в исследовании в виде метода выбора между альтернативами, требуют меньше респондентов, чтобы достичь мощности 80%. Другими словами, формальные методы не обязательно лучше неформальных методов.

3.2.3.2. Надежность суждений лингвистов

Однако, когда мы говорим о вынесении суждений самим исследователем, мы должны учитывать проблему когнитивного искажения (*cognitive bias*). В частности, лингвисты могут понимать теоретические последствия того или иного суждения, поэтому их приверженность определенной теории (в том числе своей) будет определять строгость и направление суждения. Оценить роль когнитивного искажения в синтаксической теории можно двумя способами. Во-первых, можно сравнить суждения лингвистов

¹⁷ Статистическая мощность эксперимента — вероятность обнаружить реально существующее различие между условиями в рамках эксперимента [Sprouse, Almeida 2013b]. Статистическая мощность выражается в процентах: например, мощность 80% означает, что в 80% случаев в эксперименте обнаружится различие между двумя условиями. В экспериментальной психологии этот показатель держится в среднем на уровне 60%. Статистическая мощность эксперимента определяется множеством факторов: выполняемым заданием, величиной различия между двумя условиями, количеством участников. При сравнении мощности экспериментов нужно учитывать все эти факторы.

с различными теоретическими взглядами и проверить, определяет ли суждения теоретическая приверженность человека. Во-вторых, можно было бы сравнить суждения лингвистов с суждениями наивных респондентов. Второй вариант проблематичен: непонятно, в каком направлении должны отличаться суждения лингвистов — должна ли оценка быть всегда ниже или выше, а может, для каждого явления происходит что-то свое. Влияние взглядов лингвистов на оценки было исследовано в работе [Dąbrowska 2010].

В качестве материала использовались предложения с нарушением ограничения сложной именной группы (**What did John make [the claim [that Mary bought _]]?*). Э. Дабровска исходила из следующей гипотезы: если проблема когнитивного искажения действительно существует, тогда генеративные лингвисты должны давать больше низких оценок, поскольку островные ограничения составляют ядро генеративной теории синтаксиса, в то время как функционалисты должны рассматривать эти примеры как явление второстепенное, не заостряя на нем внимание, а значит, оценивая менее строго.

В результате обнаружилось, что генеративные лингвисты оценили предложения выше, чем функциональные лингвисты, что противоречит изначальной гипотезе о когнитивном искажении.

Интересно, что во всех рассмотренных выше экспериментах, в которых сравнивались формальные и традиционные методы вынесения суждений, исследователи брали те данные, которые не вызывают больших сомнений. Любопытно, как быть с теми данными, которые вызывают споры, и как интерпретировать вариативность.

3.2.3.3. Интерпретация вариативности в данных

Представим себе, что исследователь пытается выяснить, есть ли различие между двумя типами предложений, проводя эксперимент методикой извлечения суждений. Чтобы это понять, он может собрать все суждения, затем для каждого экспериментального условия посчитать среднюю оценку, применить к средним оценкам статистические тесты. Тогда статистический тест может показать, что две синтаксические структуры действительно получили различные оценки. Тем не менее, то, что средние оценки будут статистически значимо отличаться, вовсе не означает, что каждый респондент оценил одну конструкцию как более грамматичную, а вторую — как менее грамматичную.

По-настоящему проблематичными оказываются ситуации, когда эффект устанавливается только для определенного процента испытуемых.

[Sprouse, Almeida 2012] выделяют три логически возможных подхода к интерпретации варьирования в данных. Во-первых, полученную вариативность можно интерпретировать как результат зашумленности данных. Так, в единичных измерениях возможен случайный шум (*unsystematic variation*), поэтому важна основная тенденция, устанавливаемая для всей выборки. Другими словами, при такой позиции не обязательно, чтобы эффект был для каждого респондента. Данный подход основывается на идее о том, что поведенческие реакции человека определяются не только экспериментальными манипуляциями, но и массой случайных факторов. Случайными факторами может быть настроение респондента, личные ассоциации с лексическим наполнением предложения, невнимательность при выполнении эксперимента и т.п. Как отмечают Дж. Спрауз и Д. Алмейда, такой подход чаще других встречается в литературе по экспериментальному синтаксису.

Вторая позиция заключается в выделении особого диалекта у респондентов, для которых не обнаруживается экспериментальный эффект. Логически вполне допустимо, что респонденты не различают две синтаксические конструкции просто потому, что в их грамматике нет каких-то ограничений.

Третий подход основывается на идее о том, что неграмматичные предложения должны всегда оцениваться как неприемлемые. Если теория предсказывает, что предложение должно быть неграмматичным, а респонденты не дают предложению полярных оценок, то теория неверна. Подобный подход предполагает, что никакой шум не может оказать влияние на конечный результат. А это значит, что, во-первых, оценки грамматичности/приемлемости напрямую отражают грамматичность предложения и не подвержены воздействию никаких других возможных когнитивных систем. Во-вторых, снижение внимания и усталость во время выполнения эксперимента не оказывают никакого влияния на его результаты. В-третьих, оценки на самом деле категориальные, хотя они как-то искусственно и могут быть расположены на шкале. Эта позиция имеет наибольшее количество исходных предположений, которые могут быть легко подвергнуты критике исследователей, ведущих работу в рамках психолингвистической традиции.

Наконец, мы хотели бы выделить четвертый способ интерпретации. Вполне вероятно, что вариативность в оценках связана с тем, что в грамматике происходят определенные изменения. Так, для одних респондентов инновация уже привычна, для других — все еще неграмматична. Данный подход отчасти напоминает вторую позицию: причина различий лежит в несхожести грамматик разных носителей.

Рассмотренные логические возможности наглядно демонстрируют, что то, как будет построена синтаксическая теория, определяется не только непосредственно эмпирическими данными, но и тем, как эти данные будут интерпретированы. Таким образом, проблема в области синтаксиса связана совсем не с эмпирическими данными, а скорее с тем, как эти данные интерпретируются и что каждая из интерпретаций может значить для теории грамматики.

3.2.3.4. Условия использования экспериментальных методов в синтаксисе

По итогам обсуждения приведенной литературы может сложиться впечатление, что формальные методы сбора данных оказываются ненужными для построения синтаксической теории. Тем не менее, мы хотим аргументировать противоположную точку зрения: формальные методы могут существенно обогатить эмпирическую базу для построения теории и разрешить дискуссии о наиболее адекватном направлении анализа из нескольких имеющихся. Важно при этом понимать, что выбор метода исследования зависит от того, какую информацию мы хотим получить, а также как мы собираемся ее интерпретировать.

Мы предполагаем, что экспериментальные методы могут понадобиться для синтаксического исследования в следующих случаях:

- Наблюдаются существенные различия в оценке одной конструкции разными людьми (как исследователями, так и неспециалистами). Например, необходимо опровергнуть не соответствующие действительности утверждения предыдущих исследователей о языковом материале.
- Необходимо установить величину различия между условиями. Величина различия в оценках приемлемости позволяет оценить «силу» различных грамматических ограничений.
- Проводится исследование конструкций, которые могли бы помочь уточнить некоторый раздел лингвистической теории, но редко встречаются в корпусах. В частности, необходимо установить границы варьирования или исследовать языковую инновацию.
- Необходимо установить причинно-следственную связь между определенными факторами.
- Проводится исследование взаимодействия грамматики с системами обработки языковой информации (*parsing systems*).

В настоящем исследовании мы имеем дело с ситуацией, когда отсутствие полноценных данных о варьировании стало причиной невозможности построить теорию, адекватную эмпирическим данным. В различных литературных источниках исследователи высказывают зачастую диаметрально противоположные суждения о языковом материале. При этом в обоих случаях вариативности корпусных данных оказывается недостаточно. По этим причинам мы обращаемся к экспериментальным методикам для установления основных параметров и границ варьирования. В частности, в этой главе представлены две методики: дополнение предложения путем раскрытия скобок и методика извлечения суждений о приемлемости предложений. Ниже мы рассмотрим методологические проблемы, которые возникают при выборе данных методик экспериментального исследования.

3.3. Методики экспериментального синтаксиса

В предыдущем разделе мы рассмотрели вопрос о том, какую роль экспериментальные данные играют в синтаксической теории и для ответов на какие вопросы они могут понадобиться. Были обсуждены методологические проблемы, которые возникают при проведении исследования с помощью эксперимента. Данный раздел посвящен методологическим проблемам, которые возникают при выборе двух экспериментальных методик: методики дополнения предложения путем раскрытия скобок и методики извлечения суждений о приемлемости предложений.

3.3.1. Эксперименты на порождение

Чтобы определить, что можно считать экспериментом на порождение, обратимся к понятиям внешней и внутренней валидности, которые были введены в предыдущем разделе. При проведении исследования экспериментатор решает несколько задач. Во-первых, он формулирует некоторый вопрос в виде гипотезы. Во-вторых, исследователь старается провести исследование, которое соответствует некоторым требованиям. В-третьих, экспериментатор стремится к тому, чтобы поведение объекта исследования, насколько это возможно, соответствовало поведению объекта в реальности. С точки зрения этих трех задач посмотрим на процесс порождения. Результаты эксперимента на порождение связаны с гипотезой с помощью исходного предположения исследователя о том, что респондент порождает грамматичную конструкцию, а следовательно, совокупность порожденных респондентами конструкций дает представление о грани-

цах допустимого в языке. В эксперименте, который имеет высокую внутреннюю валидность, респондент ставится в такие условия, что порождает конструкции, параметры которых строго задаются и контролируются исследователем. В эксперименте, который имеет высокую внешнюю валидность, респондент порождает предложения максимально естественно, не контролируя специальным образом свою речь; другими словами, он порождает то, что мог бы сказать в реальной ситуации общения.

3.3.1.1. Континуум контролируемого порождения

Поскольку одновременно увеличить и внутреннюю, и внешнюю валидность нельзя, правильно адекватно представлять себе то, как проводимый эксперимент соотносится с идеальными требованиями и с реальной языковой ситуацией. В работе [Eisenbeiss 2010] выделяется континуум методик на порождение, который простирается от тех методов, которые обладают наибольшей внешней валидностью, к тем, которые обладают максимально высокой внутренней валидностью. К первым С. Айзенбайс относит натуралистическое исследование (*naturalistic study*): ситуация записи, которая практически тождественна реальной ситуации использования языка. Неполное соответствие реальной ситуации возникает как следствие парадокса наблюдателя [Labov 1979]: объект исследования склонен менять свое поведение из-за продолжающегося исследовательского процесса, в частности, из-за присутствия исследователя, стимулов или аппаратов для записи/фиксации наблюдений. Минимизация вмешательства исследователя приводит к высокой внешней валидности. Однако она приводит и к тому, что на выходе получаются образцы речи, которые невозможно сравнивать между собой статистическими методами. Параметры порождаемых конструкций совсем не контролируются. Кроме того, эта методика слабо чувствительна к малым языковым различиям; она не подходит для изучения малочастотных конструкций; бесполезна при ответе на специфические вопросы об устройстве грамматики.

Далее в континууме расположена частично структурированная эlicitация (*semi-structured elicitation*), в которой общение исследователя с респондентом способствует порождению нужных конструкций, причем коммуникативная ситуация достаточно похожа на реальную. Данный метод подразумевает больший контроль над тем, что порождает испытуемый. Тем не менее к результатам все равно невозможно применить статистический анализ. Исследователь не может знать истинные причины «не-порождения» определенной конструкции: было это случайностью или так действительно не говорят.

Итак, самая большая сложность при исследованиях порождения — сделать так, чтобы респондент породил именно то, что нужно исследователю. Именно в этом и может помочь экспериментальная методика. При проведении эксперимента исследователь представляет респонденту ограниченный набор стимулов и предлагает с помощью этих стимулов породить конструкцию по образцу. Критически важным является вопрос о том, насколько строгим может быть такой образец. Наиболее строгим вариантом закрепления образца будет эксперимент, в котором респонденту нужно заполнить пропуски в представленном предложении.

3.3.1.2. Методика дополнения предложения и альтернативы к ней

Основная критика методик дополнения предложения связана с тем, что при выполнении задания респондент оценивает вероятность того или иного варианта из ограниченного множества альтернатив [Stowe, Kaan 2006]. Другими словами, он оценивает приемлемость варианта в конкретных условиях. Важно заметить, что рассуждение [Stowe, Kaan 2006] предполагает, что респондент всегда выбирает один «лучший» вариант. Иначе говоря, не может быть так, что есть несколько одинаково хороших альтернатив и респондент случайно выбирает одну из них. Сторонники этой позиции утверждают, что такая оценка приемлемости не сопоставима с тем выбором, который респондент делает в ситуации общения.

Тем не менее мы предполагаем, что использование этой методики может быть вполне оправдано. По нашим соображениям, трудности, которые возникнут при выборе альтернативных методов, могут порой перевешивать недостатки, возникающие при дополнении предложения путем раскрытия скобок. Кроме того, ограничение на абсолютную строгость образца может быть снято: в частности, исследователь может сформулировать экспериментальное задание таким образом, чтобы возможный ответ респондента не ограничивался изучаемой конструкцией.

Рассмотрим возможные методики порождения на примере конструкций с рассогласованием при гибридных существительных (*моя зубной врач сказала*). Наиболее строгой методикой является следующая. Респонденты получают сложные предложения, в которых в первой части представлен контекст, однозначно указывающий на пол лица, о котором идет речь. Вторая часть предложения представляет собой клаузу, содержащую именную группу и предикат, согласовательные показатели которых нас интересуют, с пропущенными окончаниями. Респондентам предлагается выписать слова с необходимыми окончаниями так, чтобы предложение стало полным. Итак, респондент последовательно читает предложение, прочитывает все атрибуты с пропусками. Респондент не знает согласо-

тельных характеристик до тех пор, пока не дойдет до гибридного существительного: чтобы продумать окончания, респонденту нужно вернуться обратно. Вопрос в том, когда он это делает. Наиболее вероятным вариантом является тот, при котором респондент сначала дочитывает предложение до конца, а уже потом выписывает соответствующие слова. Таким образом, у респондента слишком много времени. Проблему вызывает то, что слово с основными характеристиками появляется в предложении после слов, для которых мы хотим узнать согласовательные характеристики. Рассмотрим несколько возможных вариантов решения этой проблемы.

Первый альтернативный вариант состоит в том, что респондент слышит контекст, видит на экране прилагательные и предикат. Задача респондента состоит в том, чтобы составить полноценное предложение из элементов, представленных на экране.

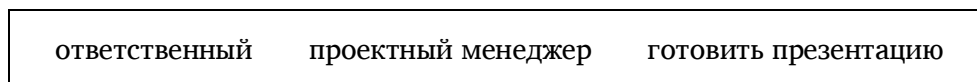


Рисунок 3.1. Сопроводительный материал для альтернативного варианта I

Сложность в этом подходе состоит в том, что исследователю трудно понять, оказывает ли исходная форма прилагательного влияние на ответ испытуемого. Для решения этой проблемы можно использовать прилагательные в разных родах, причем дополнительно можно делать так, чтобы род прилагательных не совпадал. Кроме того, чтобы понять, какое влияние оказывают различные комбинации значений грамматического рода, надо сравнивать ответы испытуемых при различном стимульном материале. Таким образом, решение приводит к экспоненциальному увеличению объема стимульного материала.

Второй вариант — это чтение вслух с методикой саморегуляции скорости. Экспериментальная процедура может быть устроена следующим образом: респондент вслух читает предложение, при этом ему заранее показывается целевое слово; прилагательное, которое нужно согласовать, подсвечивается особым образом. Таким образом, респонденту не надо «возвращаться» обратно по предложению: он сразу понимает, с чем нужно согласовать предикат.

Сложность при таком подходе заключается в том, что задание само по себе достаточно сложное и может накладывать ограничения на время реакции респондента. Более того, при таком подходе довольно легко выработать стратегию поведения: исследователь сам заостряет внимание респондента на целевом слове, и респондент может обучиться принимать одно и то же решение на протяжении всего эксперимента.

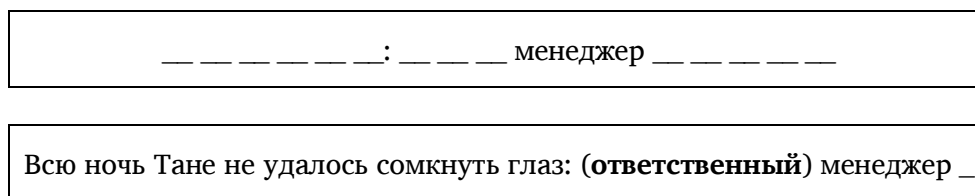


Рисунок 3.2. Иллюстрация процесса чтения с саморегуляцией скорости для альтернативного варианта II

Третий вариант эксперимента — обучение с персонажами. Так, можно представить респонденту персонажей и их изображения; затем дать инструкцию, что имена персонажей называть нельзя. Экспериментальная процедура при этом может состоять в том, что респондент видит прилагательные и предикат, а также изображение соответствующего персонажа, и должен составить естественное предложение.

Персонажи: Таня — ученый секретарь — ☺; Оля — главный редактор — ...;
Света — зубной врач — ...; Валя — финансовый аналитик — ...

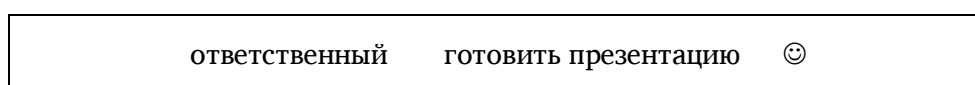


Рисунок 3.3. Сопроводительный материал для альтернативного варианта III

Такая методика является весьма трудозатратной и требует много времени для подготовки и проведения эксперимента. Кроме того, на сам эксперимент накладывается ограничение в плане объема: возможности быстрого обучения человека ограничены объемом краткосрочной памяти. Как следствие, в эксперименте с одним респондентом должно быть мало стимульного материала. Хотелось отметить также, что определить границу, после которой респондент не будет тратить время на обдумывание своего ответа, сложно, если вообще возможно. Мы не можем исключить подобное обдумывание своего ответа ни для одного из альтернативных вариантов.

Итак, мы предполагаем, что именно методика дополнения предложения является наиболее экономным решением как с точки зрения ресурсов, которые необходимо потратить на подготовку эксперимента, так и с точки зрения подверженности ответов респондента влиянию сторонних факторов.

Описанная экспериментальная процедура нацелена на то, чтобы установить распределение варьирования в речи респондентов. В то же время споры в литературе в основном затрагивают оценки конструкций с раз-

личными вариантами. Чтобы приблизиться к разрешению споров о данных, проводится другое контролируемое исследование с применением методики извлечения суждений. В следующем разделе мы подробнее обсудим особенности использования данной методики.

3.3.2. Эксперименты на оценку приемлемости/грамматичности

Эксперименты на оценку приемлемости/грамматичности можно отнести к экспериментам, в которых исследуется понимание или восприятие респондентом определенного языкового выражения.

Методика оценки грамматичности (*grammaticality judgment*) — методика, при которой исследователь явно спрашивает испытуемого, возможно ли определенное языковое выражение в языке. Так, Н. Хомский предполагает, что проверить адекватность грамматики, то есть *грамматичность* языковых выражений, которые этой грамматикой порождаются, позволяет суждение носителей о том, считают ли они данное языковое выражение приемлемым [Chomsky 1957: 13]. Как замечает [Schütze 1996], существует некоторая непоследовательность в англоязычной терминологии. В частности, грамматика есть мысленный конструкт (*mental construct*), который недоступен для сознательного восприятия. Ввиду этого респонденты не могут осознавать статус предложения относительно грамматики. Далее, методика не предполагает, что респондент как-либо рассуждает о подходящем ответе. С учетом этих двух замечаний, более адекватным термином был бы реакция на приемлемость (*acceptability reactions*). Исходно предполагается, что оценка приемлемости/грамматичности — это некоторый продукт восприятия, который самопроизвольно проявляется как ответная реакция на лингвистический стимул. Приемлемость/грамматичность похожа на любой другой продукт восприятия: яркость, громкость, температура, боль, — в том, что не существует прямых методов этот продукт измерить, ведь он существует только в рамках человеческого рассудка.

3.3.2.1. Противопоставление грамматичности и приемлемости

При обсуждении суждений респондентов неизбежно возникает вопрос о том, в чем заключается отличие грамматичности и приемлемости, и что именно оценивается в эксперименте¹⁸. [Schütze, Sprouse 2014] предлагают считать грамматически правильными (*well-formed*) цепочки слов, которые

¹⁸ Обзор исторического развития взглядов на соотношение грамматичности и приемлемости представлен в работе [Федорова 2013].

могут быть порождены некоторой (ментальной) грамматикой. Грамматика представляет собой когнитивную сущность, недоступную для непосредственного доступа. Таким образом, «грамматическая правильность» апеллирует не к поведенческой, а к внутренне-когнитивной стороне языковой деятельности. Грамматичность противопоставляется приемлемости, которая относится к поведенческому аспекту языковой деятельности и определяется через грамматическую правильность: приемлемыми считаются такие языковые единицы, которые носители языка признают в качестве правильных.

Грамматичность является лишь одним из факторов, которые определяют приемлемость предложения. Суждения относительно приемлемости предложений могут быть связаны с самыми разнообразными факторами употребления — такими как объем рабочей памяти испытуемого, частотность слова, длина предложения, сложность предложения и проч., поэтому эти суждения очень сложно правильно проинтерпретировать.

Если оценка грамматичности является осознанным продуктом человеческого восприятия, значит ли это, что уменьшается внешняя валидность эксперимента? Отчасти это действительно так: оценка грамматичности требует осознанного подхода респондента к языку. Так, респондент должен осознавать свою реакцию (продукт восприятия) и разумно сообщать о ней исследователю. В то же время оказывается, что оценка человеком собственных ощущений систематична. Именно этот факт и позволяет строить фальсифицируемые теории на основании оценок носителей языка.

Оценки приемлемости предоставляют информацию о грамматичности выражений, которые никогда не порождались в реальности. В первую очередь это касается тех явлений, которые проявляются в языке очень редко. Другими словами, нет иных методов узнать что-либо об этих явлениях. Более того, оказывается, что оценка грамматичности может быть свидетельством активности языковой способности у людей, которые не проявляют этой способности при выполнении каких-либо других заданий [Schütze 2011].

Сторонники использования данного метода также утверждают, что оценки приемлемости обладают преимуществом перед спонтанным порождением: речь говорящего/пишущего нередко содержит случайные ошибки. К ошибкам в данном случае относятся оговорки, опечатки. В результате возникают предложения, которые не должны лицензироваться грамматикой. Метод извлечения суждений позволяет исследователю задать человеку вопросы о том, что тот только что породил. Кроме того, суждения позволяют исследовать явления в ситуации ограниченных ресурсов: при отсутствии лабораторного оборудования, корпусов текстов.

Поскольку вынесение суждений является поведенческой методикой, которую довольно легко осуществить, извлечение суждений проводилось лингвистами по большей части неформально. При неформальном подходе исследователь свободен в выборе людей, которых он спрашивает, количестве предложений, формулировке инструкции и т.п. Формальный же подход предполагает ответы на ряд вопросов: в частности, нужно определиться с типом задания (шкалой оценки), количеством респондентов, допущением к участию лингвистов, количеством лексических вариантов одной конструкции, квантитативным анализом результатов.

3.3.2.2. Методики вынесения оценки

Методы оценки можно в первую очередь разделить на нечисловые и численные [Schütze, Sprouse 2014]. Это разделение имеет прямую взаимосвязь с тем, на какие вопросы может помочь ответить та или иная методика. Нечисловые методы используются для того, чтобы установить качественное различие между условиями. В то же время они не позволяют измерить степень этого различия. К таким методам относятся выбор между альтернативами (*forced-choice task*) и категориальная оценка (да/нет) (*yes-no task*). Численные методы, напротив, позволяют установить степень различия между предложениями, но могут уступать нечисловым методам в определении тонких различий между условиями. К численным методам относятся шкала Ликерта (*Likert scaling*), оценка величины стимула (*magnitude estimation*), метод термометра (*thermometer task*).

Ниже мы представим краткую характеристику каждого из указанных методов, составленную по работе [Schütze, Sprouse 2014]. При выборе между альтернативами (*forced-choice task*) респонденту представляется два предложения, из которых ему нужно выбрать одно, более (менее) приемлемое. Таким образом, методика позволяет эксплицитно сопоставить два или большее количество условий и прямо ответить на качественный вопрос о том, есть ли между этими условиями различие. При задании на категориальную оценку (да/нет) (*yes-no task*) респондентам представляется одно предложение, о котором нужно сказать, является ли оно приемлемым или нет. При использовании шкалы Ликерта (*Likert scale*) респонденту предоставляется пронумерованная шкала, крайние значения которой соответствуют приемлемому и неприемлемому предложению. Респондент должен оценить предложение по данной шкале. Обычно используются шкалы от 1 до 5 и от 1 до 7. Выбор таких шкал основан на том, что на них

можно установить середину. Тем не менее, это правило не слишком строгое: например, исследователь может использовать четную шкалу, если он сознательно исключает неопределенность, которая возникает в середине шкалы. При использовании оценки величины стимула (*magnitude estimation*) респонденту в качестве ориентира предлагается предложение, которому уже выставлена численная оценка (предложение-стандарт; абсолютное значение называется модулем (*modulus*)). Задача респондента — оценить предложение относительно стандарта. Предположение о том, что респонденты действительно сравнивают стимульное предложение со стандартом, спорно. Как указывается в работе [Sprouse 2011], респонденты используют оценку величины стимула по типу шкалы Ликерта. Почему так происходит, до сих пор не вполне понятно. Дж. Спрауз предполагает, что проблема связана с неосмысленностью нуля: ноль в таком случае подразумевает абсолютное отсутствие грамматичности. Наконец, выделяется метод термометра (*thermometer task*), который представляет собой комбинацию оценки величины стимула и шкалы Ликерта. На шкале фиксируются полярные точки, относительно которых респондент должен расположить свое суждение о предложении.

Экспериментальное сопоставление перечисленных методов было проведено в работе [Sprouse, Almeida 2017]. Дж. Спрауз и Д. Алмеида выбрали 50 синтаксических явлений, на каждое из которых было придумано два предложения — приемлемое и неприемлемое. При этом использовалось 4 метода оценки приемлемости: категориальный, шкала Ликерта, оценка величины стимула и выбор между альтернативами. При этом для каждого явления была определена степень различия между двумя условиями: по материалам предыдущего исследования Дж. Спрауза была подсчитана метрика *Cohen's d*, равная отношению разницы между средними оценками двух условий к обобщенному стандартному отклонению для двух условий. Для оценки статистической мощности эксперимента была проведена симуляция выборок от 5 до 100 респондентов. Процесс последовательной перегруппировки выборок был представлен ранее, в разделе 3.2.3.1. Экспериментальное исследование показало, что методика выбора между альтернативами особенно действенна для эффекта малого и среднего размера. Это логично, ведь при этой методике респондент фактически говорит, есть ли различие между двумя условиями. Шкала Ликерта и оценка величины стимула продемонстрировали практически равные результаты. Это подтверждает гипотезу Дж. Спрауза о том, что оценка величины стимула мало отличается от традиционного использования шкалы.

3.3.2.3. Бинарность vs. градуальность приемлемости/грамматичности

Одна из центральных проблем экспериментального синтаксиса касается источника градуальности (*gradience*) в оценках приемлемости. Как отмечается в [Featherston 2005], поскольку оценки приемлемости являются эмпирической основой для синтаксического моделирования, лингвистические теории должны объяснить градуальность приемлемости.

Выделяется две основных линии рассуждения [Schütze, Sprouse 2014]. Во-первых, можно предположить, что грамматика по своей природе категориальна. Другими словами, лингвистическая теория представляет собой модель, которая различает лишь грамматически правильные и грамматически неправильные предложения. Тогда градуальность возникает как следствие некатегориальной природы других когнитивных систем человека, которые оказываются задействованы при вынесении суждений о приемлемости предложения. Параметрами таких когнитивных систем могут выступать как такие базовые когнитивные характеристики, как объем рабочей памяти, так и отражение приобретенного опыта: вероятность описываемой ситуации, частотность лексем в составе предложения или частотность синтаксических конструкций и т.п.

Второе направление рассуждения подразумевает, что грамматика градуальна по своей природе. В пользу такой идеи К. Шютце [Schütze 2011] приводит данные о том, как человек вообще использует шкалу при оценивании. Эмпирические наблюдения исследователей, занимающихся экспериментальным синтаксисом, показывают, что респонденты последовательны в своих оценках по шкале. Кроме того, разработаны методы, которые позволяют выделить бесконечное количество уровней приемлемости, которые респонденты успешно используют (в частности, К. Шютце имеет в виду методику оценки величины стимула (*magnitude estimation*)). Тогда само понятие грамматичности есть континуум, по которому с течением времени могут перемещаться конструкции естественного языка.

К сожалению, как показывает сам К. Шютце, способность использовать градуальную шкалу относительно определенного объекта вовсе не подтверждает градуальную природу этого объекта. В частности, [Armstrong et al. 1983] и [Barsalou 1987] демонстрируют, что люди могут давать систематическую градуальную оценку относительно любого концепта, даже такого категориального концепта, как четность числа. Судя по всему, при оценке по шкале в процесс принятия решения вмешиваются какие-то другие факторы, помимо непосредственного умения человека

определять, является число четным или нет. Помимо этого, замечен следующий факт: респондент имеет склонность приравниваться к заданию, которое дает исследователь. В результате возможна реинтерпретация задания: касательно четности респондент может в действительности оценивать, насколько свойства данного числа являются типичными для четных чисел вообще. Если подобные закономерности наблюдаются при оценке четности, вполне может быть, что то же самое происходит с грамматикой, которая на самом деле является категориальной.

3.4. Заключение

Итак, в настоящей главе обосновывается использование лингвистического эксперимента для построения синтаксической теории. Для обоснования использования экспериментальных методов в синтаксисе мы определили место экспериментального синтаксиса относительно других лингвистических дисциплин, в частности психолингвистики. В сопоставительном аспекте были рассмотрены различные экспериментальные методики, были выделены их сильные и слабые стороны, а также общетеоретические проблемы, которые возникают при использовании экспериментальных данных различного типа при синтаксическом моделировании.

Экспериментальные методики различаются по «чувствительности»: методика порождения более тонко улавливает возможные отклонения от грамматического образца, а методика оценки приемлемости позволяет нивелировать элемент случайности и оценить распространенность инновации в языковом сообществе. Отдельно стоит отметить, что экспериментальные методики, особенно в сочетании с корпусными, могут быть использованы для собственно лингвистической оценки состояния языка, его статуса как находящегося вне опасности или потенциально угрожаемого. Все эти соображения позволяют рассматривать экспериментальные методики как мощный инструмент исследования внутриязыкового варьирования и его параметризации.