

© 1991 г.

ШАЛЯПИНА З. М.

**ГРАММАТИКА И ЕЕ СООТНОШЕНИЕ СО СЛОВАРЕМ ПРИ
СЛОВОЦЕНТРИЧЕСКОМ ПОДХОДЕ К ЯЗЫКУ****(На опыте формализованного лингвистического описания)**

Проблема соотношения словаря и грамматики представляет собой одну из наиболее кардинальных языковедческих проблем, и от того, как она решается в рамках конкретных лингвистических описаний, существенно зависят их состав и строение. Особенно важное значение это имеет для формализованных описаний языка, составляемых в прикладных целях, поскольку к такому описанию предъявляются более высокие, чем в иных случаях, требования последовательности, непротиворечивости и согласованности между собой всех его компонентов, как лексических, так и грамматических, и необходимые предпосылки реализации этих требований должны быть в явном виде заложены уже в самом способе организации соответствующего описания и в структуре используемых при этом формальных средств. Предметом настоящей статьи является теоретическое осмысление и обоснование, а также рассмотрение некоторых прикладных и общезыковедческих импликаций одного из возможных здесь подходов. В основных своих чертах этот подход сформировался в связи с задачей создания системы англо-русского автоматического перевода (АРАП), разрабатывавшейся в МГПИИЯ им. М. Тореза в 1969—1977 гг. (см. [1—7]). В настоящее время он развивается применительно к лингвистическому обеспечению системы японско-русского автоматического перевода (ЯРАП), работа над которой начата в Институте востоковедения АН СССР [8—10].

Обе названные системы базируются на том общем принципе (часто называемом лексикографическим), что центральным компонентом их лингвистического обеспечения должен быть словарь и ему должна принадлежать определяющая роль по отношению ко всем прочим частям этого обеспечения, включая грамматику.

Принцип этот, в теоретическом плане тесно связанный с так называемым словоцентрическим [11] направлением в языкознании, активно применялся в прикладной лингвистике по существу с самого начала разработок в области автоматического перевода. Однако вначале он нередко получал неоправданно упрощенную интерпретацию, искажающую его лингвистический смысл, и сводился на практике к заведомо безуспешным попыткам отказаться при обработке текстов от анализа их синтаксической структуры, ограничившись выявлением фигурирующих в них лексических единиц [12]. Бесперспективность этого пути привела к появлению иной трактовки рассматриваемого принципа, наиболее четко представленной в модели «Смысл ↔ Текст»: «необходим словарь, в систематической форме задающий закономерности употребления всех содержащихся в нем слов» [13, с. 31], т. е. описывающий лексику языка в неразрывной связи с ее сочетаемостными потенциями. На такое понимание лексикогра-

фического принципа мы и опирались при решении вопросов организации лингвистического обеспечения систем АРАП и ЯРАП.

Ориентация на указанный принцип поставила перед нами задачу пересмотреть тот взгляд на соотношение словаря и грамматики в формализованном лингвистическом описании, который был широко распространен (во многом, видимо, под влиянием школы Н. Хомского) в структурной лингвистике 60 — нач. 70-х годов и который эксплицитно выражен, например, в одной из работ Е. В. Падучевой: «...словарь должен служить дополнением грамматики до исчерпывающего описания соотношения „смысл ↔ текст“, т. е. задавать для каждого слова значения всех признаков, упоминаемых в грамматике» [14]. Легко видеть, что данная точка зрения вступает в противоречие с тезисом о центральной роли словаря: согласно цитируемому утверждению, состав и структура словарных описаний принципиально зависят от того, как (с помощью какого набора признаков) описана грамматика соответствующего языка, и, стало быть, именно грамматика, а не словарь оказывается центральным, определяющим компонентом всего описания. Чтобы избежать этого противоречия, необходимо было найти какой-то иной способ согласования словаря и грамматики, при котором не требовалось бы в качестве предварительного условия для составления словаря иметь полный перечень признаков, возможных в грамматике данного языка.

В поисках такого способа было естественно обратиться к опыту, накопленному в интересующей нас области теоретической лингвистики.

В общем языкознании выделен целый ряд критериев противопоставления в лингвистическом описании грамматического и неграмматического (лексического): семантический критерий, предполагающий различие в характере значений грамматических и лексических единиц, соответственно; критерий ограниченности/неограниченности инвентаря тех и других единиц; критерий обязательности/необязательности их употребления и т. п. (см., например [15, с. 200—210]). Из них для разграничения и согласования между собой словаря и грамматики в системе АРАП (и позже ЯРАП) был выбран критерий общности — наиболее традиционный и, на наш взгляд, наиболее универсальный¹.

Суть этого критерия сводится, как известно, к тому, что в отличие от лексических описаний, фиксирующих индивидуальные особенности отдельных единиц словаря, грамматика носит обобщенный характер, описывая свойства, относящиеся к целым классам однородных единиц. Впрочем, строго говоря, лексика также есть некоторое обобщение, именно — обобщение множества словоформ данного языка, для многих языков являющееся достаточно нетривиальным. Поэтому более корректным будет здесь, по-видимому, утверждение не о бинарном противопоставлении индивидуального и общего, но о различии в степени достигаемого обобщения: по сравнению со словарем, грамматика основана на обобщении существенно более высокого уровня, которое во многих случаях можно считать вторичным по отношению к словарному обобщению. В той мере, в какой имеет место такая вторичность, можно говорить, что грамматика есть обобщение словаря.

¹ В противоположность авторам [15, с. 209], рассматривающим общий и регулярный характер грамматических правил как следствие свойства обязательности выражения грамматических значений, мы полагаем, что само это свойство возникает в результате объединения в один грамматический класс тех и только тех единиц, которые обладают общим для всех них признаком регулярного выражения при них определенного значения.

В рамках формализованного лингвистического описания критерий общности легко интерпретируется в операционных терминах: возможность обобщения каких-то элементов такого описания должна проявляться, очевидно, в повторяемости соответствующих языковых фактов в языке и речи (тексте), и простейшим способом этого обобщения оказывается построение грамматики путем выделения и, так сказать, «вынесения за скобки» при описании индивидуальных языковых единиц всей той информации, которая достаточно часто повторяется в отношении различных единиц и/или их разных текстовых употреблений.

Если требуемую степень повторяемости обнаруживает некоторая конкретная единица как таковая, в грамматику выносятся сама эта конкретная единица вместе с описанием всех ее свойств (ср. традиционно включаемые в грамматику служебные слова, отличающиеся именно той особенностью, что частотность их употребления в текстах всегда существенно выше частотности в тех же текстах знаменательных слов, независимо от стиля или тематики самих текстов). Если же повторяются отдельные структурные свойства различных единиц: внутренние (= касающиеся набора образующих ту или иную единицу составных компонентов и их организации в ее пределах) или внешние (характеризующие место и роль данной единицы в построении более крупных единиц текста), — сведения о таких совпадающих для группы единиц свойствах в описании каждой из них заменяются некоторым условным символом, выполняющим две взаимосвязанные функции: по отношению к заменяемому им лингвистическим сведениям он выступает как их сокращенное обозначение; по отношению к тем единицам, в описаниях которых такие сведения выявлены, — как имя грамматического класса, объединяющего эти единицы по признаку общности для них данной группы сведений. В грамматике каждому подобному символу сопоставляется при таком подходе собственное описание, содержащее рассматриваемые сведения в полном виде, т. е. дается, так сказать, расшифровка введенного для них «сокращенного обозначения» [3, с. 10].

В обоих случаях основным способом лингвистического обобщения и формирования на его основе грамматических описаний, или, как они названы в [3] (где впервые эксплицитно излагается и иллюстрируется данный подход к грамматике), а затем в [4], «грамматических словарных статей», является, как видим, операция теоретико-множественного пересечения в своей обычной математической интерпретации. Это создает, очевидно, достаточно серьезные предпосылки для формализации и в перспективе — автоматизации на этой базе самих соответствующих процессов.

Необходимо оговорить, однако, что применение названной теоретико-множественной операции как средства грамматического обобщения характеризуется одним важным ограничением, не вытекающим из ее математического определения. Именно: пределы применения здесь данной операции диктуются не только самим фактом наличия/отсутствия в описываемых текстах или в информации к имеющимся в них единицах тех или иных повторов, но и соображениями оптимальности результирующего лингвистического описания в целом, т. е. требованиями, во-первых, его максимальной компактности и, во-вторых, максимальной простоты его применения при обработке текстов.

Выполнение первого требования — обеспечение максимальной компактности составляемого лингвистического описания при сохранении полноты охвата в нем описываемого языкового материала — сопряжено,

с одной стороны, с возможно более полным устранением в этом описании случаев дублирования одной и той же информации и, с другой стороны, с минимизацией того исходного инвентаря единиц, на который оно опирается. Изложенный выше метод грамматического обобщения посредством операции теоретико-множественного пересечения служит для решения только первой из названных двух задач — устранения дублирования информации. Напротив, инвентарь исходных единиц описания при этом, вообще говоря, увеличивается, так как, кроме конкретных единиц, в него включаются также имена грамматических классов. Для минимизации этого инвентаря требуется, по-видимому, учет взаимосвязей, имеющих между разными элементами обобщаемой информации, с тем чтобы во всех случаях, когда релевантность для некоторой языковой единицы одного из взаимосвязанных в этом смысле элементов информации влечет за собой релевантность для нее также другого из них и наоборот, в лингвистическом описании оба эти элемента соотносились, по возможности, с одной и той же единицей исходного инвентаря.

Важнейшим типом взаимосвязей такого рода является нередко наблюдаемая корреляция между наличием у рассматриваемых единиц общих внешних (сочетаемостных) свойств и присутствием общего компонента в определенной позиции их внутренней структуры. В этом случае для выделения общих свойств таких единиц можно не вводить специальный объединяющий их грамматический класс, а задавать все их общие свойства в описании, сопоставленном общему компоненту их структуры, наряду с собственными свойствами последнего, отметив только, что указанные сведения относятся не к самой той единице, в описании которой они фигурируют, а к единицам более крупным (относящимся к более «высокому» уровню синтагматического членения текста), в состав которых она может входить в качестве их структурного компонента. Единицы, свойства которых полностью представимы при этом через сведения об их структурных компонентах, можно, очевидно, исключить из исходного инвентаря, оставив в нем только эти компоненты. В той мере, в какой лингвистическое описание может быть сведено этим способом к описанию отдельных слов и их классов, и можно, по-видимому, говорить о «словоцентрическом» подходе к языку в полном смысле данного термина.

В формализованном лингвистическом описании изложенный метод позволяет, в частности, строить исчисление всех не идиоматических единиц, больших, чем отдельное слово, — словосочетаний, предложений, сверхфразовых единств, теоретически даже целых текстов — в виде описания словоформ, допустимых в их составе. Описание же словоформ может, в свою очередь, представляться как описание их морфологических компонентов: лексических корней и основ, образующих словарь такого описания, и служебных элементов (прежде всего словоизменятельных и стандартных словообразовательных показателей), которые, будучи значительно более частотными, должны, по нашим принципам, включаться в грамматику.

Понятно, что во всех случаях такое покомпонентное описание возможно, как уже оговорено выше, только при условии неидиоматичности описываемых единиц: если какая-то сложная единица — будь то сочетание лексического корня со словообразовательным аффиксом или целое предложение — имеет свойства, не выводимые из свойств составляющих ее единиц, она должна, естественно, включаться в словарь целиком. Однако указывать в ее словарном описании достаточно при этом лишь тот минимум ее свойств, который, действительно, идиоматичен; все остальные свой-

ства такой единицы могут быть опять-таки перенесены в описание, сопоставляемые ее компонентам.

Основным средством характеристики внешних свойств единиц всех отражаемых в таком описании языковых уровней: уровня морфологии слова, уровня синтаксиса предложения, уровня синтаксиса сверхфразовых единств и т. п. — становятся в нем в а л е н т н о с т и соответствующих уровней — *гесп.* морфологические, синтаксические, сверхфразовые и пр., — приписываемые непосредственно отдельным морфологическим элементам — как лексическим, так и грамматическим, — а также другим нечленным (хотя бы в отношении той или иной части их свойств) языковым единицам (таким, как, например, фразеологические сочетания).

Скажем, английской граммеме *pass(ive)* — показателю страдательного залога — на морфологическом уровне должна быть приписана при этом валентность на основу переходного глагола (переходного в широком смысле — т. е. допускающего трансформацию пассивизации), на синтаксическом же — валентности на субъект (агенса) и объект (пациенса) действия, задаваемого этой глагольной основой, с уточнением, что первая из них (субъектная, или агентивная) при ее непосредственной реализации в тексте должна оформляться в нем в виде косвенного дополнения с предлогом *by*, а вторая (объектная, пациентивная) — в виде подлежащего той словоформы, в состав морфологического представления которой входит данная граммема.

Другой пример: лексеме *similar* («аналогичный, подобный») будет приписано в рамках рассматриваемого подхода три валентности синтаксического уровня: валентность на объект, применительно к которому проводится аналогия с другим объектом (валентность реализуется определяемым существительным или — при предикативном употреблении описываемого прилагательного — подлежащим); валентность на объект, с которым проводится требуемая аналогия (реализуется на синтаксическом уровне дополнением с предлогом *to*), и валентность на признак, служащий основанием для проведения данной аналогии (реализуется дополнением с предлогом *in*). Кроме них эта лексема получит и ряд валентностей сверхфразового уровня, в частности, валентность, имеющую ту же содержательную интерпретацию, что и вторая из трех названных синтаксических (внутрифразовых) валентностей, но выражаемая рематическим компонентом предложения, предшествующего предложению со словом *similar* (ср.: *We use a magnetic amplifier essentially similar in operation to that used by our colleagues.* vs.: *Our colleagues use the F-1 type of magnetic amplifier. We use an essentially similar amplifier type.*), а также валентность на конкретизацию проводимой аналогии (реализуемую предложением/предложениями, следующим(и) за предложением со словом *similar* и уточняющим(и), в чем именно аналогичны сопоставляемые объекты и/или каковы границы проводимой аналогии, т. е. в чем эти объекты, несмотря на их сходство в других отношениях, различны) и т. д. ²

Помимо самих валентностей и возможных способов их реализации, в словарных статьях описываемых единиц должны указываться и все прочие особенности синтагматических конструкций, в которые они могут входить, — в той части, естественно, в какой соответствующие особенности обусловлены свойствами именно этих единиц. Такие особенности практически всегда могут быть определены в терминах реализуемых в той или

² Подробнее о понятии сверхфразовых валентностей и условиях их релевантности для структуры конкретного текста см. в [16].

ной конструкции валентностей ее компонентов. Это могут быть, например, ограничения на совместную реализацию разных валентностей самой описываемой единицы; ограничения на реализацию каких-то из ее валентностей одновременно с теми или иными валентностями других единиц, связанных с ней в рассматриваемой конструкции; требования определенных модификаций в оформлении валентностей этих других единиц; условия эллипсиса единиц, реализующих валентности тех или иных компонентов конструкции; закономерности, позволяющие восстанавливать эллиптированные единицы при ее обработке, и т. д.³

Например, в статье японского союза *кацу* («и, а также») будут не только указаны две его валентности и способы реализации каждой из них (одна его валентность реализуется при используемом представлении синтаксической структуры единицей, подчиняющей себе данный союз и стоящей непосредственно слева от него; если это глагол, он должен иметь при этом так называемую срединную форму; вторая валентность реализуется единицей, подчиненной союзу *кацу*, стоящей от него справа и, возможно, отделенной от него какими-то другими единицами), но будут также даны и правила согласования единиц, реализующих указанные две валентности в тексте, по части речи и по семантическому классу.

В статье английской граммы *part(iciples)* — показателя причастия — будет, в частности, даваться указание о том, что одна из валентностей, вводимых в статьи других компонентов той же глагольной словоформы, — та, для которой в соответствующей статье (скажем, в статье граммы *pass*) предусматривается ее реализация в виде подлежащего данного глагола, — в присутствии описываемой граммы *part* не может реализоваться в указанном виде, но получает те возможности реализации, которые характерны для первой валентности прилагательного (в виде определяемого данной словоформой существительного или в виде подлежащего при подчиняющем ее глаголе-связке).

Для слов, образующих конструкции с предикатными актантами, должна даваться, помимо прочего, информация о том, как соотносятся в таких конструкциях валентности главного (описываемого) и подчиненного предикатов — т. е., в частности, каковы условия и возможности кореферентности и эллипсиса реализующих эти валентности единиц в тех случаях, когда это зависит от ограничений, налагаемых самой конструкцией⁴. Такая информация релевантна, например, для английских глаголов типа *consider, think, find, want, make, see, feel* и т. п., имеющих то общее свойство, что в качестве одного из их актантов может выступать прилагательное или глагол [ср.: *We consider their amplifier (to be) essentially similar to ours; She was found guilty by the jury; I want him to specialize in electronics; They saw us enter(ing) the building* и т. д.]. Данное свойство может служить основанием для объединения их в общий лексико-грамматический класс, которому будет сопоставлена словарная статья, содержащая указание о том, что единица, реализующая первую валентность прилагательного (resp. «подлежащую» валентность глагола), подчиненного описываемому в этой статье слову (= члену описываемого в ней класса), всегда кореферентна его объекту (= пациенсу; синтаксически он может выражаться и подлежащим — см. второй из приведенных здесь примеров) и в синтаксической структуре эллиптируется.

³ Попытка исчисления основных типов синтагматических сведений подобного рода сделана в [2] в виде описания «общей структуры» словарных статей, используемых в системе АРАП.

⁴ О типах кореферентности такого рода см., например, в [17].

Другие типы сложных синтагматических сведений, помещаемых при нашем подходе в словарные статьи лексических и грамматических единиц, а также способы их формальной записи иллюстрируются, например, в работах [18—20].

Как видим, в рамках такого подхода для любой неидиоматичной единицы определенного языкового уровня, не являющейся внутренней элементарной (т. е. допускающей синтагматическое членение на более мелкие единицы), все основные аспекты ее сочетаемости на этом уровне могут быть адекватно представлены в виде суммы⁵ валентностей данного уровня, приписанных в лингвистическом описании элементарным компонентам рассматриваемой единицы, вкупе со сведениями о способах реализации и соотношении этих валентностей, также фиксируемых при соответствующих элементарных единицах. Тем самым, используя понятие валентности и дифференцируя валентности на типы в зависимости от того, сочетаемость единиц какого языкового уровня они характеризуют, можно обеспечить практическое исчерпывающее описание языка и его функционирования на всех таких уровнях. В этом смысле «словоцентрический» подход к языку при данной его интерпретации можно называть также «валентностным» подходом к лингвистическому описанию.

Другой тип взаимосвязей, учет которых имеет существенное значение для минимизации инвентаря исходных единиц лингвистического описания, — это взаимосвязи, объединяющие такие свойства языковых единиц, которые относятся либо только к их внешней сочетаемости, либо только к их внутренней структуре. Поскольку при отсутствии корреляций между сочетаемостью и структурой разных единиц обобщение их повторяющихся свойств может идти лишь путем формирования соответствующих грамматических классов, учет подобных взаимосвязей должен служить для минимизации инвентаря уже самих этих классов. Для этого сведения обо всем множестве свойств, которые предстают как взаимосвязанные в указанном выше смысле в пределах одной и той же группы единиц, следует включать в описание (статью) одного и того же грамматического класса, охватывающего всю эту группу единиц.

Здесь возникает, однако, та проблема, что при образовании того или иного класса языковых единиц и вынесении в озаглавленную его именем грамматическую статью всех общих для них сведений зачастую обнаруживается немало сходных с ними единиц, для которых большинство из этих сведений также справедливы, но отдельные из них (чаще всего разные для разных единиц) — ложны⁶ (скажем, ряд слов, по большинству своих свойств примыкающих к глаголам, может не допускать некоторых глагольных категорий, другие могут не сочетаться с обычным типом подлежащего и т. д.).

Формировать для описания общих свойств таких единиц самостоятельные грамматические классы, увеличивая тем самым их инвентарь, в свете требований компактности не всегда может быть целесообразно. Не вполне удовлетворителен при словоцентрическом подходе и традиционный спо-

⁵ Слово «сумма» употреблено здесь не совсем в том смысле, который оно имеет в математике: как было показано выше (см. пример с граммемой причастия), при формировании требуемой «суммы» некоторые из ее «слагаемых» могут в отдельных случаях модифицироваться под влиянием других, так что процедура «суммирования» оказывается осложненной соответствующими модифицирующими операциями. Однако чаще всего речь идет действительно о сумме — как теоретико-множественном объединении исходных лингвистических сведений.

⁶ Ю. А. Шрейдер говорит в этой связи о «почти-правилах», которые верны «почти для всех» единиц определенного класса и «почти верны» для всех этих единиц [21].

соб перечисления в грамматическом описании «исключений из правил», так как это противоречит основному принципу данного подхода («от словаря к грамматике», а не наоборот). С точки зрения этого принципа, более последовательно не в грамматике указывать, какие конкретные единицы не подчиняются данному общему правилу, а в словаре отмечать, какие элементы информации из релевантных для данной единицы грамматических статей к ней не относятся (или относятся в модифицированном виде).

Снажем, сведения о том, что в конструкции, где русское существительное сочетается с числительным *два, три* или *четыре*, стоящим в именительном падеже, для определяющих это существительное прилагательных не действует стандартное правило согласования с определяемым существительным по числу (и теряет свою обязательность правило согласования по падежу: падеж прилагательного может быть здесь как родительным, так и именительным), имеет смысл включать не в статью грамматического класса «прилагательное», которая будет содержать только стандартные согласовательные правила, но либо в статьи данных конкретных числительных, либо в статью очень частного лексико-грамматического класса, объединяющего ровно эти три числительных.

В формальном отношении возможность такого размещения информации об «исключениях из правил» — как и отмечавшаяся выше возможность модификации сведений, приписанных одному из компонентов некоторой элементарной единицы, при наличии соответствующих указаний в статье другого ее компонента (см. выше о свойствах граммы *part*), — это прежде всего возможность включения в словарные статьи, помимо прочих типов сведений, также требований об «отмене» тех или иных правил, содержащихся в других статьях. Чтобы обеспечить такую возможность, достаточно ввести в формализованный язык, используемый для записи лингвистической информации, специальный метасимвол (в последней версии лингвистического обеспечения системы АРАП в этой роли выступает символ *СНЯТЬ*), требующий игнорировать правила, отвечающие указанному при данном метасимволе образцу, при их обнаружении среди информации, релевантной для той единицы, в описании которой этот метасимвол фигурирует.

Второе из двух выделенных нами требований оптимальности формализованного лингвистического описания — требование максимальной простоты обращения с ним при обработке с его помощью конкретных текстов — применительно к интересующему нас соотношению словаря и грамматики подразумевает прежде всего простоту их соотнесения между собой, согласования заданной в них информации при релевантности ее для одной и той же реальной текстовой единицы. Излагаемый подход к грамматике как к набору лингвистических сведений, отличающихся от словарных только большей степенью повторяемости, позволяет обеспечить это следующим образом.

Любое правило, будь оно словарным или грамматическим, при рассматриваемом подходе можно формулировать так, как если бы оно относилось к одной конкретной словоформе — той, которая при обработке текста окажется содержащей данный лексический, словоизменительный и т. п. элемент или получит признак данного грамматического класса. Соответственно, как словарные, так и грамматические статьи могут строиться при этом по единому образцу, опираясь на одну и ту же символику (в прикладных описаниях — на один и тот же формальный язык и один и тот формат правил), а также на одну и ту же схему структурирования лингвистической информации (определяемую требуемым способом ис-

пользования этой информации при обработке текстов). В результате вся грамматика может быть представлена в виде своего рода «грамматических словарей»: словаря словоизменительных и словообразовательных показателей (описываемых вполне аналогично лексическим единицам); словарей грамматических классов разных типов и т. д.⁷

Таким образом, при нашем подходе описание лексики и грамматики оказывается полностью единообразным, и соотнесение разных лексических и грамматических статей, релевантных для той или иной текстовой словоформы, в стандартных случаях может быть сведено просто к теоретико-множественному объединению имеющейся в них информации в рамках отдельных зон и разделов общей для всех типов статей структурной схемы. Применительно к прикладным задачам это означает, среди прочего, что для процедуры и результатов функционирования системы, опирающейся на данный способ организации лингвистического описания, вообще говоря, безразлично, из словаря или из грамматики извлечена та или иная конкретная информация: для обработки этой информации в обоих случаях смогут использоваться одни и те же управляющие и интерпретирующие алгоритмы. Тем самым создается также принципиальная возможность автоматической оптимизации соотношения лексического и грамматического компонентов лингвистического описания непосредственно в ходе работы системы (например, по мере накопления количественных данных об обрабатываемых текстах) путем перераспределения имеющейся информации между двумя компонентами описания.

Несколько более сложным будет при описываемом подходе соотнесение между собой словарной и грамматической информации для тех единиц, для которых в силу их лексических особенностей действительно не все сведения из релевантных для них грамматических статей (см. выше). В подобных случаях требование компактности лингвистического описания и требование простоты работы с ним вступают между собой в конфликт, оптимальный способ разрешения которого во многом зависит от технических условий реализации всей системы и в этом отношении входит скорее в компетенцию математиков и программистов, а не лингвистов.

Итак, с позиций словоцентрического подхода к лингвистическому описанию не словарь дополняет грамматику, задавая значения грамматических признаков, а грамматика содержит в себе расшифровку тех сокращенных обозначений, которые вводятся в словарь во избежание дублирования информации. Как мы показали, такая смена угла зрения дает ряд преимуществ прикладного плана. Назовем в заключение еще некоторые следствия из этого подхода, имеющие, на наш взгляд, также определенное общеязыковедческое значение.

Прежде всего следует отметить, что рассматриваемый подход расширяет сферу применимости и словаря, и грамматики: в словаре при этом вполне допустимы, например, синтаксические правила, если они специфичны только для отдельных лексем; в грамматику естественным образом включается, среди прочих, семантический отдел, описывающий общие свойства лексико-семантических классов слов, и т. д.

⁷ В таком формате было последовательно представлено уже лингвистическое обеспечение системы АРАП [2—6]. В настоящее время он принят также в ряде других прикладных систем, разрабатываемых в СССР. В области «человечно-ориентированных» лингвистических описаний аналогичный подход предложен для описания русского словообразования В. В. Лопатиным и И. С. Улухановым в докладе «Мотивированное слово в описательной грамматике и в словаре служебных морфем», сделанном на Всесоюзной конференции «Слово в грамматике» (Звенигород, 1981 г.).

Обогащается и методическая основа описания лексики и грамматики, поскольку любое лингвистическое средство, выработанное в какой-либо из данных двух сфер, применимо при этом и к другой сфере. Например, понятия толкования, модели управления и т. п., возникшие первоначально как инструмент описания лексики, естественным образом переносятся и на описание морфолого-грамматических категорий. Выше мы уже показывали на примере английского пассива, что в описания залоговых морфем, в частности, удобно непосредственно включать синтаксическую модель управления для первых двух глагольных валентностей — т. е. сведения о допустимых для них способах их поверхностного оформления при реализации в тексте. В статье компаратива, как это демонстрировалось в [5], должна указываться, помимо синтаксической, также специфическая для компаративных форм прилагательных семантическая модель управления (определяемая набором переменных в толковании компаратива как такового). При этом в подобном описании может использоваться классификация языковых единиц как по частям речи, так и по членам предложения (такого рода классы, т. е. понятия типа *подлежащее, дополнение, определение* и т. п., уже фигурировали в приводившихся выше примерах описания способов реализации в тексте тех или иных глагольных синтаксических валентностей, описываемых в статьях, сопоставленных залоговым морфемам), — что позволяет шире привлекать в целях формализованного лингвистического описания результаты, сформулированные в терминах «человеко-ориентированной» лингвистики.

Повышается при излагаемом подходе также гибкость грамматического описания, так как рассмотренный метод грамматической классификации не предполагает таких ограничений, как, скажем, требуемая некоторыми авторами обязательная иерархичность грамматических классов или их обязательная противопоставленность по общему признаку. Так, в области семантической грамматики классификация слов может вестись не только по известному родо-видовому принципу, но и, например, по сходству семантических интерпретаций их валентностей.

Если первый способ классификации удобен преимущественно для описания селекционных ограничений и правил стандартной полисемии (таких, как приобретение смысла «единицы измерения» словами, обозначающими вместилище или замкнутое пространство, в контекстах типа *стаканом чаю, воз сена, мы видели целую поляну таких цветов* и т. п.), то сочетание его с валентностно-семантической классификацией позволяет задавать также общие правила восстановления эллипсиса, стандартные синонимические трансформации и т. п., в том числе и для таких случаев, которые, казалось бы, с трудом поддаются обобщению и формализации.

Так, этими средствами можно описать разницу в семантической обработке (в частности, при переводе) японских фраз:

(1) <i>Кэганин</i>	<i>ва</i>	<i>тэитэ</i>	<i>га</i>	<i>окурэ́тэ сима́тта</i>
Раненый	показатель темы или им. падежа	(медицинская) помощь	показатель ремы и/или им. падежа	опоздал

⇒ Раненому опоздали оказать помощь, но не *Раненый опоздал с помощью (как это предлагается в словаре [22]).

(2) <i>Гакусэй</i>	<i>ва</i>	<i>сакубун</i>	<i>но</i>	<i>тэйсюцу</i>	<i>га</i>	<i>окурэ́тта</i>
Студент	показатель темы или им. падежа	сочинение	род. падеж	подача, предложение явление	показатель ремы и/или им. падежа	опоздал

⇒ Студент опоздал подать сочинение (но не *Студенту опоздали подать сочинение).

Здесь применимы, например, следующие правила (в рамках настоящего изложения они приводятся, естественно, в неформальном виде):

1. В японской конструкции X *wa* Y *ga* Z , где X — непредикатная именная группа, X — именная группа с предикатным значением, Z — группа глагола, единица X семантически заполняет ту из валентностей предиката Y , семантическая интерпретация которой наиболее полно согласуется с толкованием единицы X . При отсутствии оснований для предпочтения какой-либо из валентностей предиката Y считается, что X заполняет его субъектную валентность.

2. Семантическая интерпретация валентности W_i согласуется с толкованием единицы X более полно, чем семантическая интерпретация валентности W_j , если при сопоставлении лексико-семантических классов, включающих единицу X , с одной стороны, и валентностно-семантических классов, включающих валентности W_i и W_j , с другой стороны, с учетом степени конкретности/абстрактности сопоставляемых классов, обнаруживается, что для валентности W_i совпадение соответствующих классов имеет место на более конкретном уровне, чем для валентности W_j .

В примере (1) $X =$ *кэганин*, $Y =$ *тэатэ*. Лексему *кэганин* «раненый» можно истолковать, например, как «человек, находящийся в таком состоянии, особенности которого препятствуют его нормальной жизнедеятельности (и нуждаются в устранении)»; лексему *тэатэ* «(медицинская) помощь» — как «действия человека W_1 (субъекта), имеющие целью устранить или уменьшить те особенности состояния человека W_2 (объекта), которые препятствуют его нормальной жизнедеятельности».

Если эти толкования представить в терминах формализованного семантического языка, предложенного в [7], то по ним с помощью имеющихся в [7] формальных правил можно построить описания интенционалов всех семантических классов, в которые могут, при данных толкованиях, входить толкуемые единицы и/или их валентности. В частности, лексема *кэганин* входит при этом в лексико-семантические классы, неформально описываемые как (а) «человек», (б) «человек, особенности состояния которого препятствуют его нормальной жизнедеятельности и должны быть устранены (= нуждаются в устранении)» и др. Валентность W_1 лексемы *тэатэ* входит в валентностно-семантические классы (а') «человек» и (в) «человек, осуществляющий действия, имеющие целью устранить...»; валентность W_2 — в классы (а') и (б') «человек, особенности состояния которого препятствуют его нормальной жизнедеятельности и должны быть устранены (или уменьшены, согласно целям действий человека W_1)».

Легко видеть, что для валентности W_1 совпадение включающих ее валентностно-семантических классов с лексико-семантическими классами, в которые входит лексема *кэганин*, наблюдается лишь на уровне наиболее общих из имеющихся классов — (а) и (а'). Что же касается валентности W_2 , то здесь совпадают (с точностью до компонентов, стоящих в скобках) значительно более конкретные классы (б) и (б'). Таким образом, по правилу 1, лексема *кэганин* в разбираемом примере семантически соответствует объектной, а не субъектной валентности лексемы *тэатэ*, что и эксплицируется в русском переводе.

Для второго примера, где в роли X выступает лексема *гакусэй* «студент», а в роли Y — группа *сакубун но тэйсюцу* «подача сочинения», таких четких семантических предпочтений не прослеживается. Поэтому действует вторая часть правила 1, и *гакусэй* интерпретируется как субъект предиката *тэйсюцу* («подача»).

В описании нашего типа находит свое место большинство понятий,

широко используемых в современных структурно-семантических исследованиях, в том числе «глубинные падежи» Ч. Филлмора [23] (если учесть связь смысловой синтагматики слова с его толкованием [7, с. 249—251], это по существу условные обозначения очень общих валентностно-семантических классов слов), аппарат лексических функций [13, с. 19—26] и др.

Существенно для описываемого подхода также то, что в его рамках отсутствует абсолютизация границы между словарем и грамматикой и можно говорить о большей или меньшей грамматичности того или иного явления в зависимости от количественных, объективно измеримых данных о его повторяемости в текстах или в системе языка. Тем самым, поскольку повторяемость, частотность языковых фактов служит одним из важных критериев их лингвистической нормативности, этот подход в явном виде утверждает и обосновывает то традиционное представление, что понятие грамматики тесно связано с понятием языковой нормы. Вместе с тем постепенность перехода между словарем и грамматикой и значимость здесь статистических факторов согласуется, по-видимому, с некоторыми механизмами языкового развития, объясняя, например, большую устойчивость по отношению к процессам грамматической аналогии тех слов, которые характеризуются также и большей частотностью, возможность влияния на грамматический строй языка массовых заимствований в него однотипной лексики и т. д.

Таким образом, предложенная трактовка противопоставления словаря и грамматики с позиций словоцентрического подхода к лингвистическому описанию оказывается достаточно универсальной и продуктивной, позволяя охватить и непротиворечивым образом увязать между собой как различные формализованные средства и методы построения структурно-грамматических описаний преимущественно прикладной направленности, так и многие традиционные лингвистические взгляды и понятия, выработанные в связи с соответствующими проблемами теоретическим языкознанием, ориентированным на человека, а не машину. Думается, что такую совместимость в рамках данной трактовки собственно прикладных и общезыковедческих принципов можно считать доводом в ее пользу, так как оптимальной для прикладных целей может быть, очевидно, лишь та методология, которая в полной мере использует достижения фундаментальных исследований в той же области науки — в нашем случае, в области общего языкознания, — и наоборот, серьезные теоретические разработки не могут игнорировать проблемы, выдвигаемые прикладными задачами, но должны включать в себя принципиальные основы их решения.

СПИСОК ЛИТЕРАТУРЫ

1. Кулагина О. С., Мельчук И. А., Эрстов К. О. Об одной возможной системе машинного перевода // Предв. публ. проблем. группы по экспериментальной и прикладной лингвистике. Вып. 21. М., 1971.
2. Шалапина З. М. Англо-русский многоаспектный автоматический словарь (АРМАС) // Машинный перевод и прикладная лингвистика. Вып. 17. М., 1974.
3. Отчет № 1 по теме «Создание англо-русского словаря по радиоэлектронике с грамматическим обеспечением» хоздоговора между ВИНТИ ГКНТ при СМ СССР и АН СССР и МГПИИЯ им. М. Тореза. Отв. исп. З. М. Шалапина. М., 1973.
4. Шалапина З. М. Семантико-синтаксический анализ в системе англо-русского автоматического перевода (АРАП) // Предв. публ. проблем. группы по экспериментальной и прикладной лингвистике. Вып. 47. М., 1974.
5. Шалапина З. М. Система АРАП: словарь, грамматика, и их использование при автоматическом анализе // Международный семинар по машинному переводу. Москва, 25—27 ноября 1975 г.: Тез. докл. и сообщ. М., 1975.

6. *Шаляпина Э. М.* Многоаспектный автоматический словарь с грамматическим обеспечением и его использование при автоматическом анализе текстов // Вопросы кибернетики. Формализация языковых данных. Вып. 17. М., 1977.
7. *Шаляпина Э. М.* Формальный язык для записи толкований слов и словосочетаний // Проблемы кибернетики. Вып. 36. М., 1979.
8. *Шаляпина Э. М.* Макет лингвистического обеспечения системы японско-русского автоматического перевода ЯРАП: общая структура и основные компоненты. М., 1980.
9. *Shalyapina Z. M.* Automatic translation as a model of the human translation activity // Intern. forum on information and documentation. 1980. V. 5. № 2.
10. *Shalyapina Z. M.* Problems of formal representation of text structure from the point of view of automatic translation // COLING 80: Proc. of the 8th Intern. conf. on computational linguistics. Sept. 30 — Oct. 4, 1980. Tokyo, 1980.
11. *Алпатов В. М.* О двух подходах к выделению основных единиц языка // ВЯ. 1982. № 6.
12. *Richens R. H., Booth A. D.* Some methods of mechanized translation // Machine translation of languages. N. Y., 1955. P. 49 ff.
13. *Жолковский А. К., Мельчук И. А.* К построению действующей модели языка «Смысл \Leftrightarrow Текст» // Машинный перевод и прикладная лингвистика. Вып. 11. М., 1969.
14. *Падучева Е. В.* Некоторые проблемы моделирования соответствия между текстом и смыслом в языке // ИАН СЛЯ. 1975. № 6. С. 557.
15. *Общее языкознание. Внутренняя структура языка.* М., 1972.
16. *Шаляпина Э. М.* Коммуникативная организация предложения: функциональное содержание и средства выражения // Восточное языкознание: грамматическое и актуальное членение предложения. М., 1984.
17. *Шаляпина Э. М.* О двух типах условий кореферентности валентностей в предикатно-актантных конструкциях // Типология конструкций с предикатными актантами. Л., 1985.
18. *Чехов А. С.* Местоимение *it* в англо-русском многоаспектном автоматическом словаре // Машинный перевод и прикладная лингвистика. Вып. 17. М., 1974.
19. *Афоняшева Л. А., Никанорова Т. Г.* Синтаксическое представление абсолютной конструкции для целей англо-русского перевода // Машинный перевод и прикладная лингвистика. Вып. 19. М., 1981.
20. *Афоняшева Л. А., Никанорова Т. Г.* Синтаксическое описание глагольных *-ing* форм для целей автоматического анализа английского текста // Машинный перевод и прикладная лингвистика. Вып. 20. М., 1980.
21. *Шрейдер Ю. А.* Что такое математическая модель языка. М., 1971.
22. *Новый японско-русский словарь.* Токио (Хакуюся), 1974.
23. *Fillmore Ch.* The case for case // Universals in linguistic theory. N. Y., 1968.