

© 2013 г. О.В. ФЕДОРОВА

ОБ ЭКСПЕРИМЕНТАЛЬНОМ СИНТАКСИСЕ И О СИНТАКСИЧЕСКОМ ЭКСПЕРИМЕНТЕ В ЯЗЫКОЗНАНИИ*

Настоящая статья представляет собой аналитический обзор методологических исследований, посвященных теории и практике использования метода вынесения/извлечения суждений о грамматической правильности/приемлемости предложений в современной синтаксической традиции¹. Хотя идея оценивать выражения с точки зрения их грамматической правильности или неправильности принадлежит не Хомскому и уходит корнями в глубокую древность², именно генеративный синтаксис в первую очередь ассоциируется с подобным способом работы с языковым материалом. В пятидесятилетней истории развития генеративной грамматики можно выделить три всплеска методологического интереса к этому вопросу: (1) 70-е годы XX века – первая волна методологических инноваций и критики; (2) середина 90-х годов XX века – вторая волна, породившая новое направление под названием «экспериментальный синтаксис»; (3) конец 10-х годов XXI века – третья волна, приведшая к возвращению ко многим исходным идеям. В настоящем обзоре мы последовательно опишем эти три периода, делая дополнительный акцент на сравнении данной парадигмы с парадигмой экспериментальной психолингвистики. В заключении будет дана авторская оценка результатов и перспектив данного направления исследований.

Ключевые слова: грамматическая правильность, грамматическая приемлемость, вынесение суждений, извлечение суждений, синтаксис, эксперимент

This article surveys judgments about the grammaticality or acceptability of sentences that are the most widely used data source in the syntactic literature, with a particular focus on the relevance of the evidence to the goals of generative grammar. After a background section overviewing the main notions and objectives of that framework, we can trace the field's history in three major periods that are considered in the three subsequent sections: (i) Formative period, (ii) Experimental syntax period, and (iii) Old ideas revisited period. Typically, syntacticians rely on their own judgments, or those of a small number of colleagues; sometimes they rely on the Experimental syntax paradigm. The perspective adopted here is considered in the two last sections.

Keywords: grammaticality judgment, acceptability judgment, elicitation, syntax, experiment

* Работа выполнена при финансовой поддержке гранта РГНФ № 11-04-00153. Автор выражает благодарность Е.А. Лютиковой, А.А. Кибрику, В.А. Плунгяну, М.В. Фаликман, а также анонимному рецензенту за критические замечания и советы, высказанные при подготовке работы.

¹ Данная работа будет ограничена областью синтаксиса. Необходимо особо отметить, что исследования с использованием методологии вынесения/извлечения суждений о приемлемости проводятся также в области семантики, однако подобные исследования представляют собой, на наш взгляд, относительно самостоятельную парадигму со своей собственной методологией.

² Согласно работе [Householder 1973: 365] деление предложений на грамматически правильные и грамматически неправильные можно найти уже у древнегреческого грамматика Аполлония Дискола (II в. н.э.). Кроме того, мы согласны с мнением анонимного рецензента настоящей статьи, что полезно различать имплицитную и эксплицитную методологию. Несомненно, имплицитное деление предложений на правильные и неправильные, а также спорадическое использование отрицательного языкового материала при изучении грамматических явлений существует с момента появления первых грамматик. В дальнейшей работе, говоря о методологии, мы будем иметь в виду только эксплицитно декларируемую методологию.

1. ВВЕДЕНИЕ: КОМПЕТЕНЦИЯ И УПОТРЕБЛЕНИЕ, СИНТАКСИС И ПСИХОЛИНГВИСТИКА, ГРАММАТИЧЕСКАЯ ПРАВИЛЬНОСТЬ И ПРИЕМЛЕМОСТЬ

Традиционное для структурной лингвистики первой половины XX в. соссюровское деление речевой деятельности (фр. *langage*) на язык (фр. *langue*) и речь (фр. *parole*) [Соссюр 2004] в генеративной грамматике Хомского трансформируется в дихотомию 'языковая компетенция/языковое употребление'³ [Chomsky 1965]. **Компетенция** (англ. *competence*), постулируемая Хомским, – это знание языка, или, иначе говоря, модель (ментальной) грамматики языка, представленной в виде системы правил, которые позволяют порождать и понимать бесконечное количество правильных синтаксических структур, имеющих в данном языке, и отвергать все неправильные. **Употребление** (англ. *performance*) – это процесс использования знания языка, то есть компетенции, в реальной речевой деятельности⁴.

Как Соссюр, так и Хомский полагают, что лингвистам «в собственном смысле слова» следует в первую очередь заниматься языком/компетенцией, а изучение речи/употребления относится скорее к области другой науки⁵. Этой другой наукой по праву считается психолингвистика, которая возникла, как и генеративный синтаксис, в середине XX в. Как и в истории генеративного синтаксиса, роль Хомского в истории развития психолингвистики трудно переоценить. Однако отношения Хомского с психолингвистикой сложнее. С одной стороны, всеми признается тот факт, что работы Хомского оказали большое влияние на современную психологию и психолингвистику; более того, с точки зрения Хомского, «на уровне универсальной грамматики лингвист пытается выявить некоторые общие свойства человеческого разума. Лингвистика в таком случае есть не что иное, как просто раздел психологии, изучающий эти аспекты мышления»⁶ [Chomsky 1968: 24]. Как отмечается практически во всех психолингвистических учебниках, первая книга Хомского, «Синтаксические структуры» [Chomsky 1957], легла в основу современной американской психолингвистики.

С другой стороны, сам Хомский относится к идее проведения экспериментальных исследований⁷ для подтверждения психологической реальности описаний языка, сделанных формальными лингвистами, весьма скептически. В частности, он пишет: «Если

³ В данной статье мы не будем касаться дихотомии E-language vs. I-language, введенной в работе [Chomsky 1986].

⁴ Многие функциональные и когнитивные лингвисты считают само противопоставление 'компетенция / употребление' не очень осмысленным (см., например, [Hopper 1987]). Апеллируя к подобным функциональным исследованиям, Р. Лангакер [Langacker 1987] ввел в практику термин «модель, ориентированная на употребление» (англ. *usage-based model*). Однако в данной работе мы исходим из того, что само по себе противопоставление «компетенция / употребление» является по крайней мере полезным.

⁵ Таким образом, на периферии лингвистики большей части XX в. оказалась такая лингвистическая дисциплина, как дискурсивный анализ, получивший развитие значительно позднее, чем другие уровневые дисциплины – фонетика / фонология, морфология и синтаксис.

⁶ Перевод взят из [Слобин, Грин 2004: 226].

⁷ Следует сразу оговориться, что ключевые термины настоящей работы – «эксперимент», «экспериментальный» – вызывают в среде лингвистов много разночтений. На наш взгляд, традиция понимания эксперимента в смысле Щербы (описанию этой традиции будет посвящен раздел 5 настоящей статьи) продолжает доминировать в лингвистических кругах и в настоящее время. В данной работе мы будем придерживаться общенаучного определения эксперимента как «метода познания, при помощи которого в контролируемых и управляемых условиях исследуются явления действительности» (цитата из БСЭ, издание 1978 г., т. 30, с. 6). Наиболее важным для понимания сути экспериментального исследования является положение, что в любом эксперименте необходимо сравнение изучаемого экспериментатором условия с контрольным условием, при котором экспериментальных манипуляций не происходит. Именно по этому критерию любой вид интроспекции или спорадический вопрос к информанту экспериментом в строгом смысле слова не является, подробнее см. раздел 5.

мы будем располагать такими операциональными тестами, которые помогут нам доказать правильность наших интроспективных суждений относительно очевидных явлений, это будет полезно для оценки других конкретных наблюдений»⁸ [Chomsky 1964: 80]. То есть если экспериментальные данные подтверждают некоторые очевидные случаи, в которых лингвисты сами не сомневаются, то аналогичные эксперименты могут быть использованы и для установления истины в менее очевидных случаях, но не более того. Подобные высказывания Хомского даже приводят некоторых психолингвистов к идее о том, что «попытки проверки валидности разных аспектов генеративной грамматики в психолингвистических экспериментах – а таков по преимуществу американский проект психолингвистики – это попытки с заведомо негодными средствами» [Фрумкина 2009].

* * *

С момента зарождения генеративизма основным способом получения генеративных языковых данных является метод вынесения суждений о грамматической правильности/приемлемости предложений. Во многих исследованиях, начиная с работы [Chomsky 1965], было показано, что термин «вынесение суждений о грамматической правильности» (англ. grammaticality judgment) не идентичен термину «вынесение суждений о приемлемости» (англ. acceptability judgment) и является неудачным, так как апеллирует не к поведенческой, а к внутренней когнитивной стороне языковой деятельности. Грамматика, являясь некоторой когнитивной сущностью, нам непосредственно недоступна. В частности, в работе [Newmeyer 2007] автор отмечает, что «вынесение суждений о грамматической правильности, строго говоря, есть теоретическое суждение, а вовсе не языковые данные» (перевод мой. – О. Ф.). В дальнейшем изложении мы будем придерживаться подобного взгляда и называть **грамматически правильными** или **грамматичными** предложениями такие цепочки слов, которые могут быть порождены некоторой (ментальной) грамматикой. **Приемлемыми** мы будем называть такие цепочки, которые носители языка (включая лингвистов) признают в качестве правильных предложений своего языка⁹.

2. ПЕРВАЯ ВОЛНА: БЕСЦВЕТНЫЕ ЗЕЛЕННЫЕ ИДЕИ СПЯТ ЯРОСТНО

Как это часто бывает с революционными идеями, несмотря на то, что некоторая критика генеративного подхода появилась практически сразу (см., например, работу [Hill 1961], русский перевод [Хилл 1962]), более серьезное критическое обсуждение началось только в 70-е годы. В этом разделе будут использованы критические работы У. Лабова, Д. Слобина и Дж. Грин¹⁰. Однако начнем мы его с более подробного описания важности в генеративном синтаксисе отрицательного языкового материала.

⁸ Перевод взят из [Слобин, Грин 2004: 59].

⁹ Необходимо отметить, впрочем, что в научной литературе существуют и другие определения терминов «вынесение суждений о грамматичности» и «вынесение суждений о приемлемости». Так, в работе [Bader, Häussler 2010] авторы используют первый термин в том случае, когда проводят с испытуемыми эксперимент с бинарной шкалой оценок, то есть с двумя значениями: 'грамматичный' и 'неграмматичный', а второй – когда шкала оценок более, чем бинарна [Bader, Häussler 2010: 278].

Кроме того, среди «рядовых» генеративных лингвистов термин «грамматичность» распространен намного шире, чем термин «приемлемость»; в частности, в работе [Featherston 2005] отмечается, что 23 из 30 авторов популярного учебника «Handbook of contemporary syntactic theory» [Baltin, Collins 2001] используют именно этот термин, даже не давая его определение, а термин «приемлемость» встречается только у одного из авторов этого учебника [Featherston 2005: 673]. Тем не менее в данной статье мы будем придерживаться вышеприведенных формулировок.

¹⁰ Работы Слобина и Грин будут цитироваться по изданию [Слобин, Грин 2004]. Переводы [Slobin 1971] и [Greene 1972] были впервые опубликованы на русском языке в 1976 г. [Слобин, Грин 1976]. Работы Лабова [Labov 1972; 1975] будут цитироваться по более поздней работе [Labov 1996].

Итак, основу языковой компетенции составляет способность носителя языка отличать грамматически правильные (англ. well-formed) предложения некоторого языка от грамматически неправильных (англ. ill-formed). Таким образом, **отрицательный языковой материал** (который может быть получен только в результате вынесения суждений относительно приемлемости, так как не может быть представлен в корпусе) является исключительно важным для исследователей-генеративистов при построении их моделей¹¹. Важность отрицательных данных в генеративных теориях определяется также следующим фактом. С одной стороны, мы считаем, что носители языка в большинстве случаев соглашались в оценке приемлемости тех или иных предложений. С другой стороны, большинство исследователей детской речи полагают, что вопреки распространенному мнению родители крайне редко исправляют грамматические ошибки своих детей (известный аргумент от **бедности стимула**, англ. poverty of the stimulus). Таким образом, отсутствие достаточных отрицательных данных вкупе с наблюдением, что усвоение языка влечет за собой способность отличать грамматически правильные предложения от неправильных, приводит Хомского и его последователей к самому главному постулату генеративизма – что существует некоторое врожденное знание языка, которое определяет взрослую способность порождать правильные предложения и не порождать неправильные. Лингвистическую модель этого врожденного компонента традиционно называют Универсальной грамматикой (далее УГ)¹².

В современной лингвистической практике неприемлемые предложения обычно обозначаются знаком звездочки перед соответствующим примером. По всей видимости, авторство этого обозначения принадлежит американскому лингвисту Ф. Хаусхолдеру – в статье «On arguments from asterisks» 1973 года он сам описывает, как летом 1958 года, читая курс морфосинтаксиса студентам Мичиганского лингвистического института, ввел в обиход это обозначение [Householder 1973: 365].

Грамматическую правильность, по Хомскому, следует отличать от семантической осмысленности. Хрестоматийный пример такого рода – грамматичное, но семантически бессмысленное (англ. nonsensical) предложение (1) *Colorless green ideas sleep furiously* (*Бесцветные зеленые идеи спят яростно*). Если прочитать это предложение в обратном порядке – (2) *Furiously sleep ideas green colorless* (*Яростно спят идеи зеленые бесцветные*) – то оно станет не только бессмысленным, но и неграмматичным [Chomsky 1957: 15; Хомский 1962: 418].

Этим примером Хомский аргументировал независимость грамматики от семантики; однако Р.О. Якобсон, полемизируя с Хомским, показал, что грамматически правильное предложение *Colorless green ideas sleep furiously* может быть метафорически прочитано и как семантически осмысленное [Jakobson 1959/1971: 330]. В качестве дополнительного аргумента Якобсон приводит опубликованное в 1957 г. вполне осмысленное стихотворение Д. Хаймса под названием «Colorless green ideas sleep furiously»¹³.

Тот факт, что грамматическую правильность предложения можно оценивать независимо от его семантики, иногда демонстрируют и при помощи так называемых «пред-

¹¹ В отечественной лингвистике идея о важности учета отрицательного языкового материала в первую очередь ассоциируется с работой Л.В. Щербы «О тройном аспекте языковых явлений и об эксперименте в языкознании» 1931 г., в которой он пишет: «Но особенно поучительны бывают отрицательные результаты: они указывают или на неверность постулированного правила, или на необходимость каких-то его ограничений, или на то, что правила уже больше нет, а есть только факты словаря...» [Щерба 1931/1974: 38].

¹² Заметим, однако, что, имея правильные и неправильные примеры одного конкретного языка, мы не можем реконструировать УГ, так как некоторые грамматические конструкции некоторого конкретного языка могут быть запрещены, потому что так устроена УГ, а некоторые – потому что так установлены параметры данного языка. Таким образом, для реконструкции УГ необходимы типологические исследования.

¹³ Данная дискуссия не закончилась и спустя сорок лет, см. документальный фильм «Discovering the human language: Colorless green ideas» (Equinox Films; produced, directed and written by G. Searchinger. New York (NY): Ways of Knowing, Inc., 1995).

ложений Jabberwocky», то есть создавая предложения из несуществующих в данном языке слов. Подобные предложения большинство носителей оценивают как приемлемые, в отличие от предложений с конверсированным порядком тех же псевдослов. Кроме известных примеров из Л. Кэррола, в этот ряд можно поставить предложение Г. Глисона (3) *The iggle squiggs trazed wombly in the harlish hoop*, а также известный русский пример Л.В. Щербы (4) *Глокая куздра штеко будланула бокра и курдячит бокрёнка*. Вышеперечисленные предложения могут быть оценены носителем с точки зрения их грамматической правильности независимо от того, что вероятность их появления в речи равна нулю. Говоря другими словами, грамматическая правильность предложения не равна не только осмысленности, но и вероятности употребления в реальной речи.

Являясь центральным пунктом концепции Хомского, идея различения грамматически правильных и грамматически неправильных предложений присутствует уже в самой первой книге Хомского «Синтаксические структуры» [Chomsky 1957], см. примеры (5), (6) и (7). В то же время различение грамматической правильности и приемлемости было введено Хомским только в «Аспектах теории синтаксиса» [Chomsky 1965].

(5) The scene of the movie was in Chicago.

(6) The scene that I wrote was in Chicago.

(7) *The scene of the movie and that I wrote was in Chicago [Chomsky 1957: 36].

Как отмечает Хомский и его последователи, 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. По мнению Хомского, такие суждения не равнозначны суждениям о грамматической правильности, поскольку часто бывает, что высказывания, которые по той или иной причине не являются приемлемыми, тем не менее грамматически безупречны. Таким образом, грамматичность является лишь одним из факторов, который определяет приемлемость предложения. По мнению Хомского, мы можем придумать строгие экспериментальные процедуры, чтобы определить приемлемость, но мы не можем аналогичным образом определить грамматичность¹⁴ [Chomsky 1965: 11].

Еще одним принципиально важным методологическим вопросом является вопрос о бинарности или градуальности¹⁵ грамматичности/приемлемости. Признавая в ранних работах тот факт, что на самом деле грамматическая правильность предложения является «a matter of degree» [Chomsky 1957: 36; 1965: 11], Хомский идеализирует ее, сводя к бинарной оппозиции и строя модель, которая различает лишь грамматически правильные и грамматически неправильные предложения [Chomsky 1957: 13]. Что касается приемлемости, то подавляющее большинство исследователей сходятся на том, что она градуальна (в частности, это отражается в многочисленных промежуточных оценках предложений с использованием диакритик: «?», «??», «*?» и «?*»). Таким образом, большинство генеративных лингвистов разделяют мнение, что сама модель ментальной грамматики, то есть компетенция, основана на бинарных ограничениях, а градуальность привносится только на уровне употребления ментальной грамматики в речи (см. наиболее авторитетное в этой области мнение Ф. Ньюмейера [Newmeyer 1983]). Однако в последние годы вопрос о бинарности или градуальности ментальной грамматики вновь становится актуальным, см., например, стохастическую теорию оптималь-

¹⁴ Некоторые авторы, которые используют термин «грамматичность» как в смысле некоторой цепочки слов, порожденной ментальной грамматикой, так и в смысле суждения о некотором предложении, вынесенного носителем языка, пишут первое из них с заглавной буквы, то есть «Грамматичность» (в частности, см. [Featherston 2005: 673]).

¹⁵ Термин «градуальность» (англ. gradience) был впервые использован в работе [Bolinger 1961]; отметим также, что в английском языке это слово используется только как лингвистический термин. В качестве антонима к термину «градуальный» мы будем использовать термин «бинарный» (англ. categorical).

ности [Boersma, Hayes 2001], линейную теорию оптимальности [Keller 2000], а также работу [Wasow 2009]¹⁶.

Перейдем теперь к критике. Первую волну критики генеративизма можно разделить на два направления – социолингвистическое и психолингвистическое. Самая, пожалуй, серьезная критика генеративного способа получения языковых данных в 70-х гг. пришла из области социолингвистики и была представлена в работах Лабова. Сравнивая объективные социолингвистические данные с субъективными генеративными, он предложил следующие принципы работы с языковым материалом.

(8) **Принцип консенсуса:**

Если нет никаких оснований думать обратное, мы предполагаем, что суждения любого наивного носителя характеризуют всех говорящих на данном языке.

(9) **Принцип экспериментатора:**

Если в интроспективных суждениях имеются рассогласования, то суждения людей, знакомых с теоретической подоплекой исследования, не принимаются во внимание.

(10) **Принцип валидности:**

Если данные, полученные в неподготовленной речи носителей, оказываются более устойчивыми, чем данные интроспективных суждений, то валидное описание языка должно опираться на употребление, а не на интуицию.

([Labov 1996: 34], перевод мой. – О. Ф.)

Данные принципы были впервые сформулированы в первой половине 70-х гг. Глядя на прошедшее с тех пор время из середины 90-х, Лабов отмечает, однако, что большинство формальных лингвистов и сейчас не согласятся с ними, предпочитая «идеалистическую позицию генеративизма материалистическому подходу социолингвистики» [Labov 1996: 35].

Второе направление критики генеративного метода получения языковых данных было представлено психолингвистами. В те годы психолингвистика только начинала развиваться, так что даже само слово «психолингвистика» чаще заменялось словом «психология». С самых первых шагов у психологов и лингвистов возникло желание поделить сферы влияния, см. цитату из Грин: «Попытки разрешить пограничный конфликт между лингвистикой и психологией всегда приводят к проблеме разграничения языковой способности¹⁷, которая считается областью лингвистики, и языковой активности, относимой к области психологии» [Слобин, Грин 2004: 222].

Многие психологи не довольствовались отведенной им ролью изучения дополнительных факторов, которые отличают грамматическую правильность от приемлемости (другими словами, компетенцию от употребления). Как пишет Слобин, другая задача психологов, не менее важная, состояла в том, чтобы «попытаться пройти через лабиринт психологических факторов, которые делают языковую способность отличной от языковой активности, чтобы показать, что языковая способность, в том виде, как ее описывают лингвисты, имеет “психологическую реальность”, то есть существует как некий психологический феномен¹⁸» [Слобин, Грин 2004: 35].

¹⁶ По нашему мнению, на современном этапе развития науки данная дискуссия носит во многом чисто теоретический характер, так как все существующие сейчас методы проверки бинарности или градуальности грамматики, как лингвистические, так и психолингвистические, носят опосредованный характер. Однако бурно развивающиеся в последнее время методы нейровизуализации могут со временем превратить этот вопрос из вопроса веры в вопрос строгого научного знания.

¹⁷ В переводе книги Слобина и Грин [Слобин, Грин 2004] «competence» передается как «языковая способность», а «performance» – как «языковая активность».

¹⁸ Термин «психологическая реальность», введенный в лингвистический обиход Э. Сепиром в его работе о психологической реальности фонем [Sapir 1933], а позднее закрепившийся в работах Хомского, представляется нам крайне неудачным. На наш взгляд, психолингвисты, которые, как считается, занимаются гораздо более психологически реальными исследованиями, чем простые теоретические лингвисты, с точки зрения близости к «реальному положению вещей» ничем не отличаются от последних, см. аналогичное мнение в статье под названием «There's no reality like psychological reality» [Dresher 1995].

Не соглашаясь со словами Хомского о второстепенной роли психологических экспериментов, тестирующих построения генеративной грамматики, Слобин суммирует свой взгляд таким образом: «Мы можем допустить, что формальная теория где-то и как-то существует в голове носителя языка и определяет его способность действовать определенным образом. Языковая способность, таким образом, есть модель того, что, по нашему предположению, существует в сознании носителя языка, модель, созданная лингвистом на основе его интуитивной способности отличать правильно построенные высказывания от неправильно построенных и т.п. Достоверность существования такой модели может быть доказана только тщательным изучением реальной активности, которую она, по нашему мнению, определяет» [Слобин, Грин 2004: 57].

Подводя итоги описанию первого этапа развития методологии вынесения суждений о грамматической приемлемости предложений, можно заключить, что общая практика той поры состояла в том, что исследователи на основании своей интуиции оценивали сконструированные ими же предложения. Не обращая особого внимания на замечание Лабова о том, что лингвисты не могут одновременно порождать и языковые данные, и теории [Labov 1996: 33], они даже иногда использовали аналогию с дегустацией вина: подобно тому, как только самые хорошие дегустаторы могут различить тончайшие оттенки вкуса вина, только самые тонкие лингвисты могут прочувствовать все оттенки грамматики родного языка [Valian 1982].

3. ВТОРАЯ ВОЛНА: ЭКСПЕРИМЕНТАЛЬНЫЙ СИНТАКСИС

Вторая волна интереса к статусу метода вынесения суждений о приемлемости пришла на вторую половину 90-х гг. и связана в первую очередь с именами двух американских лингвистов – К. Шутце и В. Коварта, чьи работы способствовали существенному изменению подхода к данной проблеме.

Рассмотрим ту же по существу процедуру вынесения суждений о приемлемости языковых выражений, которая используется генеративными лингвистами при интроспективном методе, но применительно к (i) некоторой репрезентативной группе испытуемых с (ii) достаточным количеством языковых стимулов и (iii) более чем бинарным шкалированием возможных оценок. Если мы затем (iv) обобщаем полученные данные при помощи статистики, традиционной для области когнитивной психологии (частью которой, по мнению многих приверженцев идей Хомского, была и остается генеративная лингвистика), то, согласно идеям Шутце, Коварта и многих их последователей, мы получим уже совсем другой уровень объективности и значимости проведенного исследования. В работе [Myers 2009] эта разновидность метода вынесения суждений называется **формальным суждением** (англ. *formal judgment*), в то время как рассмотренный выше метод интроспективной оценки назван **неформальным суждением** (англ. *informal judgment*). После выхода в свет книг [Schütze 1996] и [Cowart 1997] некоторые генеративные лингвисты полагают, что вынесение формальных суждений о приемлемости является уже полноценным психолингвистическим экспериментом, так как удовлетворяет основному психологическому критерию, а именно критерию объективности (см. подзаголовок книги Коварта [Cowart 1997]: «*Experimental syntax: Applying objective methods to sentence judgments*»). Более того, на методе вынесения формальных суждений о приемлемости было основано целое направление экспериментальных лингвистических исследований, получившее название «экспериментальный синтаксис»¹⁹. В дальнейшей работе мы будем называть данный метод методом **извлечения** суждений о приемлемости (англ. *elicitation*), в отличие от метода **вынесения** суждений о приемлемости, основанного на интроспекции самого исследователя.

¹⁹ Отметим, что в современных психолингвистических исследованиях подобные тесты известны и популярны под названием **опросников** (англ. *questionnaire*).

Если первые два нововведения – увеличение количества испытуемых и количества стимулов – представляют собой простое копирование давно устоявшихся нормативов экспериментальной психологии, то на возможных методах шкалирования оценок стоит остановиться подробнее. До последнего времени наиболее распространенным способом более чем бинарного шкалирования был психометрический подход, предложенный Р. Ликертом [Likert 1932]. При работе со шкалой Ликерта испытуемый обычно оценивает степень своего согласия или несогласия с предложенным стимулом (в нашем случае – степень приемлемости предложения) по пяти- или семибалльной шкале – от значения «полностью согласен» до значения «полностью не согласен»; реже применяются шкалы с тремя или девятью делениями; иногда подобная шкала состоит из только положительных значений (например, от 1 до 5), иногда в ней имеется нулевая точка перехода (от -3 до +3). Обычно в таких экспериментах испытуемый получает в качестве образца примеры предложений, соответствующих крайним точкам на шкале (в качестве примеров подобных исследований см. [Bybee, Eddington 2006] или [Culbertson, Gross 2009]).

Однако в середине 90-х гг. «золотым стандартом» (такая оценка данного метода содержится, например, в книге [Cowan 1997]) более чем бинарного шкалирования являлся метод **оценки величины стимула** (англ. *magnitude estimation*), первоначально предложенный психофизиком С. Стивенсом [Stevens 1956] для исследования величины ощущения. В работе [Bard et al. 1996] эта методика была адаптирована к синтаксическому материалу следующим образом. В ходе экспериментальной процедуры испытуемый получает анкету, в которой приемлемость некоторого предложения, например (11) *What do you wonder whether Mary bought?* (подобное предложение называется **стандартом**), оценена в 100 баллов (это абсолютное значение называется **модулем** (англ. *modulus*)). Испытуемого просят оценить приемлемость целевого предложения, например (12) *What did Lisa meet the man that bought?*, используя в качестве ориентира модуль предложения-стандарта. Так, он может оценить целевое предложение в 50 баллов, если решит, что оно в два раза менее приемлемо, или, наоборот, в 200 баллов, если оно покажется ему в два раза более приемлемым (конкретный пример взят из работы [Schütze 2011]). Таким образом, данная методика принципиально отличается от классической шкалы Ликерта тем, что здесь модуль предложения-стандарта используется в качестве некой единицы измерения, по отношению к которой испытуемые оценивают другие предложения.

За пятнадцать лет, прошедших с момента возникновения экспериментального синтаксиса, данная идея, по существу представляющая собой некоторый компромисс между теоретической и экспериментальной лингвистиками, превратилась в достаточно солидное направление со своей методологией, сторонниками и противниками. Однако она так и не стала доминирующей. Так, с одной стороны, многие лингвисты полагают, что как неформальная, так и формальная разновидности вынесения/извлечения суждений о приемлемости безусловно проигрывают методу наблюдения в корпусе [Sampson 2007] и/или подлинной экспериментальной процедуре [Bornkessel-Schlesewsky, Schlewsky 2007]. С другой стороны, большинство генеративных лингвистов до сих пор полагают, что и неформальной методики вынесения суждений вполне достаточно для получения надежных и валидных языковых данных [Phillips 2009].

4. ТРЕТЬЯ ВОЛНА: ДВИЖЕНИЕ ПО СПИРАЛИ

Третья волна методологического интереса и разнообразной критики генеративного метода вынесения/извлечения суждений о приемлемости развивается на наших глазах. Одной из ее отличительных особенностей можно считать появление крупномасштабных теоретических дискуссий. В частности, широкий резонанс вызвало обсуждение статьи Ф. Ньюмейера «Grammar is grammar and usage is usage», опубликованной в 2003 г. в журнале «Language» [Newmeyer 2003]. В этой статье Ньюмейер выступает против подхода, ориентированного на употребление (в частности, см. [Langacker 1988;

Bybee 2006]), особое внимание уделяя критике **стохастических грамматик**, то есть грамматик, которые включают в себя информацию о вероятностях. Учитывая вероятности употребления тех или иных конструкций, мы тем самым стираем четкие границы между компетенцией и употреблением. В серии из трех ответных статей (например, [Clark 2005]) оппоненты Ньюмейера приводят примеры удачных проектов, выполненных в парадигме подхода, ориентированного на употребление (в частности, стохастическую теорию оптимальности [Boersma, Hayes 2001]).

Однако самой, на наш взгляд, интересной представляется дискуссия по поводу статьи под заголовком «Data in generative grammar: The stick and the carrot» С. Фетерстона. Этой проблематике целиком посвящен третий номер 33-го тома журнала «Theoretical linguistics» (2007 г.): в заглавной статье Фетерстон призывает генеративных лингвистов раз и навсегда отказаться от использования интуитивных суждений в пользу более серьезных методов, а шесть его оппонентов высказывают свои точки зрения в ответных статьях. Рассмотрим содержание этой дискуссии более подробно.

В заглавной статье Фетерстон сначала описывает отрицательные стороны современного положения вещей в методологии сбора материала генеративной лингвистики (апеллируя к методу кнута и пряника, данную часть он сравнивает с кнутом), а затем перечисляет бонусы-пряники, которые может получить генеративный лингвист, воспользовавшись экспериментальной методологией. Фетерстон заключает, что генеративная лингвистика, используя ненаучные методы получения языковых данных, по существу отказывается от своих собственных амбициозных намерений стать частью современной психологической науки, которая основана на экспериментальной методологии. Описывая текущее состояние дел в области взаимодействия генеративной лингвистики с психологией, Фетерстон сетует, что, как и 50 лет назад, большинство генеративных лингвистов при построении своих теорий продолжают опираться на собственные интуитивные примеры, не привлекая других носителей и основываясь на единичных примерах и бинарной оппозиции «можно-нельзя». Приводя цитату из известной работы 1982 г., он отмечает, что эта проблема остается нерешенной и теперь: «когда психологические данные не подтверждают лингвистическую теорию, психологи заключают, что лингвистическая теория ошибочна, а лингвисты – что психологические данные нерелевантны» ([Roeger 1982: 468], перевод мой. – О. Ф.).

Фетерстон приводит конкретные примеры генеративных исследований, которые не выдерживают экспериментальной проверки. В частности, он описывает собственный эксперимент, посвященный тестированию теории связывания на материале немецкого языка, в котором был использован стимульный материал из работы [Grewendorf 1988]. В работе [Grewendorf 1988: 58] автор, основываясь на собственной интуиции, предлагает такое распределение форм под звездочкой и грамматичных высказываний, см. пример (13). Однако эксперимент на извлечение суждений с привлечением 26 испытуемых, которые оценивали 16 синтаксически однотипных примеров, подобных (13), используя метод оценки величины стимула, показал прямо противоположные результаты (подробнее см. в [Featherston 2002]).

- | | | | | | |
|---------|------------|---------------|----------------------------|---------------------------------------|---------------|
| (13) a. | Der | Arzt zeigte | den Patienten _j | sich _j / *ihn _j | im Spiegel |
| | the | doctor showed | the patient.ACC | himself / him | in.the mirror |
| b. | Der Arzt | zeigte | dem Patienten _j | ihn _j / *sich _j | im Spiegel |
| | the doctor | showed | the patient.DAT | him / himself | in.the mirror |

Все вышеперечисленные недостатки генеративных работ могут быть ликвидированы, по мнению Фетерстона, если ввести в генеративную практику контролируемый психолингвистический эксперимент. Фетерстон выделяет необходимые и желательные требования к подобной экспериментальной процедуре:

- (14) необходимо:
- (1) много информантов;
 - (2) много однотипных стимулов;

желательно:

- (1) инструкция «Насколько естественно звучит это предложение?»;
- (2) более чем бинарное шкалирование²⁰ [Featherston 2007: 282].

Три из четырех перечисленных пунктов Фетерстона – за исключением особого требования к формулировке инструкции – разделяют почти все приверженцы идей экспериментального синтаксиса. Что касается инструкции, то Фетерстон аргументирует именно такую формулировку следующим образом. Более частым вариантом инструкции является вариант «Используете ли вы подобное предложение в своей речи?», однако он, как отмечает Фетерстон, недостаточно корректен по двум причинам. Во-первых, при такой постановке вопроса в голове у носителя могут возникнуть представления о норме и он может не захотеть положительно оценить синтаксическую конструкцию, которая по каким-то причинам ассоциируется у него с низким стилем речи. Во-вторых, в языке могут существовать более синтаксически простые аналоги изучаемой конструкции, которые вытесняют из практического употребления данную, хотя она и сохраняет грамматическую правильность.

Говоря о более чем бинарном шкалировании оценок, Фетерстон предлагает свою собственную процедуру подобного шкалирования, которую он называет **термометром** (англ. *thermometer judgments*). Данный метод напоминает обычную шкалу Ликерта, однако между крайними точками шкалы Фетерстона 20 (нижней точкой) и 40 (высшей точкой) испытуемый может сам для себя установить любое количество делений (подробнее см. [Featherston 2009]).

Итак, в статье Фетерстона приводятся достаточно убедительные аргументы в пользу использования формального метода извлечения суждений. Однако статьи шести его оппонентов оказываются не менее убедительными. Одна из основных идей ответной статьи Г. Фанзелова с не менее говорящим названием «Carrots perfect as vegetables, but please not as a main dish» состоит в следующем: с одной стороны, большинство лингвистов легко разделяют тезис о том, что грамматическая правильность является лишь одним из факторов, который влияет на суждение о приемлемости. Между тем в настоящее время проведено совсем мало исследований, в которых бы это строго доказывалось с перечислением основных дополнительных факторов. С другой стороны, те же самые лингвисты, разделяющие этот тезис, легко переносят данные, собранные методом извлечения суждений, непосредственно в свои модели грамматики, даже не пытаясь отсеивать эти дополнительные факторы [Fanselow 2007: 363]. Эта проблема, по мнению Фанзелова, не может быть разрешена при помощи простого усовершенствования экспериментальной методики.

Идея Фанзелова находит дальнейшее развитие в статье психолингвистов И. и М. Шлезевских «The wolf in sheep's clothing: Against a new judgement-driven imperialism». Сравнивая метод извлечения суждений с нейролингвистическими методами нейровизуализации (такими как метод регистрации вызванных потенциалов мозга и функциональная магнитно-резонансная томография), авторы приходят к заключению, что данные, полученные методом извлечения суждений, представляют собой результат взаимодействия внутренних и внешних факторов (другими словами, факторов компетенции и употреб-

²⁰ В работе [Myers 2009] описана другая разновидность формальной экспериментальной процедуры извлечения суждений, получившая название «миниатюрный экспериментальный синтаксис» (англ. *small-scale experimental syntax*). По мнению автора, данная идея должна прийтись по душе тем генеративистам, которые не хотят тратить много времени на полномасштабное экспериментальное исследование, но будут рады возможности соблюсти все формальные требования без особых затрат времени и человеческих ресурсов. Для таких исследователей Майерс создал бесплатную программу MiniJudge (адрес в Интернете <http://www.ccunix.ccu.edu.tw/~lngproc/MiniJudge.htm>), в которой любой синтаксист может провести исследование с минимальным количеством испытуемых и стимульных предложений, необходимых для статистически надежных результатов. Кроме того, в отличие от методики Фетерстона, в данной программе используется бинарное шкалирование оценок.

ления), которые трудно развести между собой [Bornkessel-Schlesewsky, Schlesewsky 2007: 331]. Таким образом, Шлезевские ставят под сомнение ценность методики, предложенной Фетерстоном, так как она, на их взгляд, не помогает получить языковые данные «чистой» грамматичности.

Наконец, в статье Ньюмейера основное внимание уделяется вопросу бинарности или градуальности грамматичности/приемлемости. Как и большинство других авторов ответных статей, он не соглашается с идеей Фетерстона о градуальности грамматики, однако справедливо заключает, что увеличение числа испытуемых никак не приблизит нас к ответу на этот сложный вопрос [Newmeyer 2007: 399].

Таким образом, с одной стороны, в рамках методологических усовершенствований третьей волны некоторые лингвисты пытаются придумать все более экспериментально выверенные методы работы с языковым материалом. Однако, с другой стороны, лингвисты, придерживающиеся противоположных взглядов, также не останавливаются на достигнутом. Так, рассмотренные ниже работы Дж. Спрауса с коллегами продвигают методологию в обратном направлении.

Спраус замечает, что в последние годы появилось немало работ, написанных в основном психолингвистами, в которых утверждается, что метод неформального вынесения суждений не обладает необходимым уровнем надежности получаемых результатов. Ошибочные результаты таких исследований могут быть двух типов: ошибочно положительные и ошибочно отрицательные. Первые случаются, когда исследование показывает наличие некоторого значимого различия, хотя на самом деле его нет (подобные претензии к неформальным суждениям предъявляются в психолингвистических статьях [Edelman, Christiansen 2003; Ferreira 2005; Wasow, Arnold 2005; Gibson, Fedorenko 2010; Gibson, Fedorenko in press]), вторые – когда исследование не показывает различия, которое на самом деле существует [Keller 2000; Featherston 2007].

Спраус с коллегами решил проверить интуитивные суждения лингвистов, приводимые в различных научных работах, используя два метода: метод оценки величины стимула и обычный для формальных лингвистов метод бинарного выбора. Сначала авторы взяли для проверки все 469 примеров из популярного учебника синтаксиса [Adger 2003] и провели с этими стимулами эксперимент с участием 440 испытуемых [Sprouse, Almeida submitted]. Различие между неформальной интуицией лингвиста и формальными методами извлечения суждений оказалось равным 2%, т.е. меньше стандартного уровня значимости психолингвистических экспериментов, равного 5%. В следующей работе Спраус с коллегами провел аналогичное масштабное (264 человека) тестирование 292 предложений, выбранных из генеративных статей, опубликованных в одном из самых престижных лингвистических журналов «Linguistic inquiry» с 2001 по 2010 г. Полученное различие оказалось ровно на стандартном уровне значимости психолингвистических экспериментов в 5% [Sprouse et al. submitted]. Таким образом, выяснилось, что по крайней мере на проверенном материале метод неформального вынесения суждений не уступает формальным методам извлечения суждений, то есть при общепринятой генеративной практике вынесения суждений особой методологической катастрофы не происходит.

Отметим еще одну важную методологическую особенность третьей волны, которая состоит в переосмыслении роли метода оценки величины стимула в качестве «золотого стандарта». О важности этого вопроса говорит уже тот факт, что две статьи, посвященные этой теме, – [Weskott, Fanselow 2011] и [Sprouse 2011], опубликованные в журнале «Language», – были признаны лучшими статьями этого журнала за 2011 г. В работе [Weskott, Fanselow 2011] авторы экспериментально проверили гипотезу о том, что метод оценки величины стимула дает лучшие результаты по сравнению с бинарным методом или использованием семибалльной шкалы Ликерта.

- (15) Peter hat erzählt dass der Präsident den Scheich empfangen hat.
 Peter has reported that the.NOM president the.ACC sheik received has
- (16) Peter hat erzählt dass den Scheich der Präsident empfangen hat.
 Peter has reported that the.ACC sheik the.NOM president received has

В серии экспериментальных исследований, описанных в этой статье, одна группа испытуемых сначала проходила эксперимент, направленный на изучение особенностей порядка слов в немецком языке (ср. примеры (15) и (16)), с использованием бинарного метода оценки; через две недели те же испытуемые оценивали те же стимулы при помощи метода оценки величины стимула. Вторая группа испытуемых, начиная с того же бинарного метода, через две недели использовала семибалльную шкалу Ликерта. Авторы пришли к выводу, что метод оценки величины стимула не дает никаких преимуществ, а при некоторых обстоятельствах его результаты оказываются даже хуже тех, которые получаются при использовании традиционных методик.

В работе [Sprouse 2011] делается попытка дать объяснение этим и подобным им результатам. Спраус полагает, что основная проблема состоит в том, что в отличие от психофизики (где возможно, например, полное отсутствие звука) синтаксис не имеет настоящей нулевой точки, которая необходима для полноценного использования данной методики. Испытуемые, получая задание оценить величину стимула, выполняют его как задание оценки по шкале Ликерта; при этом неясность инструкции создает дополнительный шум, ухудшая в целом результаты экспериментов.

Подведем итоги. Анализируя динамику развития методологии вынесения/извлечения суждений о приемлемости в последние годы, мы наблюдаем существенный отход к исходным позициям, так как «золотой стандарт», на который возлагались большие надежды, себя не оправдал, а сам метод извлечения суждений в целом также не дает никаких особых преимуществ, при том что трудо- и временные затраты, которые приходится на него расходовать, оказываются очень велики. Как кажется, в целом развитие этой методологии за 50 лет представляет собой движение по спирали, когда, испробовав более «прогрессивные» методы работы с языковым материалом, исследователи постепенно возвращаются к старым проверенным средствам.

5. СИНТАКСИЧЕСКИЙ ЭКСПЕРИМЕНТ С ТОЧКИ ЗРЕНИЯ ПСИХОЛИНГВИСТИКИ

В этом разделе мы оценим эксперимент на вынесение/извлечение суждений о приемлемости с точки зрения психолингвистики. Начнем с самого определения термина «эксперимент». Отечественная традиция экспериментирования в лингвистике идет от известной работы Щербы «О трояком аспекте языковых явлений и об эксперименте в языкознании» 1931 г., в которой автор пишет: «Большинство лингвистов обыкновенно и к живым языкам подходит, однако, так же, как к мертвым, т.е. накапливают языковой материал, иначе говоря – записывают тексты, а потом их обрабатывают по принципам мертвых языков. Я утверждаю, что при этом получают мертвые словари и грамматики. Исследователь живых языков должен поступать иначе. Конечно, он тоже должен исходить из так или иначе понятого языкового материала. Но, построив из фактов этого материала некую отвлеченную систему, необходимо проверять ее на новых фактах, т.е. смотреть, отвечают ли выводимые из нее факты действительности. Таким образом, в языкознание вводится принцип эксперимента. Сделав какое-нибудь предположение о смысле того или иного слова, той или иной формы, о том или ином правиле словообразования или формообразования и т.п., следует пробовать, можно ли сказать ряд разнообразных фраз (который можно бесконечно множить), применяя это правило. Утвердительный результат подтверждает правильность постулата и, что любопытно, сопровождается чувством большого удовлетворения, если подвергшийся эксперименту сознательно участвует в нем» [Щерба 1931/1974: 38]. Другими словами, эксперимент по Щербе – это и есть вынесение суждений о приемлемости на основании собственной интуиции, то есть именно та методика, которую мы описывали выше.

Однако по сути данный метод является методом интроспекции²¹. Данная «экспериментальная» процедура не удовлетворяет ни одному из приводимых ниже обязательных признаков естественнонаучного эксперимента. В строгом смысле слова **экспериментом** называется проверка некоторой научной гипотезы каузального характера²² на основе применения определенных правил. Наиболее важными правилами экспериментирования являются контролируемость экспериментальных условий (об этом см. сноску 8 выше) и воспроизводимость результатов. В эксперименте исследователь манипулирует одним или несколькими факторами (которые принято называть независимыми переменными), определяя их воздействие на другой фактор (зависимую переменную). Проведя экспериментальное исследование с необходимым для получения статистически значимых результатов количеством испытуемых и стимульных выражений (в нашем случае предложений), он получает некоторые результаты, из которых впоследствии выводит надежные и валидные заключения. Надежность эксперимента – это воспроизводимость и стабильность его результатов в другом подобном исследовании, которое будет проведено тем же самым или иным исследователем. Валидность эксперимента – это степень достоверности выводов и заключений, сделанных на основе этого эксперимента. Традиционно экспериментальное исследование (в частности, в области психолингвистики) имеет высокую внутреннюю валидность, непосредственно связанную с контролируемостью условий эксперимента, и низкую экологическую валидность, так как часто те результаты, которые мы получаем в лабораторных условиях, плохо экстраполируются на реальное языковое поведение людей в реальной жизни.

Мы видим, что метод вынесения суждений о приемлемости совсем мало похож на строгую экспериментальную процедуру. Насколько ей соответствует метод извлечения суждений о приемлемости? Если мы проведем эксперимент по всем правилам парадигмы экспериментального синтаксиса, будет ли он соответствовать стандартам, принятым в экспериментальной психологии? Для ответа на этот вопрос нам сначала нужно определиться, можем мы привлекать к подобным исследованиям «тренированную интуицию» лингвистов или нам стоит ограничиться интуицией рядовых носителей языка.

Среди лингвистов генеративного направления до сих пор преобладает мнение, согласно которому максимально достоверные суждения о приемлемости дают сами лингвисты, иногда основываясь на собственной интуиции, иногда привлекая интуицию своих коллег [Newmeyer 1983: 50]. Однако что об этом говорят экспериментальные лингвисты? За пятидесятилетнюю историю генеративизма мы имеем не так много работ, в которых бы эти вопросы тестировались строго экспериментально. Первая значительная статья на эту тему была написана в 1973 г. Н. Спенсером, который взял 150 примеров из шести лингвистических статей того времени и сравнил «да/нет» – суждения 43 наивных носителей английского языка и 22 лингвистов [Spencer 1973]. Оказалось, что среди наивных носителей совпадало 80% суждений, а среди лингвистов – только 50%. В работе [Snow, Meijer 1977], однако, была получена высокая корреляция между ответами испытуемых лингвистов и нелингвистов; в работе [Gordon, Hendrick 1997], наоборот, были зафиксированы значимые различия. В недавней работе [Dąbrowska

²¹ Так как в этом разделе речь идет о психологической составляющей, необходимо специально оговорить использование терминов «интуиция» и «интроспекция». В психологической традиции, идущей от В. Вундта, считалось, что при помощи интроспекции (интуиции) человек может осуществлять прямой доступ к своей когнитивной системе. Современная парадигма генеративного синтаксиса исходит из того, что метод вынесения/извлечения суждений является поведенческим продуктом работы когнитивной системы, а никак не методом прямого доступа к ментальной грамматике. Таким образом, мы отождествляем метод интроспекции и вынесения/извлечения суждений о приемлемости не совсем корректно.

²² То есть гипотезы о причинно-следственной связи одного явления с другим. Если мы проверяем просто гипотезу о связи одного явления с другим, то, в соответствии с общепризнанной психологической практикой, подобное корреляционное исследование экспериментом в строгом смысле слова не является.

2010] автор сравнила суждения о приемлемости 38 наивных носителей английского языка и 38 лингвистов, дополнительно разделив вторую группу на генеративистов и функционалистов. Оказалось, что полученные значимые различия между лингвистами и нелингвистами могут быть объяснены тем, что лингвисты были более чувствительны к грамматической структуре предложения и менее чувствительны к его лексическому наполнению, в то время как наивные носители, наоборот, были более чувствительны к конкретной лексике, чем к грамматической структуре. Что касается различий между двумя группами лингвистов, то в целом функционалисты оказались ближе к наивным носителям языка, чем генеративисты. В результате Е. Домбровска делает общий вывод о нежелательности использования суждений лингвистов в синтаксических статьях.

Еще одним аргументом против привлечения лингвистов может служить тот факт, что металингвистические знания, которые испытуемый привлекает в ходе теста на извлечение суждений, могут повлиять на результаты эксперимента. Так, в медицине, биологии и психологии давно известен **эффект плацебо**, когда испытуемый констатирует улучшения в состоянии своего здоровья, принимая вместо лекарства вещество без лечебных свойств (обычно лактозу), только благодаря тому, что верит в эффективность данного препарата. Менее известен факт, что сами врачи чаще находят несуществующие улучшения в состоянии здоровья тех своих пациентов, которые принимают хорошие, на их взгляд, лекарства, по сравнению с ситуацией, когда врач не осведомлен в деталях о назначенном лечении [Noseworthy et al. 1994]. Подобные факты приводят медиков, биологов и психологов к необходимости использования процедуры **слепого** (англ. blind) исследования, когда испытуемый не знает, что именно интересует исследователя в данном конкретном исследовании, или даже **двойного слепого** (англ. double-blind) исследования, когда человек, проводящий исследование, не знаком с его целями. По замечанию Е. Домбровской, лингвист-синтаксист, полагающийся в своих исследованиях на свои собственные суждения о приемлемости, совмещает в одном лице и врача, и пациента описанного выше медицинского исследования: он не только описывает свое внутреннее состояние, но еще и интерпретирует его [Dąbrowska 2010]²³.

Но в не утихающей до сих пор дискуссии есть и противоположные мнения. Так, М. Девит полагает, наоборот, что лингвисты-синтаксисты являются самыми лучшими испытуемыми подобных тестов. Он рассуждает следующим образом: (1) любой испытуемый выносит суждения, руководствуясь своей внутренней теорией относительно родного языка; (2) лингвисты обладают лучшей (по крайней мере, более полной) теорией, чем нелингвисты; (3) таким образом, суждения лингвистов представляют собой «более достоверные индикаторы синтаксической реальности» (подробнее см. [Devitt 2006]). Однако с этим не согласны авторы статьи [Culbertson, Gross 2009], которые провели эксперимент с 42 испытуемыми, разделенными на четыре группы: (1) лингвисты, имеющие значительный опыт в области синтаксиса; (2) студенты лингвистических специальностей, прослушавшие курс синтаксиса; (3) студенты когнитивных наук без синтаксического бэкграунда; (4) люди, не имеющие знаний в области когнитивных наук. Они не обнаружили различий между разными группами испытуемых²⁴ и пришли к выводу, что гипотеза Девита не верна. Однако точка в данной дискуссии еще не постав-

²³ Похожая критика содержится в работе [Wasow, Arnold 2005], в которой интроспективные суждения лингвистов называются **первичной интуицией** лингвистов, а интерпретация полученных интроспективных суждений – **вторичной интуицией** лингвистов. Авторы работы выступают против использования вторичной интуиции, что совпадает с известным высказыванием Лабова о том, что лингвисты не могут одновременно порождать и языковые данные, и теории, см. выше.

²⁴ Вернее, авторы работы [Culbertson, Gross 2009] обнаружили отсутствие высокой корреляции в группе 4, то есть среди испытуемых с отсутствием общекогнитивных знаний. Однако после изучения неформальных отчетов испытуемых, проведенных после завершения эксперимента, они сделали осторожный вывод о том, что данные испытуемые не отличаются от других по степени владения синтаксическими структурами, – скорее, они не имеют навыков работы с подобными опросниками [Culbertson, Gross 2009: 732].

лена, о чем можно судить по продолжающейся полемике на эту тему (см. [Devitt 2010; Gross, Culbertson 2011]).

Некоторые исследователи считают, что более серьезное различие между суждениями лингвистов и нелингвистов может быть связано с различиями в их предшествующем опыте. Когда лингвист часто сталкивается с неграмматичными или пограничными примерами, они начинают казаться ему более приемлемыми. Это явление, получившее название **синтаксического пресыщения** (англ. syntactic satiation) или **лингвистического недуга** (англ. linguists' disease), было экспериментально исследовано в работах [Stromswold 1986; Hiramatsu 1999; Snyder 2000]. Авторы отмечают, что многие лингвисты признают, что вместо того, чтобы чувствовать неграмматичность некоторой часто цитируемой синтаксической конструкции, они просто помнят суждения, приводимые в лингвистической литературе (см., например, [Snyder 2000: 575]).

Таким образом, существует довольно много разнообразных доводов против использования суждений лингвистов в экспериментальной практике. Однако что произойдет, если мы откажемся от использования лингвистов в качестве испытуемых и будем тестировать только наивных носителей языка? Во-первых, мы все равно не сможем решить проблему влияния металингвистических знаний испытуемого – даже самый лингвистически неподготовленный испытуемый, зная цель эксперимента, не может от этого абстрагироваться. Другими словами, внутренняя валидность таких исследований оказывается весьма низкой²⁵ (подробнее см. [Bresnan 2007]). Во-вторых, некоторые исследователи упоминают в связи с этой проблемой так называемый «феномен умного Ганса»²⁶ [Wasow, Arnold 2005]. Получается, что и метод извлечения суждений о приемлемости не может считаться экспериментом в строгом смысле слова, сохраняя все черты субъективности, присущие методу вынесения суждений о приемлемости.

В заключительном разделе мы подведем общие итоги дискуссии и обозначим свою позицию относительно перспектив развития данного направления исследований.

6. ЗАКЛЮЧЕНИЕ: ОБЩИЕ ПРИНЦИПЫ РАБОТЫ С ЯЗЫКОВЫМ МАТЕРИАЛОМ

Подытоживая описанную выше дискуссию, мы сначала сформулируем вопросы, на которые надо так или иначе ответить, а затем опишем некоторые общие принципы работы с языковым материалом, которые представляются нам наиболее перспективными из всех вышеописанных. Разделим все вопросы на четыре блока:

(17) Проблема использования интуиции лингвиста:

- (i) можно ли полагаться на интуицию лингвистов, (ii) нужно ли целиком полагаться на интуицию лингвистов или (iii) суждения лингвистов должны быть раз и навсегда исключены?

(18) Проблема количества языковых данных:

- (i) можно ли использовать одиночные суждения носителей, (ii) нужно ли всегда заменять их масштабной процедурой в духе экспериментального синтаксиса или (iii) к этой процедуре стоит прибегать только в определенных ситуациях?

²⁵ Что касается экологической валидности извлечения суждений, то тут мнения исследователей в очередной раз расходятся: некоторые полагают, что человек в повседневной жизни довольно часто сталкивается с необходимостью оценивать предложения с точки зрения их правильности, – например, редактируя какой-либо документ или поправляя ошибки иностранца, изучающего данный язык; другие же считают, что такая деятельность не типична для повседневной жизни.

²⁶ Он состоит в том, что животные, которые как бы совершают трюки, требующие высоких умственных способностей, на самом деле могут реагировать на скрытые сигналы, бессознательно подаваемые человеком, проводящим исследование. Эффект, первоначально полученный при работе с животными, часто распространяют и на испытуемых-людей.

(19) Проблема статуса теста на вынесение/извлечение суждений:

(i) насколько тест на вынесение/извлечение суждений о приемлемости является экспериментом в строгом смысле слова, (ii) нужно ли дополнять его другими экспериментами?

(20) Проблема противоречивости данных:

как лучше выходить из ситуации противоречивых данных в тех случаях, когда не совпадают (i) суждения лингвистов и наивных носителей, (ii) суждения людей, знакомых и незнакомых с теоретической подоплекой исследования, или (iii) данные тестов на приемлемость и данные, полученные из других источников?

По нашему мнению, дальнейшего прогресса в этой области удастся достичь, руководствуясь следующими тремя принципами работы:

(21) Принцип консенсуса:

Если нет никаких оснований думать обратное, мы предполагаем, что одиночные суждения любого носителя (включая лингвиста) характеризуют всех говорящих на данном языке.

В предыдущих разделах мы показали, что в некоторых случаях суждения лингвистов отличаются от суждений наивных носителей языка, а суждения наивных носителей не являются надежными, то есть не воспроизводятся при проведении тестов с другими носителями. Однако такие рассогласования составляют мизерный процент от общего числа суждений и в любом случае должны, на наш взгляд, рассматриваться более пристально с использованием всего спектра доступных методов. Кроме того, до сих пор, насколько нам известно, не существует ни одного исследования, в котором было бы продемонстрировано, что различия в суждениях лингвистов и нелингвистов приводят к различиям в теоретических построениях.

(22) Принцип языковых данных:

Как тест на вынесение суждений о приемлемости, так и тест на извлечение суждений о приемлемости не является строгой экспериментальной процедурой. В случае рассогласования суждений исследователю стоит использовать как корпусные данные, так и строго экспериментальные.

Традиционно в лингвистике выделяется три метода сбора языкового материала – метод интроспекции, метод наблюдения в естественных условиях, составной частью которого является корпусный метод, а также метод эксперимента. Интроспекция, на которой основаны тесты на вынесение/извлечение суждений о приемлемости, не дает нам прямого доступа к когнитивной деятельности, поэтому мы не можем целиком полагаться только на нее. Наблюдение и эксперимент предоставляют дополнительные свидетельства, которые в отличие от интроспекции являются более объективными, хотя также не обеспечивают прямого доступа к ментальной грамматике.

(23) Принцип валидности:

Если в интроспективных суждениях имеются рассогласования, то валидное описание языка должно опираться на (i) суждения людей, незнакомых с теоретической подоплекой исследования; (ii) суждения лингвистов, а не наивных носителей языка; (iii) данные наблюдения и эксперимента, а не на интуицию.

Мы разделяем мнение Лабова и его сторонников о том, что знание теоретической подоплеки может оказывать влияние на выносимые лингвистом суждения, поэтому в сложных случаях исследователю стоит воздержаться от собственных оценок. В то же время именно в сложных случаях мнение других лингвистов может оказаться более ценным, чем мнение наивных носителей, а данные интроспекции следует дополнять данными наблюдения и эксперимента.

Таким образом, на наш взгляд, методология экспериментального синтаксиса не находит своей ниши – она неоправданно усложняет сбор примеров в тривиальных

случаях, которые составляют подавляющее большинство всего языкового материала, и мало помогает в сложных случаях, в которых нужно использовать подлинные экспериментальные техники, к числу которых методы экспериментального синтаксиса не относятся.

По нашему мнению, для достижения дальнейшего прогресса в этой области генеративным лингвистам не стоит расходовать силы на усложнение процедуры сбора языкового материала, а стоит сосредоточиться на задаче, которая до сих пор остается по сути нерешенной, а именно задаче вычленения «чистой» грамматичности из приемлемости. Именно в этой задаче, на наш взгляд, усилия лингвистов должны быть поддержаны усилиями психолингвистов, которым еще только предстоит найти и тщательным образом изучить все те дополнительные факторы, которые оказывают влияние на грамматические правила при использовании предложений в реальной речи.

СПИСОК ЛИТЕРАТУРЫ

- Слобин, Грин 1976 – Д. Слобин, Дж. Грин. Психолингвистика. М., 1976.
Слобин, Грин 2004 – Д. Слобин, Дж. Грин. Психолингвистика. М., 2004.
Соссюр 2004 – Ф. де Соссюр. Курс общей лингвистики. М., 2004.
Фрумкина 2009 – Р.М. Фрумкина. Психолингвистика как проект // Текст доклада XVI Симпозиума по психолингвистике и теории коммуникации «Психолингвистика в XXI веке: результаты, проблемы, перспективы». 2009. <http://polit.ru/article/2009/06/15/psyling/>
Хилл 1962 – А. Хилл. О грамматической отмеченности предложений // ВЯ. 1962. № 4.
Хомский 1962 – Н. Хомский. Синтаксические структуры // Новое в лингвистике. Вып. 2. М., 1962.
Щерба 1931 (1974) – Л.В. Щерба. О тройном аспекте языковых явлений и об эксперименте в языкознании // Л.В. Щерба. Языковая система и речевая деятельность. Л., 1974.
Adger 2003 – D. Adger. Core syntax: A minimalist approach. Oxford, 2003.
Bader, Häussler 2010 – M. Bader, J. Häussler. Toward a model of grammaticality judgments // Journal of linguistics. 2010. V. 46. № 2.
Baltin, Collins 2001 – M.R. Baltin, C. Collins. The handbook of contemporary syntactic theory. Blackwell, 2001.
Bard et al. 1996 – E.G. Bard, D. Robertson, A. Sorace. Magnitude estimation of linguistic acceptability // Language. 1996. V. 72.
Boersma, Hayes 2001 – P. Boersma, B. Hayes. Empirical tests of the gradual learning algorithm // Linguistic inquiry. 2001. V. 32.
Bolinger 1961 – D.L.M. Bolinger. Generality, gradience, and the all-or-none. The Hague, 1961.
Bornkessel-Schlesewsky, Schlewsky 2007 – I. Bornkessel-Schlesewsky, M. Schlewsky. The wolf in sheep's clothing: Against a new judgment-driven imperialism // Theoretical linguistics. 2007. V. 33. № 3.
Bresnan 2007 – J. Bresnan. Is syntactic knowledge probabilistic? Experiments with the English dative alternation // S. Featherston, W. Sternefeld (eds.). Linguistics in search of its evidential base. Berlin, 2007.
Bybee 2006 – J.L. Bybee. From usage to grammar: The mind's response to repetition // Language. 2006. V. 82. № 4.
Bybee, Eddington 2006 – J.L. Bybee, D. Eddington. A usage-based approach to Spanish verbs of becoming // Language. 2006. V. 82. № 2.
Chomsky 1957 – N. Chomsky. Syntactic structures. The Hague; Paris, 1957.
Chomsky 1964 – N. Chomsky. Degrees of grammaticality // J. Fodor, J. Katz (eds.). The structure of language: Readings in the philosophy of language. New York, 1964.
Chomsky 1965 – N. Chomsky. Aspects of the theory of syntax. Cambridge, 1965.
Chomsky 1968 – N. Chomsky. Language and mind. New York, 1968.
Chomsky 1986 – N. Chomsky. Knowledge of language: Its nature, origin, and use. New York, 1986.
Clark 2005 – B. Clark. On Stochastic grammar // Language. 2005. V. 81. № 1.
Cowart 1997 – W. Cowart. Experimental syntax: Applying objective methods to sentence judgments. London, 1997.
Culbertson, Gross 2009 – J. Culbertson, S. Gross. Are linguists better subjects? // The British journal for the philosophy of science. 2009. V. 60. № 4.

- Dąbrowska 2010 – *E. Dąbrowska*. Naive v. expert intuitions: An empirical study of acceptability judgments // *The linguistic review*. 2010. V. 27. № 1.
- Devitt 2006 – *M. Devitt*. Ignorance of language. Oxford, 2006.
- Devitt 2010 – *M. Devitt*. Linguistic intuitions revisited // *British journal for the philosophy of science*. 2010. V. 61.
- Dresher 1995 – *E. Dresher*. There's no reality like psychological reality // *Glott international*. 1995. V. 7.
- Edelman, Christiansen 2003 – *S. Edelman, M. Christiansen*. How seriously should we take Minimalist syntax? // *Trends in cognitive sciences*. 2003. V. 7.
- Fanselow 2007 – *G. Fanselow*. Carrots perfect as vegetables, but please not as a main dish // *Theoretical linguistics*. 2007. V. 33. № 3.
- Featherston 2002 – *S. Featherston*. Coreferential objects in German: Experimental evidence on reflexivity // *Linguistische Berichte*. 2002. V. 192.
- Featherston 2005 – *S. Featherston*. Universals and grammaticality: Wh-constraints in German and English // *Linguistics*. 2005. V. 43.
- Featherston 2007 – *S. Featherston*. Data in generative grammar: The stick and the carrot // *Theoretical linguistics*. 2007. V. 33. № 3.
- Featherston 2009 – *S. Featherston*. A scale for measuring well-formedness: Why syntax needs boiling and freezing points // *S. Featherston, S. Winkler (eds.). The fruits of empirical linguistics: Process*. Berlin; New York, 2009.
- Ferreira 2005 – *F. Ferreira*. Psycholinguistics, formal grammars, and cognitive science // *The linguistic review*. 2005. V. 22.
- Gibson, Fedorenko 2010 – *E. Gibson, E. Fedorenko*. Weak quantitative standards in linguistics research // *Trends in cognitive sciences*. 2010. V. 14.
- Gibson, Fedorenko in press – *E. Gibson, E. Fedorenko*. The need for quantitative methods in syntax and semantics research // *Language and cognitive processes*. In press.
- Gordon, Hendrick 1997 – *P. Gordon, R. Hendrick*. Intuitive knowledge of linguistic co-reference // *Cognition*. 1997. V. 62. № 3.
- Greene 1972 – *J. Greene*. Psycholinguistics: Chomsky and psychology. London, 1972.
- Grewendorf 1988 – *G. Grewendorf*. Aspekte der Deutschen Syntax. Eine Rektions-Bindungs-Analyse. Tübingen, 1988.
- Gross, Culbertson 2011 – *S. Gross, J. Culbertson*. Revisited linguistic intuitions // *British journal for the philosophy of science*. 2011. V. 62.
- Hill 1961 – *A.A. Hill*. Grammaticality // *Word*. 1961. V. 17.
- Hiramatsu 1999 – *K. Hiramatsu*. What syntactic satiation can tell us about islands // *Papers from the regional meetings, Chicago linguistic society*. 1999. V. 35.
- Hopper 1987 – *P. Hopper*. Emergent grammar // *Berkeley linguistics society*. 1987. V. 13.
- Householder 1973 – *F.W. Householder*. On arguments from asterisks // *Foundations of language*. 1973. V. 10. № 3.
- Jakobson 1959/1971 – *R. Jakobson*. Boas' view of grammatical meaning // *R. Jakobson. Selected writings*. V. 2. The Hague; Paris, 1971.
- Keller 2000 – *F. Keller*. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Ph.D. thesis. Edinburgh, 2000.
- Labov 1972 – *W. Labov*. Where do grammars stop? // *R. Shuy (ed.). Georgetown monograph on languages and linguistics*. 1972. V. 25.
- Labov 1975 – *W. Labov*. What is a linguistic fact? (PdR Press publications in linguistic theory. V. 1.) Lisse, 1975.
- Labov 1996 – *W. Labov*. When intuitions fail // *L. McNair, K. Singer, L.M. Dobrin, M.M. and AuCoin (eds.). CLS 32: Papers from the parasession on theory and data in linguistics*. Chicago, 1996.
- Langacker 1987 – *R.W. Langacker*. Foundations of cognitive grammar. V. 1. Theoretical prerequisites. Stanford, 1987.
- Langacker 1988 – *R.W. Langacker*. A usage-based model // *B. Rudzka-Ostyn (ed.). Topics in cognitive linguistics*. Amsterdam, 1988.
- Likert 1932 – *R. Likert*. A technique for the measurement of attitudes // *Archives of psychology*. 1932. V. 140.
- Myers 2009 – *J. Myers*. The design and analysis of small-scale syntactic judgment experiments // *Lingua*. 2009. V. 119.
- Newmeyer 1983 – *F.J. Newmeyer*. Grammatical theory, its limits and its possibilities. Chicago, 1983.
- Newmeyer 2003 – *F.J. Newmeyer*. Grammar is grammar and usage is usage // *Language*. 2003. V. 79. № 4.

- Newmeyer 2007 – *F.J. Newmeyer*. Commentary on Sam Featherston «Data in generative grammar: The stick and the carrot» // *Theoretical linguistics*. 2007. V. 33. № 3.
- Noseworthy et al. 1994 – *J.H. Noseworthy, G.C. Ebers, M.K. Vandervoort, R.E. Farquhar, E. Yetisir, R. Roberts*. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial // *Neurology*. 1994. V. 44.
- Phillips 2009 – *C. Phillips*. Should we impeach armchair linguists? // *S. Iwasaki et al. (eds.). Japanese/Korean linguistics (CSLI Publications)*. 2009. V. 17.
- Roeper 1982 – *T. Roeper*. Review of Halle, Bresnan and Miller (eds.). *Linguistic theory and psychological reality* // *Language*. 1982. V. 52.
- Sampson 2007 – *G.R. Sampson*. Grammar without grammaticality // *Corpus linguistics and linguistic theory*. 2007. V. 3.
- Sapir 1933 – *E. Sapir*. La réalité psychologique des phonèmes // *Journal de psychologie normale et pathologique*. 1933. V. 30.
- Schütze 1996 – *C.T. Schütze*. *The empirical basis of linguistics*. Chicago, 1996.
- Schütze 2011 – *C.T. Schütze*. Linguistic evidence and grammatical theory // *Cognitive science*. 2011. V. 2. № 2.
- Slobin 1971 – *D.I. Slobin*. *Psycholinguistics*. Glenview, 1971.
- Snow, Meijer 1977 – *C. Snow, G. Meijer*. On the secondary nature of syntactic intuitions // *S. Greenbaum (ed.). Acceptability in language*. The Hague, 1977.
- Snyder 2000 – *W. Snyder*. An experimental investigation of syntactic satiation effects // *Linguistic inquiry*. 2000. V. 31.
- Spencer 1973 – *N.J. Spencer*. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability // *Journal of psycholinguistic research*. 1973. V. 2. № 2.
- Sprouse 2011 – *J. Sprouse*. A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments // *Language*. 2011. V. 87. № 2.
- Sprouse, Almeida submitted – *J. Sprouse, D. Almeida*. A formal experimental investigation of the empirical foundation of generative syntactic theory. Submitted.
- Sprouse et al. submitted – *J. Sprouse, C.T. Schütze, D. Almeida*. Assessing the reliability of journal data in syntax: *Linguistic inquiry* 2001–2010. Submitted.
- Stevens 1956 – *S.S. Stevens*. The direct estimation of sensory magnitudes – loudness // *American journal of psychology*. 1956. V. 69.
- Stromswold 1986 – *K. Stromswold*. Syntactic satiation. Manuscript. MIT; Cambridge, 1986.
- Valian 1982 – *V. Valian*. Psycholinguistic experiment and linguistic intuition // *T.W. Simon, R.J. Scholes (eds.). Language, mind, and brain*. Hillsdale, 1982.
- Wasow 2009 – *T. Wasow*. Gradient data and gradient grammars // *Proceedings of the 43rd Annual meeting of the Chicago linguistics society*. 2009.
- Wasow, Arnold 2005 – *T. Wasow, J. Arnold*. Intuitions in linguistic argumentations // *Lingua*. 2005. V. 115. № 11.
- Weskott, Fanselow 2011 – *T. Weskott, G. Fanselow*. On the informativity of different measures of linguistic acceptability // *Language*. 2011. V. 87. № 2.

Сведения об авторе:

Ольга Викторовна Федорова
МГУ им. М.В. Ломоносова
olga.fedorova@msu.ru