

Прикладные разработки

А.Ш. Каганов

Инструментальное исследование спектральных характеристик речи в задаче криминалистической идентификации говорящего: теоретические основания и технология исследования

Темой настоящей статьи является технология исследования спектральных характеристик речи диктора в задаче криминалистической идентификации говорящего. В статье рассматриваются также те теоретические основания, которые связывают спектральные характеристики речи с биометрическими параметрами речевого аппарата человека. Статья основана на литературных источниках и материале реальных криминалистических экспертиз.

Ключевые слова: звучащая речь, криминалистика, судебная экспертиза, идентификация, форманта, спектральный анализ, биометрия.

Введение

Характерной особенностью задачи криминалистической идентификации личности по голосу и звучащей является её многоаспектность и (как следствие) комплексность. Целью настоящей статьи не является рассмотрение комплексной универсальной методики идентификации личности по голосу и звучащей речи. Заметим только, что успешное решение этой задачи требует интеграции знаний из целого ряда гуманитарных, естественных и технических наук. Именно концепция “стыка”, интегрируя знания различных наук, позволяет многоаспектно проанализировать голос и звучащую речь диктора-фигуранта экспертизы и расширить тем самым доказательную базу криминалистического идентификационного исследования¹.

В процессе решения задачи криминалистической идентификации среди множества факторов, характеризующих конкретного говорящего, необходимо проанализировать и работу органов речевого аппарата, придающих голосу индивидуальную тембровую окраску и формирующих поток звуков речи. Иными словами, необходимо исследовать вторую независимую компоненту процесса речеобразования в модели Гельмгольца: формирование акустического качества звука за счёт возбуждения резонансных частот артикуляционного тракта (у Гельмгольца), или фильтрации (в современном рассмотрении)².

Метод фонетико-спектрального анализа

Процесс анализа спектральных характеристик речи диктора-фигуранта начнем с рассмотрения традиционного метода фонетико-спектрального анализа. Для этого, прежде всего, обсудим вопросы использования спектральных характеристик речи в качестве устойчивых идентификационных признаков в условиях ограниченного объема речевого материала, поступившего на исследование, и неполной сопоставимости исходной и сравнительной записей³.

Необходимость использования стабильных показателей в сравнительном идентификационном исследовании объясняется тем требованием, что только устойчивые признаки могут быть задействованы для установления наличия или отсутствия тождества того или иного материального объекта – в данном случае человека – по его отображениям, т.е. для идентификации в её криминалистическом понимании.

В практике традиционного формантного анализа широко используется подход, при котором для обеспечения стабильности результатов измерений форманты⁴ гласных оцениваются в одинаковом фонетическом контексте (в «триадах») с учётом комбинаторных и позиционных изменений⁵. Однако, в реальных задачах криминалистической идентификации личности объем речевого материала обычно ограничен. По этой причине эксперт-исследователь далеко не во всех ситуациях может воспользоваться указанной технологией и отобрать достаточное для представительной выборки количество «триад». В данном случае следует вести речь о решении задачи с ограничениями, наложенными на начальные условия.

Таким образом, появилась необходимость развития теории применительно к речевому материалу ограниченного, или малого

объема. Эффективной в этих случаях может оказаться методика получения и оценки значений формантных характеристик гласных, позволяющая сравнивать фонетические контексты, традиционно определяемые как несовпадающие, но такие, в которых сохраняется акустическое качество исследуемых звуков. Например, при сопоставлении ударных и безударных гласных необходимо учесть качества согласных по таким параметрам, как твердость-мягкость (предшествующего согласного, последующего согласного, т.е. позиции типа *tat* – *ta't*, *t'at* – *t'at'*, в которых изменяется степень сдвинутости фокуса артикуляции гласного вперед). Так, для гласного [а] в словах *x[a]тят*, *T[a]тьяна*, *c[a]дись*, *sig[a]рета* общей является позиция *ta't*; для гласного [ы] в словах *пассаж[ы]р*, *ж[ы]вой*, *тяж[ы]ло* общей является позиция типа *tat*.

В современных исследованиях формантного состава ударных гласных показано⁶, что в позициях типа *t'at* и *t'at'* степень продвижения гласных вперед практически совпадает. Поэтому в общем ряду можно рассматривать фонетические позиции гласного в таких словах, как «слинял» и «пяты»; «нюхать» и «люди»; «тётка» и «тётя» и др.

Далее следует *усреднить* значения формант исследуемого гласного по анализируемой выборке, ибо, как указывал Л.Р. Зиндер «.. результаты (измерений – А.К.), даже если перед нами несколько произнесений одного и того же диктора, будут колебаться»⁷. Близость числовых значений *усредненных* значений формант исходной и сравнительной записей явится в этом случае одной из разновидностей стабильности данного показателя, а сами средние значения формант можно рассматривать в качестве устойчивого идентификационного признака.

Продолжая обсуждение стабильных спектральных характеристик, заметим, что в настоящее время не получено убедительных доказательств того, что значения формант являются единственными индивидуализирующими признаками для определения акустического качества звуков. Как показывает анализ речевого материала в многообразии конкретных криминалистических экспертиз идентификации личности по голосу и звучащей речи, точнее было бы говорить о том, что существует несколько параллельно функционирующих систем, позволяющих выделить индивидуализирующие признаки спектральной группы, наподобие того как существует по крайней мере две параллельно действующие системы голосообразования⁸.

Именно за счёт существования нескольких систем признаков различного типа обеспечивается устойчивость речевой коммуни-

кации, в том числе в условиях помех, шумов и искажений (что особенно важно в идентификационном исследовании говорящего по реальным звукозаписям). Указанные параллельные системы признаков могут быть реализованы, как отмечается в работе⁹, в рамках допустимой для речевого аппарата человека акустической теории речеобразования.

Возвращаясь к параллельно функционирующим системам, позволяющим выделить идентификационные признаки спектральной группы, следует сказать, что устойчивые признаки могут иметь разную природу. К числу таких устойчивых признаков относятся и формантные соотношения – F_2/F_1 , F_3/F_1 , F_3/F_2 и т.д. Анализ этих соотношений необходим при идентификации говорящего, который находится в разном эмоциональном состоянии (спокоен, возбужден, подавлен, испуган, оживлен и др.), в разных речевых условиях (деловой разговор, выступление перед аудиторией, разговор со следователем и др.), т.е. при наличии ограничений, связанных с неполной сопоставимостью исходной и сравнительной записей. В этих ситуациях формантные соотношения обладают большей устойчивостью по сравнению с абсолютными значениями формант и поэтому являются более доказательными идентификационными признаками. Такой вывод основан на нашем опыте решения задачи идентификации говорящего по реальным фонограммам, который показывает, что при изменениях значений формант в силу тех или иных причин (например, в зависимости от ситуации речевого общения, эмоционального состояния говорящего и др.) соотношения формант практически не меняются.

Рассмотрим сравнительные характеристики речи (по признаку соотношения формант F_2/F_1) в разных речевых условиях и в разном эмоциональном состоянии диктора-фигуранта экспертизы¹⁰. Различия обобщены в таблице I и озаглавлены как ситуация 1 и ситуация 2. В первом случае говорящий эмоционально собран, осторожен, краток. Голос звучит сухо, деловито, приглушенно. Во втором случае артикуляторно четко и полно представлена реализация гласных, согласные не напряжены, речь нетороплива (темп речи снижен по сравнению с ситуацией 1 примерно на 10%).

Как видно из таблицы I в ситуации 1 и в ситуации 2 усредненные формантные показатели речи значительно расходятся, но соотношение F_2/F_1 остается практически неизменным – устойчивым, т.е. как отмечалось выше, соотношения формант меняются в незначительной степени или практически не меняются.

Итак, соотношение формант остается стабильным, а данный идентификационный признак – формантные соотношения – оказывается устойчивым даже на материале ограниченного объема.

Таблица I.

параметры формантного анализа		F_2/F_1	
гласный	средняя частота формант, Гц	ситуация 1	ситуация 2
а	исх. – $F_1=535$ и $F_2=1390$ срав.– $F_1=580$ и $F_2=1500$	2,6	2,6
и	исх. – $F_1=310$ и $F_2=2015$ срав.– $F_1=300$ и $F_2=1970$	6,5	6,6
о	исх. – $F_1=457$ и $F_2=945$ срав.– $F_1=390$ и $F_2=840$	2,0	2,2

Продолжая обсуждение таблицы I, заметим, что наилучшее совпадение относительных формантных показателей исходной и сравнительной записей по звуку [а] может быть объяснено тем фактом, что среди всех гласных русского языка гласный [а] наиболее устойчив в спектральном отношении к шумам акустической обстановки и искажениям. В силу вышеприведенных причин спектральные характеристики гласного [а] оказались наиболее устойчивыми к различиям в речевой ситуации, в контекстном окружении, в объёме и качестве речевого материала исходной записи и фонограммы-образца.

Оценивая приведенные результаты исследования акустического качества звуков речи с помощью абсолютных и относительных формантных показателей, отметим, что если средневзвешенное относительное отклонение исходного (ситуация 1) и сравнительного (ситуация 2) речевого материала, подсчитанное по средним значениям формант, в данном примере составило 8,4%, то средневзвешенное относительное отклонение, подсчитанное по отношениям F_2/F_1 (столбцы 2 и 3 таблицы 4.3.1), составляет всего 3,5 %¹¹.

Хотя и тот, и другой показатели находятся в пределах внутридикторской вариативности¹², из приведенных результатов видно, что соотношение формант явилось в данном случае более “сильным” идентификационным признаком, чем средние значения формант. Можно сказать, что как идентификационный признак соотношение формант обладает бóльшим «весом», чем средние значения формант.

Возвращаясь к анализу соотношения второй и первой формант, отметим, что данное соотношение вызывает интерес ещё и по другой причине. Оно не только отражает такое важное для идентификации

личности свойство говорящего, как сдвиг фокуса артикуляции, но и позволяет установить соответствие между отдельными идентификационными признаками перцептивной и инструментальной групп¹³.

Известно, что свойство продвинутости фокуса артикуляции вперёд или назад по сравнению с нормой можно рассматривать в качестве одной из региональных особенностей говорящего. Оно находит отражение в идентификационных признаках перцептивной и акустической групп и характеризует функционально-динамические комплексы (ФДК) артикуляторных навыков индивидуума.

В качестве иллюстрации этого положения рассмотрим данные, полученные в результате одного из проведенных нами экспериментов¹⁴. В эксперименте исследовался голос диктора-москвича, речь которого не имела заметных дефектов и характеризовалась как нормативная (московская норма). В процессе исследования были получены описания перцептивных характеристик качества звучания, а также проведены инструментальные измерения формантных частот. Голос фигуранта относится к группе мужских голосов со средней частотой основного тона $F_0 \approx 130$ Гц. В качестве предствительного показателя вновь были использованы не только абсолютные значения частоты первой, второй и третьей формант F_1 , F_2 , F_3 , но и их отношения F_2/F_1 , F_3/F_2 . Результаты измерения частоты формант на ударном гласном [а] в основных позициях при чтении фигурантом предложенного ему текста дали следующие результаты: $F_1 = 600$ Гц, $F_2 = 1250$ Гц, $F_3 = 2270$ Гц и отношение $F_2/F_1 \approx 2,1$, а $F_3/F_2 \approx 1,8$. Эти данные достаточно близки к соответствующим данным для нормативной русской речи (известно, что значение F_2/F_1 по многочисленным измерениям для мужских голосов близко к $2,0$ ¹⁵).

Затем на фразе: «Синий – синий тата взвился ввысь» и «Эх, тутку тата ухватил» были измерены формантные частоты ударного гласного [а] в слове «тата». В первой фразе преобладает артикуляция звуков, образующихся в передней полости речевого аппарата, а во второй – в задней. Это в какой-то мере имитирует эффект продвинутого вперед/назад фокуса артикуляции и сопровождается произнесением гласного [а] в слове «тата» как продвинутого вперед/назад. Оказалось, что отношение F_2/F_1 для гласного [а] в первой фразе равно 3,5, а во второй – 1,6. Это даёт в первом случае значительное превышение над измеренной величиной этого соотношения для чтения указанным диктором текста (приблизительно на 65%), а во втором – снижение почти на 24% по сравнению с чтением.

Расширим рамки анализа и рассмотрим соотношения не только первой и второй, но и первой и третьей, второй и третьей формант, распространив исследование не только на соседние, но и на более отдаленные друг от друга максимумы в спектре речевого сигнала.

В таблице II приведены сравнительные характеристики речи (по признаку соотношений формант F_2/F_1 , F_3/F_2 , F_3/F_1) в разных речевых условиях и в неодинаковом эмоциональном состоянии говорящего. В первом случае для речи диктора-фигуранта криминалистической экспертизы характерно беспокойство в сочетании с растерянностью. Оттенки голоса варьируются от умоляющих до требовательных и даже гневных. В ситуации 2 речь более сдержана, обдуманна, достаточно логична.

Из таблицы II видно, что характеристики соотношений формант, полученные в процессе формантного анализа гласных [а], [и], [о] в сильных позициях для голоса и речи фигуранта в записи исходных телефонных разговоров (т.е. в ситуации 1) и в сравнительной записи разговора со следователем (ситуация 2), меняются в незначительной степени или практически не меняются.

Таблица II.

гласный	статистические значения параметров формантного анализа		
	параметр	ситуация 1	ситуация 2
[а]	F_2/F_1	2,5	2,5
	F_3/F_2	1,7	1,7
	F_3/F_1	4,4	4,4
[и]	F_2/F_1	7,2	7,1
	F_3/F_2	1,3	1,2
	F_3/F_1	9,9	9,1
[о]	F_2/F_1	1,9	2,1
	F_3/F_2	2,6	2,5
	F_3/F_1	5,3	5,5

Таким образом, данные идентификационные признаки – формантные соотношения F_2/F_1 , F_3/F_2 , F_3/F_1 – вновь показали свою устойчивость.

Оценивая полученные результаты по выбранной метрике, отметим, что средневзвешенное относительное отклонение исходного (ситуация 1) и сравнительного (ситуация 2) речевого материала,

подсчитанное по отношениям F_2/F_1 , F_3/F_2 , F_3/F_1 (столбцы 3 и 4 таблицы II), составляет около 3,5%, что существенно меньше среднестатистической внутридикторской вариативности.

Интересно отметить, что вновь, как и в предыдущем примере, наилучшее совпадение параметров исходной и сравнительной записей (теперь уже и с учетом F_3) получено по звуку [а]. Таким образом, характеристики гласного [а] вновь оказались наиболее устойчивыми к различиям речевого материала, связанным не только с наличием шумов акустической обстановки и искажений (в данном примере на фонограмме-образце), но и с несовпадением речевой ситуации 1 и речевой ситуации 2, т.е. к ограничениям, наложенным на речевой материал реальных криминалистических экспертиз.

Теоретические основания выделения устойчивых идентификационных признаков спектральной группы

Вопрос о теоретических основаниях полученных нами результатов требует специального рассмотрения. По сути дела речь идет о теоретическом фундаменте успешного выделения устойчивых идентификационных признаков спектральной группы в условиях ограничений, накладываемых на речевой материал реальных криминалистических экспертиз.

Ещё в середине 50-х годов прошлого века отечественными исследователями Л.А. Варшавским и И.М. Литваком была высказана гипотеза о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра¹⁶. При этом форманты (т.е. максимумы в спектре речевого сигнала) являются лишь доступным для речевого аппарата способом достижения необходимых полосных соотношений.

Время показало, что высказанная авторами гипотезы мысль оказалась справедливой, фундаментальной и обладающей большой объяснительной силой. Позднее, при расширении прикладных криминалистических исследований звучащей речи, возникли новые вопросы. Потребовалось *развитие этой теории* применительно к речевому материалу ограниченного объема, к речевому материалу на т.н. «этнических» языках и др., т.е. применительно к решению задачи идентификации говорящего при наличии ограничений, налагаемых на начальные условия.

Кроме того, возникла необходимость теоретического обоснования такой важной задачи медико-криминалистической идентификации, как задача определения биометрических характеристик речевого аппарата диктора-фигуранта экспертизы¹⁷.

Рассматривая под ракурсом гипотезы Л.А. Варшавского и И.М. Литвака результаты проведенных нами идентификационных исследований и лабораторных экспериментов, легко убедиться в том, что данная гипотеза в первых двух из приведенных выше примеров реализовалась с помощью отношений средних значений частот формант (F_2/F_1). Таким образом, подтвердилось предположение о том, что форманты являются доступным для речевого аппарата способом достижения необходимых полосных соотношений.

С другой стороны, результаты, идентификационного исследования, приведенные в таблице II, показывают, что гипотеза получила дальнейшее подтверждение при сопоставлении характеристик речи диктора-фигуранта экспертизы не только в соседних, но и в более отдаленные друг от друга полосах спектра.

Таким образом, теоретическое положение о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра, получает новое развитие при сопоставлении фонограмм речи одного и того же говорящего (т.е. при установлении наличия или отсутствия индивидуально-конкретного тождества в криминалистическом понимании).

Распространим теперь гипотезу Л.А. Варшавского и И.М. Литвака на решение задачи идентификации диктора методом опорных сегментов (методом формантного выравнивания).

Поясним, что первоначально данный метод был предложен в рамках идентификации дикторов, говорящих на т.н. «этническом» языке. Актуальность этой задачи криминалистической идентификации вызвана тем обстоятельством, что в многообразии речевого материала, поступающего на экспертные исследования, всё больший удельный вес приобретают звукозаписи, в которых фигуранты говорят на т.н. этнических языках. В основном это языки юга России и бывших южных республик СССР.

Единственным выходом из создавшегося положения является создание методики, позволяющей опытным экспертам-исследователям, входящим в экспертную комиссию, но не владеющим тем языком, на котором говорит фигурант экспертизы, идентифицировать говорящего на этническом языке.

Указанная методика, которая была разработана специалистами Центра речевых технологий (г. Санкт-Петербург) и описана в целом ряде работ¹⁸. В настоящее время она успешно применяется в ряде государственных и негосударственных судебно-экспертных учреждений Российской Федерации.

Целью настоящей статьи не является рассмотрение комплексной универсальной методики идентификации личности по голосу и звучащей этнической речи. Отметим только, что данная задача, обладая определенной спецификой, во-первых, по-прежнему является типичной задачей речевой кибернетики и требует сочетания (в рамках концепции «стыка»¹⁹) знаний из различных областей гуманитарных, естественных и технических наук; во-вторых, предлагаемая методика может быть применена и для идентификации личности дикторов, говорящих на русском языке.

В наших исследованиях нас, прежде всего, интересовал вопрос о *теоретических основаниях* метода опорных сегментов. При этом речь идет о том разделе идентификационной экспертизы, в котором опосредованно происходит сравнение геометрической конфигурации речеобразующего тракта дикторов-фигурантов экспертиз на сопоставимых фрагментах исходной и сравнительной фонограмм²⁰. В рамках метода опорных сегментов указанное сравнение проводится путем сопоставления формантного состава на участках реализации в речевом сигнале артикуляторно подобных акустических событий.

Таким образом, в рамках метода формантного выравнивания экспертами отбираются не одинаковые *фонетические* реализации (т.е. «триады» в словах, в слогах), как это происходит в традиционном методе фонетико-спектрального анализа, описанном выше, а *артикуляторно* подобные события. Такие события соответствуют ситуации, когда речеобразующие органы человека находятся в положении, реализующем одно и то же фонетически значимое звучание. Иными словами, для сравнения подбираются не формально, но реально фонетически одинаковые звуки. Фонетически незначимая часть формантного спектра таких звуков и образует идентифицирующие признаки сравниваемой артикуляции²¹.

По мысли разработчиков основная идея метода формантного выравнивания основана на предположении, что каждый диктор в процессе производства речи может изменять конфигурацию своего речевого тракта лишь в рамках жёстких анатомических ограничений, позволяющих изменять геометрические размеры акустического волновода – артикуляторного тракта только в определённой

степени. Каждая конфигурация управляется диктором только по своим основным геометрическим размерам, которые обеспечивают реализацию целевых акустических резонансных свойств для низкочастотной части спектра или, вообще говоря, только первых двух-трёх формантных максимумов.

Резонансные свойства каждой конфигурации вокального тракта для четвёртой и более высоких формант обычно диктором не контролируются и задаются имеющимися анатомическими ограничениями на возможные изменения конфигурации артикуляторного тракта диктора. На языке формантного описания это приводит к тому, что при фиксированных значениях первых формант более высокие по частоте форманты у данного диктора могут занимать только более-менее стабильные индивидуальные положения.

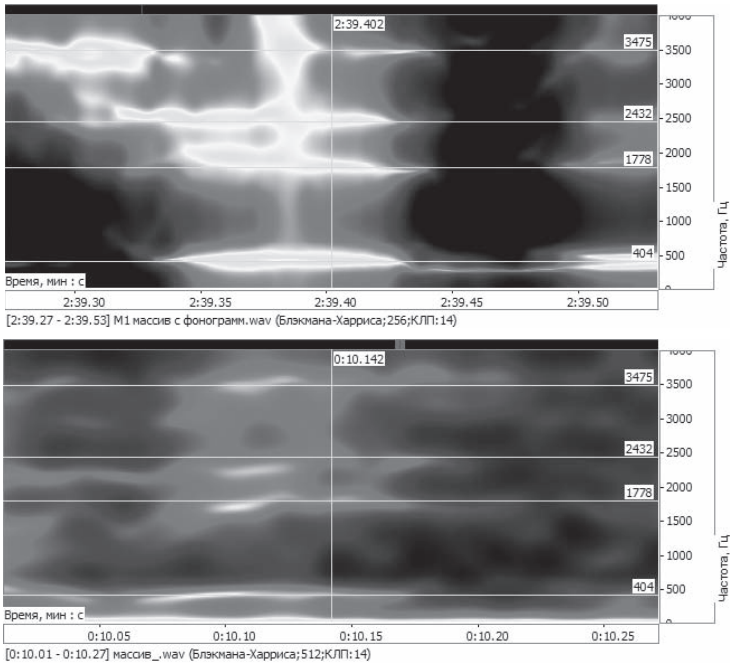
Таким образом, высокочастотная формантная структура речи диктора при фиксированной низкочастотной формантной структуре часто является стабильной²², неизменной во времени и не подлежит изменению (умышленному или неумышленному) диктором-фигурантом экспертизы, т.е. высокочастотная формантная структура – это, по выражению разработчиков метода, «устойчивая биометрическая характеристика» индивидуума. Сравнительный анализ этих высокочастотных формантных структур (а через них и геометрических характеристик вокального тракта) позволяет идентифицировать дикторов с высокой степенью надёжности²³.

По своей сути метод формантного выравнивания соответствует косвенным измерениям и сравнению мелких деталей геометрического строения артикуляторного тракта дикторов. В процессе идентификации эксперт-исследователь ищет совпадающие конфигурации речевых трактов в исходной и в сравнительной записях. При достаточном числе найденных совпавших конфигураций такое совпадение считается, по мысли авторов метода, не случайным, и может быть принято решение об идентификации дикторов.

Вопрос о *теоретических основаниях* метода опорных сегментов требует специального рассмотрения. Хотя в случае использования данного метода задача криминалистической идентификации личности по голосу и звучащей речи приобретает свою специфику, теоретические основания, которые служат фундаментом выявления устойчивых идентификационных признаков инструментальной группы, связанных со спектральными характеристиками речи индивидуума, и в этом случае сохраняют свою справедливость и объяснительную силу.

Проиллюстрируем справедливость гипотезы о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра применительно к методу опорных сегментов (методу формантного выравнивания) на конкретном примере²⁴.

Рис. 1 Динамические спектрограммы [i]-образного гласного.



На рис. 1 приведены динамические спектрограммы [i]-образного гласного в исходной (слева) и в сравнительной (справа) записях этнической речи. В приведенных реализациях совпадают положения четырех формант исследуемого звука: $F_1=400$ Гц, $F_2=1780$ Гц, $F_3=2430$ Гц, $F_4=3480$ Гц.

На рис. 2 показаны т.н. спектральные срезы, соответствующие тем временным отметкам в исходной и в сравнительной фонограммах, которые отмечены на рис. 1 вертикальными курсорами. Черная (верхняя) кривая соответствует произнесению неизвестного М, серая (нижняя) произнесению подозреваемого, т.е. фонограмме-образцу.

Рис. 2. Мгновенные спектры анализируемого [i]-образного гласного.

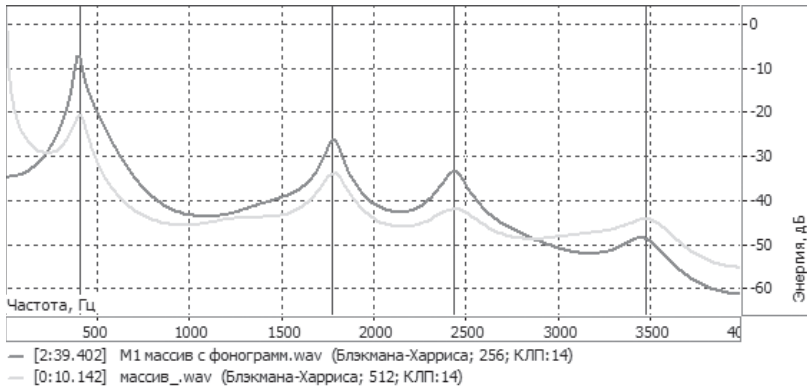


Рис. 1 иллюстрирует высокую степень совпадения характерных особенностей реализации спектрально-временной картины при произнесении анализируемого [i]-образного звука, а рис. 2 высокую степень совпадения *соотношений уровней сигнала в полосах спектра*. И то и другое косвенным образом свидетельствует о совпадении геометрии речевого тракта исследуемых дикторов.

Итак, гипотеза Л.А. Варшавского и И.М. Литвака о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра, показала свою справедливость и объяснительную силу и в случае метода опорных сегментов.

При этом форманты (т.е. максимумы в спектре речевого сигнала) вновь – как и в случае традиционного фонетико-спектрального анализа, когда речь идет о поиске и сравнении (как слуховом, так и инструментальном) реализаций одинаковых фонем в одинаковом речевом контексте²⁵, – выступают в качестве доступного для речевого аппарата способа достижения необходимых полосных соотношений.

Таким образом, анализ соотношений уровня сигнала в полосах спектра является тем *общим теоретическим фундаментом*, на котором базируется методология выявления устойчивых идентификационных признаков инструментальной группы, связанных со спектральными характеристиками речи, применительно к решению задачи идентификации личности говорящего при наличии ограничений, наложенных на начальные условия.

Определение биометрических характеристик речевого аппарата

Обратимся теперь к задаче определения биометрических характеристик речевого аппарата диктора-фигуранта экспертизы.

Среди нескольких определений биометрии с точки зрения криминалистики и судебной экспертизы наиболее интересным, на наш взгляд, является то, в котором биометрия определяется как идентификация человека по уникальным, присущим только ему биологическим признакам.

При решении задачи медико-криминалистической идентификации личности говорящего естественным образом возникает вопрос о путях использования биометрических показателей речевого тракта диктора в качестве индивидуализирующих идентификационных признаков.

Некоторые из таких идентификационных признаков могут быть, как мы убедились выше, выявлены в результате спектрального анализа речевого сигнала. Рассмотрим связь спектральных характеристик речи и отдельных биометрических параметров речевого тракта и проанализируем те *теоретические основания*, которые являются основой медико-криминалистической идентификации личности по биометрическим характеристикам речевого тракта.

Сравнительный анализ речевого сигнала в норме и при определенных патологиях (например, при болезнях гортани), проведенный в работе²⁶, показал, что при исключении в силу тех или иных обстоятельств (например, в результате хирургического вмешательства) из речеобразующего тракта Морганиева желудочка²⁷ в спектральной картине исчезает третья форманта F3 для всех русских гласных звуков кроме [и]²⁸. Из этого наблюдения можно сделать вывод, что Морганиев желудочек отвечает за формирование третьей форманты в таких русских гласных, как [а], [е], [о], [у].

Другими словами исчезновение из спектральной картины речевого сигнала третьей форманты F3 означает, что в той полосе спектра, где находится F3, исчезает спектральный максимум, который соотносится с соответствующим резонансом. Таким образом, происходит перераспределение соотношений уровней речевого сигнала в полосах спектра, что подтверждает правильность нашего наблюдения о применимости гипотезы Л.А. Варшавского и И.М. Литвака к оценке биометрических характеристик речевого тракта диктора.

При этом форманты (т.е. максимумы в спектре речевого сигнала) по-прежнему выполняют роль доступного для речевого аппарата способа достижения необходимых полосных соотношений.

Продолжая обсуждение вопросов использования биометрических характеристик речевого тракта диктора в качестве идентификационных признаков, распространим теперь гипотезу о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра, на теоретическое обоснование той части задачи построения речевого психологического портрета, которая связана с определением размера ротового отверстия говорящего фигуранта.

Интуитивно ясно, что при построении психологического портрета по голосу и звучащей речи информация о дикторе-фигуранте экспертизы передается акустическим путем. Д.п.н., проф. В.П. Морозов в своих работах не только выделил целый ряд видов и разновидностей этой информации, но и доказал их достаточно объективную связь с психофизическими особенностями говорящего.

Это (как указывает автор исследований) касается, прежде всего, таких чисто физических свойств человека, которые легко проверить, например, возраст, пол и даже его рост и вес. Последнее объясняется тем фактом, что акустические особенности голоса зависят от размеров голосовых складок, величины резонаторов голосового тракта «обследуемого» (чем больше эти размеры, тем ниже голос и «гуще» его тембр; и наоборот – чем меньше эти размеры, тем выше голос и светлее тембр).

В своих недавних исследованиях В.П. Морозову удалось показать, что субъект восприятия способен по голосу даже воссоздать общий облик говорящего, нарисовать его портрет с довольно точным изображением деталей, например, относительных размеров рта²⁹.

Рис. 3. Сравнительные интегральные спектры речи при разной ширине рта.³⁰



Для оценки ширины рта диктора-фигуранта такой экспертизы В.П. Морозовым был введен специальный коэффициент k . Данный коэффициент характеризует процентное отношение акустической энергии высокочастотной части спектра звука в *полосе* 1–11 кГц к энергии всего звука в целом (т.е. к энергии в *полосе* 0,1–11 кГц). На рис. 3 показаны результаты одного из экспериментов, проведенных В.П. Морозовым. Диктору с широким ротовым отверстием соответствует значение $k=28,6\%$, с узким – $k=16,0\%$.

Следовательно, по значениям коэффициента k (т.е. по соотношениям уровней сигнала в полосах спектра) мы можем судить о таком биометрическом показателе диктора-фигуранта экспертизы, как размер его ротового отверстия.

Таким образом, и в этом случае, т.е. в задаче определения размера ротового отверстия говорящего, гипотеза о том, что акустическое качество звуков определяется соотношением уровней сигнала в полосах спектра, показала свою справедливость и объяснительную силу.

Заключение

Итак, анализ речевого материала экспертно-криминалистических исследований, проведенный с учетом ограниченности его объема, неполной сопоставимости по речевой ситуации и качеству записи, показал, что полученные результаты полностью согласуются с гипотезой о том, что акустическое качество звуков определяется соотношением уровня сигнала в полосах спектра. При этом форманты (т.е. максимумы в спектре речевого сигнала) выступают в качестве доступного для речевого аппарата индивидуума способа достижения необходимых полосных соотношений.

С другой стороны, эксперименты, проведенные нами как в лабораторных условиях, так и на речевом материале реальных экспертов, наглядно продемонстрировали связь идентификационных признаков перцептивной и инструментальной групп, показали, что указанные признаки находятся в отношениях взаимодополнения, (частично) взаимосоответствия и взаимопроникновения, часто неразделимых.

Наконец, результаты экспериментов по определению биометрических характеристик речеобразующего тракта диктора-фигуранта экспертизы, полученные к.т.н. А.Н. Квасовым и д.п.н., проф. В.П. Морозовым, получили свое теоретическое обоснование в том, что акустическое качество звуков речи индивидуума определяется соотношением уровней сигнала в полосах спектра.

- ¹ *Каганов А.Ш.* Криминалистическая идентификация личности по голосу и звучащей речи. Монография. 2-е изд. перераб. и доп. М.: Юрлитинформ, 2012. С. 51.
- ² *Фант Г.* Акустическая теория речеобразования / Пер. с англ. под ред. В.М. Григорьева. М.: Наука, 1964. С. 27.
- ³ Для значительного числа фонограмм, поступающих на экспертно-криминалистическое исследование, характерны недостаточный объем, зашумленность, низкая разборчивость, недостаточная (на первый взгляд) сопоставимость исходной и сравнительной записей, различие технических характеристик каналов записи и другие ограничения, которые с математической точки зрения можно рассматривать как ограничения, налагаемые на начальные условия решения задачи.
- ⁴ Форманта – акустическая характеристика звука речи, связанная с уровнем энергии сигнала в соответствующей полосе спектра и участвующая в образовании тембра звука. Формантный – относящийся к форманте, формантам.
- ⁵ *Каганов А.Ш.* Указ. соч. С. 152.
- ⁶ *Златоустова Л.В.* Роль мягких согласных при правом и левом контексте: Доклад на «Ломоносовских чтениях» (25 апреля 2005 г. в МГУ, Москва).
- ⁷ *Зиндер Л.Р.* Общая фонетика, М.: Высшая школа, 1979. С. 35.
- ⁸ *Жинкин Н.И.* О теориях голосообразования // Мышление и речь / Под ред. Жинкина Н.И. и Шемякина Ф.Н. М., 1963.
- ⁹ *Галунов В.И., Гарбарук В.И.* Акустическая теория речеобразования и система фонетических признаков // Материалы международной конференции «100 лет экспериментальной фонетике в России» 1–4 февраля 2001 г. СПб.: Филол. ф-т Санкт-Петербургского гос. ун-та, 2001. С. 58–62.
- ¹⁰ Здесь и далее приводятся примеры из наших экспертных исследований, выполненных по постановлениям следственных органов и определениям судов.
- ¹¹ Для оценки близости формантных показателей использовалась метрика $\delta_2 = (\sum_j (1/|y_j|)(|x_j| - |y_j|) \cdot 100\%) / N$.
- ¹² Допустимая внутридикторская вариативность составляет по разным данным от 15 до 20%.
- ¹³ Напомним, что идентификационные признаки, по которым устанавливается наличие (или отсутствие) индивидуально-конкретного тождества между голосами неизвестного и подозреваемого делятся на перцептивные (аудитивные, лингвистические) и инструментальные, или акустические: признаки, связанные с параметрами основного тона, и спектральные.
- ¹⁴ По причине ограниченности объема настоящей статьи результаты других наших экспериментов на материале реальных экспертных исследований, не приводятся. Интересующихся этими результатами, отсылаем к монографии *Каганов А.Ш.* Указ. соч.
- ¹⁵ *Halle M.* The Sound Pattern of Russian. A linguistic and acoustical investigation/ Ed. by R. Jakobson, Mouton & Co. S-Gravenhage, 1959. 206 p.
- ¹⁶ *Варшавский Л.А., Литвак И.М.* Исследование формантного состава и не-

- которых других физических характеристик звуков русской речи // Проблемы физиологической акустики. М.–Л.: АН СССР, 1955. Т.3. С. 5–17.
- ¹⁷ Интуитивно ясно, что измерение указанных характеристик по материалам звукозаписей будет носить косвенный, опосредованный характер.
- ¹⁸ Коваль С.Л. и др. Использование метода формантного выравнивания для проведения инструментальной части идентификационного исследования говорящего // Современные методы, технические и программные средства, используемые в криминалистической экспертизе звукозаписей: Методическое пособие для экспертов под общ. ред. А.Ш. Каганова. М.: РФЦ СЭ при Минюсте России, 2003. С. 107–132; Зубова П.И., Коваль С.Л. Идентификации личности по голосу и звучащей речи на основе комплексного анализа фонограмм // Теория и практика судебной экспертизы: Научно-практич. журнал. 2007. № 3 (7).
- ¹⁹ Каганов А.Ш. Указ. соч. С. 49.
- ²⁰ Поясним, что целью задачи криминалистической идентификации с юридической точки зрения является установление *индивидуально-конкретного тождества* между голосом неизвестного, записанного на исходной фонограмме (которая могла быть получена, например, в процессе оперативно-розыскных мероприятий или предоставлена одной из сторон в судебном разбирательстве), и голосом подозреваемого, записанного на фонограмме-образце.
- ²¹ Таким образом, в методе формантного выравнивания находит своё подтверждение мысль целого ряда авторов, что в задачах КЭЗ (криминалистической экспертизы звукозаписей) наиболее информативными часто оказываются те составляющие звукового сигнала, которые вообще не воспринимаются на слух.
- ²² Напомним, что для идентификации говорящего могут быть использованы только *стабильные* показатели (см. например, Каганов А.Ш. Указ. соч.).
- ²³ Коваль С.Л. и др. Указ. соч. С. 107–132.
- ²⁴ Приведенный пример взят из реального экспертного исследования фонограмм на этническом языке. Исследование было выполнено экспертами судебно-экспертного учреждения «ФоренЭкс» (г. Санкт-Петербург).
- ²⁵ Каганов А.Ш. Указ. соч. С. 152.
- ²⁶ Квасов А.Н. Модель голосообразования и анализ речевого сигнала в норме и при патологии: Автореф. дисс... канд. тех. наук. Томск: ТУСУР, 2007.
- ²⁷ Морганиев желудочек [ventriculus laryngis, ventriculus laryngis (Morgagnii); син.: желудочек гортанный, желудочек гортани] – парное углубление слизистой оболочки гортани между преддверной и голосовой складками.
- ²⁸ Квасов А.Н. Указ. соч. С. 10.
- ²⁹ Морозов В.П. Невербальная коммуникация. Экспериментально-психологические исследования. М.: Ин-т психологии РАН, 2011. С. 169–170.
- ³⁰ Иллюстрация из книги: Морозов В.П. Указ. соч. С. 170.

З.М. Шаляпина

Некоторые возможности автоматической оптимизации лингвистического обеспечения системы обработки текстов при сущностном подходе к его организации

На опыте разработки экспериментального комплекса для японско-русского автоматического перевода ЯРАП (Институт востоковедения РАН) рассматриваются возможности, предоставляемые сущностным подходом к языку для моделирования при создании лингвистического обеспечения систем обработки текстов процессов усвоения языковых знаний человеком.

Ключевые слова: автоматическая обработка текстов, лингвистическое обеспечение, оптимизация, сущностный подход к языку

1. Введение

Лингвистическое обеспечение любой системы автоматической обработки текстов, ориентированной на получение результатов достаточно хорошего качества, должно содержать настолько много разнообразных и сложным образом организованных лингвистических сведений, что проблемы слежения за этим обеспечением при его пополнении и совершенствовании становятся весьма нетривиальными. Столь же нетривиальный характер приобретает и задача оптимального использования такого лингвистического обеспечения при обработке конкретных текстов. Практически невозможно априори предусмотреть и жестко зафиксировать все мыслимые комбинации имеющихся лингвистических сведений, которые могут оказаться релевантными для определенного отдельного текста, а также однозначно задать наиболее экономичный и эффективный в каждом конкретном случае порядок их вызова и применения.

Это связано, как представляется, не только с ограниченностью имеющихся в нашем распоряжении средств описания лингвистической информации и управления ею, но и с принципиальными

особенностями самого объекта описания – естественного языка и его функционирования в коммуникативных процессах. Иными словами, проблемы, о которых идет речь, не сводятся только к общим проблемам управления универсальными базами данных и слежения за такими базами, входящим в компетенцию программистов. Эти проблемы имеют также и весьма существенные собственно лингвистические аспекты. В данной, собственно лингвистической, части общие возможности решения соответствующих проблем должны устанавливаться, очевидно, уже в рамках самой лингвистической концепции, положенной в основу той или иной конкретной системы обработки текстов и определяющей состав и принципы организации в ней ее лингвистического обеспечения. Здесь мы хотим рассмотреть некоторые из таких возможностей применительно к задаче автоматического перевода (АП).

В связи с данной задачей нельзя не учитывать то обстоятельство, что лингвистическое обеспечение системы АП играет в ней ту же роль, которую для человека-переводчика выполняют его переводческие знания и навыки, т.е. представляет собой функциональную модель этих знаний и навыков. В той мере, в которой такую интерпретацию лингвистического обеспечения системы АП можно признать справедливой, указанные выше проблемы сближаются, очевидно, с проблемами формирования соответствующих знаний и навыков у человека, обучающегося переводу, и для их эффективного решения целесообразно иметь в виду общие особенности процесса такого обучения.

Важнейшая из них, на наш взгляд, состоит в том, что человек в этом процессе не является пассивным реципиентом информации. Даже если полностью исключить случай самообучения (которое, безусловно, представляет собой процесс гораздо более сложный, чем обучение, и более активный, включающий значительный элемент творческой деятельности) и говорить только об обучении как таковом, оно также не может быть сведено лишь к запоминанию учащимся сообщаемой ему информации. Помимо **запоминания**, при успешном обучении должно обязательно иметь место также **усвоение** получаемых от преподавателя лингвистических сведений. При этом учащийся, по-видимому, каким-то образом преобразует сообщаемые ему сведения и инструкции, соотносит их с ранее имевшейся у него информацией, устанавливает определенную иерархию различных ее фрагментов, выявляет и, насколько это ему доступно, нейтрализует возможные противоречия и т.д. – т.е. сводит всю связанную между собой информацию в единую систему, организует ее

так, как это для него оптимально при заданном составе и количестве ее элементов.

Разумеется, эти операции далеко не всегда осуществляются человеком осознанно. Не исключено даже, что чаще всего они бывают как раз неосознанными. Можно предположить, что именно с такой бессознательной реорганизацией и оптимизацией накопленной языковой информации связано то явление (наблюдавшееся, вероятно, многими из тех, кто глубоко изучал какой-либо иностранный язык), когда человек, возвращающийся к занятиям языком после сравнительно длительного (в несколько недель) перерыва, обнаруживает неожиданно большую свободу владения этим языком, чем непосредственно перед перерывом, после предшествовавшего ему периода интенсивного обучения.

Однако независимо от степени осознанности операций, происходящих при такой «самоорганизации» учащимся усваиваемых им данных, важно то, что такие операции носят активный характер и этим принципиально отличаются от простого запоминания – «зубрежки» – готовой информации.

Представляется, что и в системе АП было бы полезно предусматривать некоторый аналог таких активных процессов «усвоения» поступающей информации – т.е. включать в систему определенные средства самоорганизации и тем самым автоматической оптимизации в ней как самих вводимых лингвистических данных (подготовленных лингвистами фрагментов формализованного описания входного и выходного языков, различных переводных соответствий и трансформаций и т.п.), так и порядка вызова и применения этих данных при обработке текстов.

Попытаемся показать, какие возможности может дать в этом плане сущностный подход к лингвистическому описанию¹ в той его модификации, которая принята в рамках работ по созданию второй очереди экспериментального комплекса ЯРАП для исследований по японско-русскому автоматическому переводу в Институте востоковедения РАН².

2. Общие принципы организации лингвистического обеспечения в системе ЯРАП

В самом общем плане лингвистическое обеспечение системы ЯРАП состоит из описательной части, содержащей декларативную информацию о свойствах языковых единиц и их классов, и дина-

мической (операционной, процедурной) части, эксплицирующей лингвистические механизмы функционирования этой описательной информации при обработке текстов.

Все основные сведения о каждом из двух рассматриваемых языков – японском и русском, – т.е. прежде всего сведения о составе и структуре допустимых в них единиц, а также о существующих между ними переводных соответствиях, представляются при нашем подходе в описательной части лингвистического обеспечения. Организуются они таким образом, что для каждой языковой единицы, которая считается в рамках нашей задачи далее неразложимой (т.е. с точки зрения своей собственной сочетаемости и перевода является идиоматичной относительно своих возможных составных компонентов), задается исчисление всех допустимых для нее в тех или иных контекстных условиях способов ее сочетания с другими единицами при функционировании в составе более крупных единиц языка.

Поскольку перевод, представляя собой коммуникативную деятельность типа «текст↔текст», предполагает работу с естественно-языковыми текстами прежде всего на уровне собственно знакового плана языка³, в рамках данной коммуникативной задачи наиболее существенным является исчисление единиц и контекстов именно данного плана. Таким образом, для каждой единицы этого плана должны быть определены все собственно знаковые структуры, в составе которых она может употребляться, и те ограничения, которые налагаются при этом на саму эту единицу и на ее контекст. В первой очереди рассматриваемой системы ЯРАП/1 такое исчисление было ограничено в основном структурами лексико-морфологического и морфолого-синтаксического языковых ярусов⁴. При разработке ее второй очереди ЯРАП/2 основное внимание предполагается уделять структурам следующего, семантико-синтаксического яруса.

В рамках интересующего нас варианта сущностного подхода исчисление таких структур строится целиком на основе расширенного понятия **структурной валентности** языковой сущности⁵. Согласно определению этого понятия, как лексические, так и грамматические компоненты в составе словоформы способны обладать в собственно знаковом плане языка тем или иным специфическим для них набором структурных валентностей семантико-синтаксического характера, которые в сумме (с учетом морфологической структуры конкретной словоформы) образуют модель управления (= набор семантико-синтаксических структурных валентностей) самой словоформы.

В терминах таких валентностей, приписанных отдельным лексемам и морфологическим характеристикам слов, а также определяемых этими валентностями типов отношений зависимости, способов поверхностного оформления соответствующих отношений при их реализации в тексте и характеризующих их при этом семантических ограничений может быть описана любая языковая структура собственно знакового плана языка.

В системе перевода, помимо представленного указанным способом описания сочетаемости, для каждой единицы задается также перечень всех возможных для нее в разных контекстах переводных эквивалентов (в том числе трансформационных), за исключением тех, которые должны формироваться уже в пределах выходного (русского) языка с помощью его внутриязыковых перифрастических средств.

Правила, в форме которых строится требуемое исчисление, в общем случае могут быть сведены к логическим утверждениям типа:

(1) Если X , то Y ,

где X – лингвистически осмысленный фрагмент семантико-синтаксической структуры текста, содержащий рассматриваемую языковую сущность (скажем, лексему или морфологическую характеристику), Y – элементы и характеристики того же или других фрагментов структуры текста, которые могут или должны обнаруживаться в ней в присутствии фрагмента X .

Подчеркнем, во избежание недоразумений, что речь идет именно об **исчислении** языковых структур и их характеристик, а не об алгоритме их обработки – утверждения приведенного типа задают чисто декларативную информацию и никак не определяют последовательность шагов при анализе или синтезе текстов. Это приходится подчеркивать в связи с тем, что в сходной форме – форме логической импликации – часто записываются как раз алгоритмы, что может вызывать непонимание со стороны специалистов в этой области, пытающихся интерпретировать подобные утверждения в более привычном для них алгоритмическом смысле. Эти утверждения, однако, отличаются от алгоритмических правил (скажем, от правил, определяющих процедуры анализа) своей логической направленностью. Именно: компонент X в них, как уже отмечено, – это фрагмент не выражения текста, а его собственно знаковой, семантико-синтаксической структуры. Т.е., если брать, скажем, процесс анализа, в виде данного компонента задается не подаваемая на вход этого процесса информация о поверхностном составе обрабатываемого текста, по которой может формироваться та или иная гипотеза

о его собственно знаковой структуре, но сама эта гипотеза. Иными словами, приведенное выше правило следует читать:

- (2) «Если предположить, что в данном тексте реализована структура X, то в его составе следует ожидать наличия элементов и характеристик Y».

В этом смысле в связи с подобными правилами точнее говорить, по-видимому, не об импликации, а о секвенции.

Сказанное не означает, что все конкретные правила должны представляться непосредственно в форме секвенции. Она характеризует только логическую структуру используемых правил, конкретные же средства их формализованного представления могут иметь самый разный формат и различную степень внутренней сложности.

Так, данной логической структуре соответствует по своему содержанию стандартная «табличная» запись моделей управления в теории «Смысл ↔ Текст». Например, запись в третьем столбце таблицы, задающей модель управления русского глагола *опасаться*⁶:

(3)

Места		
1	2	3
S _{им}	1) S _{род} 2) V _{inf} 3) что + t 4) как бы + (t)	за + S _{вин}

с данной точки зрения имеет смысл: «Если обрабатываемая единица E₀ употреблена в контексте словоформы, реализующей лексему *опасаться*, причем в структуре зависимостей этого контекста занимает позицию глубинного «слуги» данной словоформы, заполняющего место с номером 3 в модели управления ее лексемы, то от этой единицы E₀ ожидается наличие следующих свойств (или, иначе, на эту единицу E₀ налагаются следующие ограничения): она должна входить в грамматический класс «существительное», иметь морфологическую характеристику «винительный падеж» и вводиться (оформляться) предлогом *за*».

В системе ЯРАП/2 для описания русского языка в модуле русского синтеза RussLan используется скобочный формализм⁷. Но интерпретироваться он может вполне аналогичным образом. Например, запись вида:

- (4) D3(ВЛ[ЛК[в2/за2/под2/]])
D3(ВЛ[ФМ[предл/Н] СМ[*лок/*направл]])

в статье глагола *бросить* может быть прочитана: «Если обрабатываемая единица употреблена в контексте глагола *бросить*, причем

занимает в структуре зависимостей этого контекста позицию «слуги» D (*Dependent*) данного глагола, реализующего его 3-й актант (D3), то для содержащей эту единицу группы в качестве вводящей ее лексемы ВЛ ожидается одна из конкретных лексем (ЛК) *в2, за2, под2*. Если эти ожидания не выполняются, вводящей лексемой может быть одна из единиц с формальным признаком (ФМ) предлога (*предл*) или наречия (*Н*) и семантической характеристикой (СМ) локализации (**лок*) или направления (**направл*)».

Первое правило (с конкретным перечислением вводящих лексем) описывает контексты типа *бросить что-либо в корзину / за ширму / под стол*, второе (с указанием их формальных и семантических признаков) – контексты типа *бросить что-либо перед входом / позади крыльца / назад / вправо / наверх / сзади дома, сбоку от себя, слева от машины / в лесу / на дороге* и т.п. Записанная таким образом информация может быть использована и при анализе – для формирования гипотезы о соответствующей связи зависимости, и при синтезе – для верификации оформления этой связи.

В японской части системы ЯРАП/2 для записи аналогичной по своему логическому устройству информации о языковых структурах семантико-синтаксического яруса разработан еще более сложный формализм, в терминах которого могут задаваться соответствующие более детальные описания контекста рассматриваемых единиц. В нем, в частности, для каждой структурной валентности описываемой единицы могут различаться признаки, возможные в реализующих ее контекстах, – ПВозм, сведения о допустимых для этой валентности «прямых» вариантах реализации ПВР, сведения о ее «косвенных» вариантах реализации КВР⁸ и т.п.

В прямых вариантах реализации единица, заполняющая рассматриваемую валентность, предстает как непосредственный синтаксический спутник носителя этой валентности в структуре зависимостей (т.е. как его синтаксический «слуга» при нексусной реализации валентности и как «хозяин» – при конквизивной). Соответственно, в описании такого варианта реализации даются сведения о направлении зависимости, связывающей соответствующие две единицы, о порядке их следования в данном контексте, об их лексических, семантических и морфолого-синтаксических свойствах. При необходимости могут вводиться сведения и о других, периферических участниках описываемого контекста, с уточнением, если требуется, уже их признаков и позиций в его структуре.

В косвенных вариантах реализации валентностей определяется прежде всего тот участник контекста, который выступает в нем как

контекстный валентностный наследник КВН заглавной единицы статьи (это могут быть, например, предлоги и их аналоги, а также служебные, ползузнаменательные, а иногда и знаменательные слова других частей речи). Такая единица-наследник обычно конкретизируется с помощью лексических, реже морфолого-синтаксических и семантических ограничений, и для нее задается тип контекстно-валентностного соответствия – указание на то, какая валентность единицы-наследника используется в качестве КВ-представителя для описываемой валентности единицы-наследодателя. Как и в случае прямых вариантов, могут добавляться и сведения о периферических участниках описываемых структур.

Так, в статье глагола *оконау* «проводить; осуществлять; выполнять» часть сведений о контекстах, где этот глагол выступает в функции определения к своему второму актанту, может быть задана в следующем виде:

- (5) ИМЯ *оконау*₁
 А-2
 ПВозм-1
 Сем = *действие
 МС = Ш/субст
 ПВР
 ПВР-2 = C₀ <-ю2- Сп
 Структ-1:
 ЛокВозм-1 = C₀, Сп
 Структ-2 : ОО = П1
 ДВозм-1 = П1 <<- Сп
 ЛокВозм-1 = C₀, П1+, Сп

Приведенное описание можно прочесть следующим образом. «Если некоторая единица употреблена в контексте глагола *оконау*, являющегося заглавной единицей C₀ данной статьи, в качестве его структурного спутника Сп, реализующего его 2-ю актантную валентность (А-2), то во всех вариантах такого контекста ожидается, что этот спутник имеет следующие ограничения на его возможные признаки (ПВозм): семантический признак (Сем) – *действие, морфолого-синтаксические признаки (МС) – существительное Ш или субстантиватор субст».

Из прямых вариантов реализации (ПВР) данного актанта в приведенной записи дан только второй вариант (ПВР-2). Он отличается тем, что заглавный глагол *оконау* (C₀) непосредственно подчинен своему спутнику (Сп) по юнктивной (определяющей) связи зависимости с индексом ю2 (C₀ <-ю2- Сп). Контекстное окружение этой связи может иметь несколько вариантов структуры, из которых

здесь указаны два: Структ-1 и Структ-2. В случае Структ-1 единственное дополнительное ограничение касается возможной взаимной локализации ЛокВозм-1 заглавной единицы C_0 и ее спутника: C_0 стоит непосредственно слева от Сп, без каких-либо промежуточных элементов, как в примере: *идзэн ни оконатта*(C_0) *кэнкю*:(Сп) «ранее проводившиеся(C_0) исследования(Сп)». В случае Структ-2 в описываемом окружении ОО единицы C_0 имеются дополнительные периферические элементы Π_1 , а сама структура имеет два ограничения: на возможное положение этих дополнительных элементов в дереве зависимостей ДВозм-1 и на возможную взаимную локализацию участников структуры ЛокВозм-1. Именно: всякий дополнительный элемент Π_1 прямо или косвенно подчиняется спутнику Сп глагола C_0 ($\Pi_1 \ll - \text{Сп}$); заглавный глагол C_0 , как и в Структ-1, стоит слева от своего спутника Сп, но может быть отделен от Сп любым ненулевым числом (+) единиц, отвечающих требованиям к Π_1 ($C_0, \Pi_1+, \text{Сп}$). Пример: *12 гацу ни оконатта*(C_0) “*данкай*($\Pi_{1.1}$) *сэдай-но*($\Pi_{1.2}$) *конго-но*($\Pi_{1.3}$) *раифусутаиру*($\Pi_{1.4}$) *то*($\Pi_{1.5}$) *сумаи-ни*($\Pi_{1.6}$) *кансуру*($\Pi_{1.7}$) *тё:са*(Сп)” «проводившееся(C_0) в декабре “обследование(Сп) представителей поколения($\Pi_{1.2}$) бума рождаемости($\Pi_{1.1}$) относительно($\Pi_{1.7}$) их последующего($\Pi_{1.3}$) образа жизни($\Pi_{1.4}$) и($\Pi_{1.5}$) местожительства($\Pi_{1.6}$)”».

Легко видеть, что приведенное описание, как и предыдущие, имеет декларативный характер, обеспечивая исчисление описываемых контекстов.

3. Словарь и грамматика при сущностном подходе к лингвистическому обеспечению

Те правила рассматриваемого исчисления, которые специфичны для отдельных конкретных языковых единиц, указываются в словаре системы, те, которые являются общими для многих разных единиц, – в ее грамматике.

По существу, этот метод представления описательной части лингвистического обеспечения системы АП аналогичен обычному способу подачи лингвистических сведений при обучении языку (как иностранному, так и родному), а также интересующему нас искусству перевода. Как правило, операционные указания при таком обучении имеют лишь самый общий смысл и ограничиваются инструкциями типа: «Не начинайте переводить, пока не прочтете весь абзац до конца»; «Заголовок лучше переводить по окончании

перевода всего текста»; «Закончив перевод текста, перечитайте его с самого начала как самостоятельный текст» и т.п. Подавляющее же большинство конкретных лингвистических сведений предлагается учащимся именно в описательной форме, в виде перечня существующих в тех или иных контекстных условиях возможностей и запретов, например:

- (6) «Если японский глагол выступает в предложении в качестве определения к существительному, он не может быть употреблен в “вежливой” форме»;
- (7) «Между японским существительным и относящейся к нему “падежной” частицей иногда может вставляться пояснение к этому существительному, выделенное скобками или тире»;
- (8) «Английское слово *until* может употребляться как предлог и переводиться русским предлогом *до* только при условии, что управляющий им глагол имеет при себе отрицание».

При таком способе организации описательной части лингвистического обеспечения системы АП о возможностях ее самоорганизации можно говорить в двух аспектах. Во-первых, это возможности автоматического осуществления взаимной корректировки, а также при необходимости – перераспределения элементов информации, включаемых в два отдела требуемого лингвистического обеспечения: словарь и грамматику. Во-вторых, это возможности интеграции в составе рассматриваемого лингвистического обеспечения, без дезорганизации этим его функционирования, элементов информации, формально находящихся в противоречии друг с другом – скажем, двух утверждений об английской единице *until* (8) и (9):

- (9) «Если *until* употреблено как предлог и управляющий им глагол не имеет при себе отрицания, то этот глагол сам подчиняется такой цепочке глагольных единиц (в частном случае состоящей из одного глагола), которая образована единицами восприятия (типа *notice* “замечать”) эмоционального отношения (типа *want* “хотеть”, *like* “любить”), модальности (*be going to*, *intend* “намереваться”) или интеллектуального состояния (*think* “думать”, *expect* “ожидать”, *suppose* “предполагать”) и одна из этих глагольных единиц имеет при себе отрицание».

Обе названные группы операций самоорганизации лингвистического обеспечения системы АП, по-видимому, имеют свои аналоги в процессах, связанных с усвоением языкового материала при обучении.

Для первой из них косвенным подтверждением существования такого ее аналога может служить тот общеизвестный факт, что усвоение грамматики происходит наиболее успешно, когда оно осу-

ществляется параллельно с расширением словарного запаса учащегося.

Для второй группы операций о том же свидетельствует распространенность такой методики преподавания, когда сведения, сообщаемые на продвинутых этапах обучения, частично опровергают сведения начального этапа. Скажем, для английского языка в первый период обучения, как правило, сообщается, что в нем имеет место жесткий порядок слов и подлежащее всегда ставится перед сказуемым, а дополнение – после него. В дальнейшем же учащийся получает информацию о том, что, вообще говоря, допустимы и другие варианты порядка слов, в том числе вариант «дополнение – сказуемое – подлежащее», а также вариант «дополнение – подлежащее – сказуемое». Соотнесение между собой таких, частично противоречащих друг другу сведений и нейтрализация соответствующих противоречий при непосредственной работе с языком составляет, очевидно, органический элемент всего процесса обучения.

Таким образом, включение в систему обработки текстов данных двух групп операций по самоорганизации описательной части ее лингвистического обеспечения означало бы отражение в этой системе реально существующих аспектов моделируемого ею типа языковой деятельности человека.

В системе ЯРАП принципиальную возможность операций первой группы – автоматической проверки соответствия между словарем и грамматикой и, при необходимости, автоматического перераспределения между ними содержащейся в них информации – предполагается обеспечивать на основе общих принципов сущностного подхода к языку. Согласно этим принципам, грамматика трактуется как обобщение словаря, т.е. как массив, объединяющий те элементы словарной по своему содержанию информации, которые оказываются совпадающими для разных словарных единиц и которые поэтому в целях экономии есть смысл не повторять в статьях всех этих словарных единиц, а вынести в отдельное описание, оставив в исходных словарных статьях только соответствующую отсылку.

Если словарные единицы, характеризующиеся такими общими для всех них элементами информации, достаточно разнообразны, роль требуемой отсылки выполняет имя грамматического класса, в который они объединяются по признаку этой общности, а сама общая для них информация организуется в рамках грамматики как целостное описание, сопоставляемое данному грамматическому классу. Если же какие-то элементы информации совпадают для

единиц, обладающих также и сходством состава и структуры своего поверхностного выражения, отсылкой к общей для таких единиц информации может служить сама совпадающая часть их поверхностного выражения, для которой в этом случае формируется особая, «обобщающая» словарная статья, содержащая всю соответствующую информацию.

Если совпадающие фрагменты обнаруживаются, скажем, в описаниях всех лексем одной и той же вокабулы (это возможно, например, при полисемии, при словообразовании путем конверсии и т.д.), в качестве их обобщения может выступать сама вокабула. Ей может приписываться в этом случае самостоятельная статья, в которую и будут выноситься из статей отдельных лексем совпадающие для них фрагменты описания⁹.

Другой пример такого рода – это возможность частичного совпадения семантико-синтаксических свойств русских лексем, имеющих в своей морфологической структуре одинаковые префиксы. В русской части системы ЯРАП/2 для учета данной возможности предусмотрен особый тип словарных статей – статьи отдельных префиксов, где и задаются свойства, общие для соответствующих префиксальных образований. Приведем пример такой статьи для одного из значений приставки *вс* (*вз*), реализуемого, например, в лексемах *вскинуть*, *встащить*, *вздернуть*, *взбросить* и т.п.:

(10) $\text{Ы}_{\text{вс}+} + {}^{10}$
 2 преф
 ## VV(Pf(Uф[ФМ[св/нсв] СМ[*движение.*ед] Нф[ФМ[нп]]
 ВКМ[: % FS(св,св,)]
 ВКV[ZZ[*направл] GV(Вскинуть)]
 УДV[ZZ[*длит] VS[v_[СИ[маршрут]]]])

Статья может быть прочитана следующим образом: «Если верна гипотеза, что в рассматриваемом приставочном образовании представлен префикс *вс* в своем 2-м значении, то ожидается, что входящая в это образование бесприставочная лексема отвечает следующим условиям (*Vф*): она имеет формально-грамматический (*ФМ*) признак глагола совершенного (*св*) или несовершенного (*нсв*) вида и семантические (*СМ*) признаки **движение.*ед* (признак **ед* указывает здесь на однонаправленность движения, отличая, например, *ползти* от *ползть*); кроме того у нее не должно быть (*Нф*) формально-грамматического признака непереходности (*ФМ[нп]*). При выполнении указанных условий статья рассматриваемого префиксального образования строится из словарной статьи бес-

приставочной лексемы следующим образом. В части морфологии в исходную статью включаются (*BKM*) указания о принадлежности префиксального образования к совершенному виду (*св, св*) и о релевантности для него способа образования видовой пары, вводимого символом *%*. В части информации, имеющей отношение к семантико-синтаксическим валентностям, в статью включаются (*BKV*): в список признаков, задающих значение (*ZZ*) данной лексической единицы, – признак направленности (**направл*), в список релевантных для нее обобщающих статей (*GV*) – статья *Вскинуть*. При этом удаляются (*УДV*): из списка признаков, задающих значение (*ZZ*), – признак длительности **длит*; из списка структурных валентностей – валентности с любым номером (*v_*), для которых в зоне валентностной семантики (*VS*) дана семантическая интерпретация маршрута (*СИ[маршрут]*»).

Отметим, что статьи данного типа позволяют не только более компактно представлять лексические сведения о единицах, включенных в словарь системы, но и формировать предположительную информацию для слов, которые в этом словаре отсутствуют. В том числе приведенная префиксальная статья обеспечивает корректную обработку, например, следующего входного задания¹¹:

- (11) этот
 чудовище RD1[вспендюрить RG2[корзина] RG3[крыша]]
 вспендюрить прош RG1[чудовище] RG2[корзина] RG3[крыша]
 корзина RD2[вспендюрить RG1[чудовище] RG3[крыша]]
 крыша RD3[вспендюрить RG1[чудовище] RG2[корзина]]

Хотя в составе этого задания содержится отсутствующая в словаре лексема *вспендюрить*, по нему синтезируется вполне удовлетворительный текст: *это чудовище вспендюрило корзину на крышу*.

Обобщающие словарные статьи образуют своего рода «грамматику в словаре», так что грамматический и лексический компоненты описательной части лингвистического обеспечения системы не требуют при рассматриваемом подходе строгого разграничения, но характеризуются, напротив, способностью, так сказать, к взаимопроникновению. Строятся они, как и естественно ожидать при такой трактовке их соотношения, достаточно единообразно. И в лексических, и в грамматических описаниях используется одна и та же структурная схема распределения информации по отдельным зонам и разделам описания, а сведения, помещаемые в одноименные зоны и разделы как в словаре, так и в грамматике записываются на одном и том же формальном языке и с помощью правил одних и тех же форматов.

4. Два типа грамматических классов в системе ЯРАП

Как видно из предыдущего, грамматика при сущностном подходе к лингвистическому описанию предстает как совокупность описаний отдельных грамматических классов, и конкретный состав ее существенно зависит от того, какие классы слов считается целесообразным выделять при ее составлении.

Наиболее очевидными являются при этом классы того типа, который можно назвать *парадигматическим*. Они объединяют конкретные языковые единицы, взятые как объект самостоятельного описания, и формируются преимущественно по признаку (частичной) общности характерного для соответствующих единиц набора их свойств определенного яруса. В традиционном языкознании этому соответствуют такие средства грамматического описания, как части речи и аналогичные им более мелкие синтаксические классы и подклассы, с одной стороны, и лексико-семантические классы слов, с другой. В информацию к классам первого типа выносятся преимущественно сведения о формальных характеристиках входящих в эти классы единиц, в информацию к лексико-семантическим классам – сведения о соответствующих семантических свойствах. Могут выделяться и классы смешанного характера.

В лингвистическом обеспечении первой очереди описываемой системы ЯРАП/1 предусматривались только классы данного, парадигматического типа. Это морфолого-синтаксические классы японских единиц, в статьях которых задаются ограничения на линейную сочетаемость, общие для единиц того или иного класса, а также сведения о характерных для этих единиц межъязыковых преобразованиях, прежде всего позиционных¹². Например, для класса японских субстантиваторов дается статья:

- (12) < *120 субст
 *231 :зс
 *244 Ш/Ц/пост
 *800 субст >.

В ней указано, что слева от единицы, имеющей в зоне *120 признак *субст*, допускаются, согласно ограничениям из зоны *231, словоформы только одного типа – с глагольной морфологической характеристикой «заклучительный статус» (:зс). Справа от такой единицы, согласно ограничениям из зоны *244, недопустимо появление единиц, чьи лексемы входят в морфолого-синтаксические классы «существительное» (Ш), «арабская цифра» (Ц) или «постпо-

зитив» (*post*). При межъязыковой обработке, согласно сведениям из зоны *800, для единицы данного грамматического класса релевантна процедура с тем же заглавием, что и сам этот класс.

В разрабатываемой в настоящее время системе ЯРАП/2 помимо парадигматических грамматических классов допускается также выделение грамматических классов другого типа – **синтагматических**, или функциональных. В такие классы объединяются языковые единицы, рассматриваемые как элементы структурного контекста других единиц. Различие классов этих двух типов состоит в следующем. Для парадигматического класса релевантность сопоставленных ему лингвистических сведений для той или иной конкретной единицы определяется только ее собственной принадлежностью к данному классу. Для синтагматического же класса это зависит от структурной позиции, занимаемой данной единицей относительно какой-то другой единицы того же текста. В частности, при анализе текста релевантность для той или иной единицы таких синтагматических сведений может быть установлена для нее только после того, как сформирована гипотеза о наличии между этой и некоторой другой единицей текста соответствующего структурного отношения.

Одним из наиболее ярких примеров синтагматических грамматических классов могут служить классы, объединяющие такие языковые единицы и характеристики, которые, участвуя в оформлении структурных отношений, реализующих валентности других единиц, обуславливают наличие у результирующих словосочетаний определенных стандартных свойств. В терминах классов такого рода естественным образом интерпретируются такие традиционные грамматические понятия, как, например, подлежащее, дополнение, обстоятельство и другие члены предложения.

Так, русское подлежащее можно охарактеризовать с этих позиций как элемент семантико-синтаксической структуры предложения, определяемый относительно личного глагола как единица, заполняющая одну из валентностей этого глагола – первую или вторую, в зависимости от залога, – и представляющая собой с точки зрения поверхностного оформления либо имя в именительном падеже, либо инфинитив, либо придаточное предложение определенной структуры и т.д. Стандартные структурные свойства, которые задает такой элемент для соответствующего сочетания единиц и которые должны быть включены в описание грамматического класса «подлежащее», – это, например, существующие в языке правила расположения подлежащего относительно других актантов и сирконстантов

того же глагола; сведения о его возможном месте в коммуникативной структуре предложения; сведения о трансформационных потенциях сочетания подлежащего со своим сказуемым и т.д.

В модуле русского синтеза системы ЯРАП/2 для сведений такого рода выделен компонент грамматического описания, где в качестве заглавий статей выступают указания на роль описываемой единицы E_1 в системе зависимостей как «слуги» D (Dependent) или «хозяина» G (Governor) другой единицы E_2 , определяемой своей частью речи, с уточнением типа реализуемой при этом валентности: если это валентность единицы E_2 , она задается своим номером, если единицы E_1 – индексом юнкции j . Чтобы отличить такие грамматические заглавия от имен обычных лексем (в которых вполне допустимы латинские буквы, цифры и знаки препинания), они вводятся двойным твердым знаком и заканчиваются двойным плюсом. Приведем фрагменты двух таких статей: для дополнения, реализующего третий актанта глагола $D3(G)$ (пример (13)), и для определения к существительному (пример (14)):

(13) $\text{ъ}D3(G)++$
 $\#\# VV(D3(U\phi[FM[G]] VKM[_L:инф])$
 $D3(U\phi[FM[III/II:прич]] VKM[_L:дат-п]))$

Данный фрагмент статьи $D3(G)$ можно прочесть следующим образом: «Если рассматриваемая единица является синтаксическим “службой” D глагола G и выступает в системе зависимостей в качестве 3-го актанта ($D3$) этого глагола, то при условии ($U\phi$), что данная единица сама имеет формальный признак (FM) глагола (G), от нее ожидается, что ее морфологическая характеристика включает (VKM) признак инфинитива ($:инф$), а аналитический показатель управления для данной связи отсутствует – представляет собой “пустую лексему” ($_L$); если же рассматриваемая единица имеет формальный признак существительного (III), прилагательного (II) или причастия ($:прич$), то ожидается, что в ее морфологическую характеристику включен признак дательного падежа ($:дат-п$), опять-таки с “пустой лексемой” ($_L$) в качестве аналитического показателя управления». Первая часть данного утверждения может быть проиллюстрирована примером: *Родителей просят(G) зайти($D3$) в школу*, вторая – примером: *Дайте(G) детям($D3$) задание*.

(14) $\text{ъ}Dj(III)++$
 $\#\# VV(Dj(U\phi[FM [G] BA[2]] H\phi[FM[NoPASS]]$
 $VKM[:пасс.прич] CG[t0[род.чис.пад]])$

Приведенный фрагмент статьи $Dj(III)$ можно прочесть: «Если рассматриваемая единица является синтаксическим “службой” D

существительного *Щ* и подчинена ему по юнктивной связи *j* (т.е. выступает в системе зависимостей в качестве его определения), то при условии (*Уф*), что данная единица имеет формальный признак (*ФМ*) глагола *Г*, причем имеющаяся между ним и существительным юнктивная связь реализует 2-ю актантную валентность (*ВА[2]*) этого глагола, а сам глагол не имеет (*Нф*) формального признака недопустимости пассива (*NoPASS*), то от рассматриваемого глагола ожидается, что в его морфологическую характеристику включен (*ВКМ*) признак пассивного причастия (:пасс.прич) и имеет место его согласование (*СИ*) с определяемой им текстовой единицей (*mθ*) по роду, числу и падежу (*род.чис.пад*)». Пример: *выполняемое(Dj) задание(Щ)*.

Синтагматические грамматические классы предусматриваются и в японской части системы ЯРАП/2. Проиллюстрируем их применительно к фрагменту описания валентности А-2 японского глагола *оконау* в примере (5). Приведенные в этой записи сведения о возможных реализациях второй валентности данного глагола в определенных конструкциях справедливы не только для глагола *оконау*, но и для других японских переходных глаголов. Поэтому практически все содержание этого описания (исключая только семантические ограничения, которые зависят от лексико-семантических свойств конкретных глагольных лексем) может трактоваться как грамматическая информация, характеризующая синтагматический класс «определяемое». Более того, если индекс реализуемой валентности в данном описании заменить переменной, то эти сведения будут применимы и к другим актантным валентностям, которые допускают реализацию посредством определяемого. Поэтому для всех них вместо самих соответствующих вариантов реализации достаточно указывать только признак синтагматического класса «Определяемое», а развернутое описание требуемого варианта реализации давать в статье этого класса:

(15) ИМЯ Определяемое

А-і

ПВозм-1

МС = Щ/субст

ПВР

ПВР-2 = С0 <-юі- Сп

Структ-1:

ЛокВозм-1 = С0, Сп

Структ-2 : ОО = П1

ДВозм-1 = П1 <<- Сп

ЛокВозм-1 = С0, П1+, Сп

Заметим, что сам факт допустимости определяемого как варианта реализации той или иной валентности, в свою очередь, носит общеграмматический характер и тоже может быть вынесен в описание соответствующих классов, но на этот раз уже парадигматических. Например, в статью парадигматического класса «Глагол» может быть помещена информация:

- (16) ИМЯ Глагол
 А-1
 ПВР
 ПВР-1 == Подлежащее
 ПВР-2 == Определяемое

в статью парадигматического класса «Переходный Глагол» – информация:

- (17) ИМЯ Переходный Глагол
 ГраммКласс Глагол
 А-2
 ПВР
 ПВР-1 == Прямое Дополнение
 ПВР-2 == Определяемое

а в статье конкретного глагола останутся только ссылки на самые узкие из включающих его парадигматических грамматических классов (из которых по транзитивности можно будет извлечь ссылки на все более широкие классы данного типа) плюс информация, специфическая именно для данного глагола. В частности, для глагола *окнау* достаточно будет дать ссылку на класс «Переходный Глагол» и на семантические ограничения, налагаемые лексемой *окнау* на каждый из двух ее актантов:

- (18) ИМЯ окнау1
 ГраммКласс Переходный Глагол
 А-1
 ПВозм-1
 Сем = *человек/*лицо
 А-2
 ПВозм-1
 Сем = *действие

Поскольку в статье «Переходный Глагол» есть своя грамматическая ссылка на парадигматический класс «Глагол», для *окнау* тем самым становится релевантной информация из статей обоих этих классов, включая сведения о том, что каждый из двух его актантов может реализоваться, помимо других возможностей, во всех вариантах, представленных в статье синтагматического грамматического класса «Определяемое» (15).

Таким образом, грамматическое описание рассматриваемого типа в современной его версии объединяет в себе как грамматику частей речи, так и грамматику членов предложения. Такое расширение его сферы, однако, не меняет того уже указанного нами принципа, что описание грамматических классов любых типов строится по тому же образцу, что и описание любой конкретной лексической или морфологической единицы – т.е. что словарь и грамматика принципиально едины и в отношении структуры включаемых в них описаний, и в отношении используемых в них формальных средств.

Благодаря этому любое грамматическое правило может быть без каких-либо существенных перестроек в способе его представления перенесено в словарную статью той конкретной языковой единицы, для которой оно релевантно. И наоборот, при обнаружении в словарных статьях ряда единиц одинаковых правил или (в случае синтагматической грамматической классификации) частей правил эти правила или части правил могут быть заменены в исходных статьях сокращенным обозначением, понимаемым как имя соответствующего грамматического класса, и вынесены в грамматику в виде описания этого класса также без каких-либо изменений в способе их формулировки.

Такое прозрачное и единообразное соотношение словаря и грамматики и создает, как представляется, лингвистическую базу для выполнения принципа самоорганизации системы АП в данном ее аспекте, т.е. обеспечивает принципиальную возможность разработки средств автоматизированного пополнения и уточнения грамматики на основе поступающей в систему дополнительной лексической информации, а также автоматизированной проверки полноты и адекватности лексических словарных статей с учетом имеющихся грамматических описаний.

5. Принципы снятия конфликтов при расширении и уточнении лингвистической информации

Второй отмеченный нами аспект формирования описательной части лингвистического обеспечения системы АП, требующий включения в нее элементов самоорганизации, связан, как мы говорили, с возможностью поступления в нее на разных этапах ее разработки и от разных участников частично противоречивой информации. Эта возможность не только реальна, но для больших систем, в класс которых входят промышленные системы АП, практически неизбежна.

Поэтому целесообразно предусмотреть какие-то средства, которые позволяли бы системе работать и с такой, частично противоречивой информацией (речь не идет, разумеется, о грубых ошибках со стороны разработчиков, приводящих к искажению всего лингвистического описания).

В системе ЯРАП/1 в качестве таких средств используется прежде всего противопоставление трех типов ограничений: «разрешений», «запретов» и «исключений из запретов»¹³. При этом в каких бы из релевантных для данной единицы частей описания ни встретились такие ограничения, запреты действуют «сильнее», чем разрешения, исключения из запретов – «сильнее», чем запреты.

Например, японскому существительному в его морфолого-синтаксической грамматической статье *Щ* приписаны две зоны «разрешений» относительно его непосредственного правого контекста – *241 и *242:

- (19) < *120 Щ
 *241 дэ1
 *242 Щ/Ч/Ц/мод/пост/соч/ЗП >

Информация, указанная в этих зонах, означает, что непосредственно справа от единицы, входящей в класс *Щ*, можно ожидать появления только таких словоформ, которые имеют лексему с именем *дэ1* (зона *241) или с грамматическими признаками существительного (*Щ*), частицы (*Ч*), арабской цифры (*Ц*), модификатора (*мод*), постпозитива (*пост*), показателя сочинения (*соч*) или знака препинания (*ЗП*).

В класс существительного входит, в частности, японская лексема *бааи* «случай, обстоятельства», о чем свидетельствует наличие признака *Щ* в зоне *120 ее словарной статьи:

- (20) < *100 бааи1_1
 *110 [случай]/[обстоятельство:мн]
 *120 Щ.информ.адв
 *244 Ч
 *245 ни2/ва1 >;

Соответственно, для *бааи* должны действовать все разрешения, предусматриваемые в вышеприведенной статье *Щ*. Однако в статье данной лексемы имеются еще две зоны ограничений на правый контекст: зона запретов *244 и зона исключений из запретов *245. В зоне запретов указан признак частицы *Ч*, следовательно, *бааи* в отличие от стандартных японских существительных не допускает появления частиц непосредственно справа от себя. Но в зоне исключений из запретов при этом даны имена двух конкретных частиц:

ни2 и *ва1*, что уточняет имеющийся запрет: справа от *бааи* не допускаются никакие частицы, кроме *ни2* и *ва1*.

При разработке второй очереди данной системы ЯРАП/2 рассмотренный способ согласования между собой частично противоречащих друг другу сведений предполагается дополнить еще одним, который, как представляется, позволит обеспечивать нейтрализацию возможных конфликтов в разрабатываемом лингвистическом обеспечении более системным образом. Этим целям должен служить последовательный учет в записи включаемой в систему языковой информации того общеизвестного принципа, что в языке не так много правил и законов, носящих абсолютный характер. Чаще всего можно говорить лишь о большей или меньшей **нормативности** тех или иных выражений, конструкций, трансформаций и т.д., но не об их обязательности или недопустимости в абсолютном смысле.

Для отражения в лингвистическом обеспечении ЯРАП/2 этого свойства естественного языка предполагается систематически использовать такой тип информации, как оценки степени нормативности языковых структур и характеристик, представленных отдельными элементами лингвистического описания. Оценки эти должны быть основаны на статистике и могут служить для динамической градации имеющихся лингвистических сведений по степени их надежности и перспективности для решения той или иной конкретной задачи текстовой обработки¹⁴. Сами соответствующие сведения при этом сохраняют ту же основную логическую структуру, которая была представлена выше формулами (1) и (2). Она приобретает лишь несколько более сложный вид:

(21) Если X , то Y с оценкой P_{XY} / P_X ,

т.е.: «Если предположить, что в данном тексте реализована структура X , то в его составе можно ожидать наличия элементов и характеристик Y с оценкой нормативности P_{XY}/P_X , где P_X – общее число употреблений структуры X , P_{XY} – число контекстов, включающих обе структуры X и Y ».

Подчеркнем, что общий характер лингвистического описания от введения в него оценок нормативности не меняется: оно по-прежнему остается исчислением языковых структур и их характеристик, а не набором процедур их обработки при анализе или синтезе.

При дополнении используемых лингвистических сведений статистическими оценками предложенного вида отмеченная выше опасность появления в системе, по мере ее роста и совершенствования, частично противоречивой информации, теряет свою остроту.

Так, правила (8) и (9), которые были приведены выше для английского *until* в предложном употреблении, получают разные статистические оценки степени их надежности (определяемой величиной дроби P_{XY}/P_X) и сферы их применимости (определяемой знаменателем этой дроби P_X). Соответственно, нейтрализовать возникающее между этими правилами противоречие можно путем сравнения сопоставленных им оценок. Например, если два противоречащих друг другу правила имеют одну и ту же сферу применимости (одинаковый знаменатель P_X), но одно из них является при этом менее точным (имеет меньший числитель P_{XY}), естественно отказаться от этого менее точного правила в пользу более точного.

Поскольку такое сравнение в принципе может быть автоматизировано, постольку рассматриваемый подход к организации описательной части лингвистического обеспечения системы АП дает принципиальную возможность ее самоорганизации и в этом аспекте.

6. Функции оптимизации и самоорганизации применительно к процедурной части лингвистического обеспечения системы ЯРАП

Остается рассмотреть особенности организации и обусловленные ими возможности самоорганизации системы ЯРАП применительно к операционной, процедурной части ее лингвистического обеспечения. В своей операционной интерпретации принцип самоорганизации можно понимать в первую очередь как способность управляющего алгоритма варьировать состав описательной информации, привлекаемой при обработке текста, и порядок обращения к отдельным ее элементам в зависимости от параметров самого обрабатываемого текста.

При сущностном подходе, когда все особенности семантико-синтаксической структуры предложений и текстов в целом представляются, благодаря понятию семантико-синтаксической структурной валентности, через свойства образующих эти предложения и тексты слов (вернее, компонентов их морфологической структуры), одним из основных параметров, которые должен учитывать соответствующий алгоритм, является, очевидно, сам словарный состав обрабатываемого текста. Понятно, что применять при обработке текста следует только ту информацию, которая релевантна для имеющихся в нем слов.

При изложенном выше способе организации описательной части лингвистического обеспечения системы это означает, что речь

должна идти об информации, которая записана либо непосредственно в словарных статьях лексем и морфологических признаков, входящих в структуры обрабатываемых на том или ином этапе слов, либо в описаниях (имеющих, как уже указывалось, также словарную форму) тех грамматических классов, имена которых упоминаются в этих статьях.

Один из простейших способов организации алгоритма, управляющего процессом анализа, так, чтобы он использовал в каждом конкретном случае только такую, релевантную для этого конкретного случая описательно-лингвистическую информацию, состоит в том, чтобы перед началом обработки очередного фрагмента текста формировать так называемые «рабочие» словарно-грамматические статьи всех словоформ этого фрагмента текста, объединяющие всю релевантную для соответствующих словоформ языковую информацию. Для этого из словаря и грамматики должны прежде всего извлекаться словарные статьи лексем и морфологических характеристик, входящих в морфологическое представление каждой словоформы, а также описания грамматических классов, в которые, согласно перечню, имеющемуся в их словарных статьях, входят эти лексемы и морфологические характеристики. Далее для информации, записанной в одноименных зонах и разделах всех статей и описаний, соответствующих одной и той же словоформе, если в них нет на этот счет каких-либо специальных указаний, должно рассматриваться ее объединение. Существенной переформулировки этой информации или других ее преобразований, как явствует из предыдущего изложения, при таком ее объединении не требуется, так что практически оно происходит только в том смысле, что информация, содержащаяся в одноименных зонах и разделах разных статей, релевантных для одной и той же текстовой единицы, становится доступна на одних и тех же этапах обработки и применяется одинаковым способом.

Непосредственно в процессе обработки текста должна использоваться информация только из таких «рабочих» словарных статей. Сама эта обработка для случая анализа может быть организована, например, таким образом, что на основе сведений о моделях управления словоформ данного предложения – т.е. о совокупностях семантико-синтаксических структурных валентностей, приписанных в словаре и грамматике лексическим и морфологическим компонентам этих словоформ, – алгоритм строит все возможные гипотезы об отношениях между отдельными словоформами, способных реализовать каждую из рассматриваемых валентностей, после чего все

остальные лингвистические сведения используются для фильтрации полученного, заведомо избыточного набора гипотез.

Отметим, что даже при таком, достаточно прямолинейном способе организации операционной части системы автоматического перевода требование самоорганизации ее функционирования в смысле автоматического приспособления, адаптации управляющего алгоритма к специфическим особенностям обрабатываемого текста в некоторой степени выполняется, поскольку информация, не релевантная ни для одного из допустимых вариантов морфологического представления словоформ текста, при обработке этих словоформ не только не проверяется, но даже доступ к ней управляющим алгоритмом не предусматривается.

Тем не менее на практике для значительной части информации, которая таким управляющим алгоритмом вызывается как релевантная для того или иного текста, по окончании его обработки выясняется, что в конечном итоге она была для него «лишней» и только усложняла процесс обработки – приводила к формированию ложных, как оказывалось в дальнейшем, гипотез о его структуре (вернее, о структуре отдельных его фрагментов).

Чтобы по возможности уменьшить участие в обработке текстов такой «лишней» для них информации и тем самым усилить способность алгоритма обработки «приспосабливаться» к свойствам обрабатываемого текста, при организации операционной части второй очереди системы ЯРАП предполагается разработать такие процедуры вызова и использования описательно-лингвистической информации, являющейся с точки зрения изложенного подхода релевантной (а вернее, условно релевантной) для того или иного текста), которые будут последовательно учитывать приписанные отдельным элементам этой информации оценки степени ее нормативности. Например, управляющий алгоритм должен будет активизировать вначале только наиболее частотные из соответствующих элементов информации, а к более редким свойствам слов обращаться лишь при условии, что обращение к свойствам, характерным для них в большей мере, не дает удовлетворительных результатов – т.е. не позволяет построить достаточно нормативную семантико-синтаксическую структуру обрабатываемого текста.

Существенным представляется при этом то свойство «хорошего» текста, что в нем, как правило, не допускается скопление в одной и той же точке его структуры сразу нескольких конструкций, каждая из которых характеризуется, с точки зрения способа ее выражения в имеющемся контексте, малой нормативностью. Появление подобно-

го скопления в каком-либо из вариантов анализа некоторого текста должно рассматриваться как указание на меньшую перспективность этого варианта для получения приемлемого результата по сравнению с имеющимися альтернативными вариантами.

Один из возможных здесь подходов разработан А.В. Костыркиным, предложившим трактовать процедуру такого анализа как задачу удовлетворения ограничений (Constraint Satisfaction Problem) и использовать при решении этой задачи метод динамических возвратов М.Л. Гинсберга¹⁵.

В заключение хотелось бы отметить тот важный момент, что, поскольку сведения о лингвистической нормативности имеют при рассматриваемом подходе статистическую природу, в процессе работы системы они могут в принципе корректироваться автоматически путем накопления соответствующей статистики непосредственно по результатам произведенной обработки – возможно, с той оговоркой, что эти результаты первоначально должны быть подтверждены человеком.

Таким образом, возникает принципиальная возможность создания средств автоматической оптимизации функционирования системы обработки текстов, опирающейся на рассматриваемый подход, применительно к текстам различных типов. Разработка таких средств позволила бы, по-видимому, в какой-то мере смоделировать в рамках систем автоматической обработки текстов еще один аспект усвоения языковых знаний и навыков человеком – накопление им опыта в процессе проработки (в частности, перевода) реальных конкретных текстов.

Примечания

- ¹ Шаляпина З.М. Трехмерная стратификационная модель языка и его функционирования. К общей теории лингвистических моделей. М.: Вост. лит., 2007. Глава 5. С. 150-250.
- ² О комплексе ЯРАП см., например: Экспериментальный комплекс ЯРАП для лингвистических исследований в области японско-русского автоматического перевода: первая очередь / Авторы: З.М. Шаляпина, Л.С. Модина, М.И. Канович, В.И. Любченко, А.С. Панина, Н.И. Сенина, В.И. Сивцева, Е.С. Тарасова, И.М. Хайлова, О.А. Штернова: Рук. деп. в ИНИОН РАН 5.11.01 № 56804. М.: ИВ РАН, 2001. 404 с.; Шаляпина З.М., Костыркин А.В., Канович М.И., Модина Л.С., Панина А.С., Тарасова Е.С. Экспериментальный комплекс ЯРАП для исследований по японско-русскому автоматическому переводу: 2008 г. // Бюллетень Общества востоковедов РАН. Вып. 17: Труды меж-

- институтской научной конференции «Востоковедные чтения 2008»: Москва, 8-10 октября 2008 г. М.: ИВ РАН, 2010. С. 359-397.
- ³ См. об этом в работах: *Шаляпина З.М.* Трехмерная стратификационная модель ... Глава 7. С. 327-348; *Shalyapina Zoya M.* Understanding and Generation of Texts from the Standpoint of the Three-Dimensional Stratification Linguistic Framework // Understanding by Communication / Eds. E. Borisova, O. Souleimanova. Newcastle-upon-Tyne: Cambridge Scholars Publishing, 2013. P. 6-32.
- ⁴ См.: *Модина Л.С.*, *Шаляпина З.М.* Принципы анализа японских текстов при моделировании японской лексико-морфологической системы // Труды международного семинара «Диалог 1996» по компьютерной лингвистике и ее приложениям (Пушино 4-9 мая 1996 г.) М., 1996. С. 169-174; *Канович М.И.*, *Шаляпина З.М.* Русский морфологический и морфолого-синтаксический синтез по неполной и/или неточной входной информации // Русский язык как иностранный: лингвистические проблемы (к 20-летию кафедры современного русского языка). М.: ИКАР, 1997. С. 137-147.
- ⁵ *Шаляпина З.М.* Структурные валентности как универсальный инструмент описания языковой сочетаемости (при сущностном подходе к ее моделированию) // Московский лингвистический журнал, 2001, том 5, № 2. С. 35-84; *Шаляпина З.М.* Трехмерная стратификационная модель ... С. 150-154 и далее.
- ⁶ *Жолковский А.К.*, *Мельчук И.А.* О семантическом синтезе // Проблемы кибернетики. Вып. 19. М.: Наука, 1967. С. 195.
- ⁷ О модуле русского синтеза комплекса ЯРАП и используемых в нем формализмах см.: RUSSLAN: A System of Russian Language Generation // Investigations into Formal Slavic Linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages – FDSLIV held at Potsdam University, November 28-30, 2001. Part I. Linguistik International. Band 10. Frankfurt am Main et al.: Peter Lang Verlag, 2003. P. 385-403; *Канович М.И.*, *Шаляпина З.М.* Аппарат R-отсылки как универсальное средство синтаксического синтеза (на опыте разработки системы русского синтеза RussLan) // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции «Диалог 2006». М.: РГГУ, 2006. С. 207-213; *Канович М.И.*, *Шаляпина З.М.* Система русского синтеза RussLan в двух разных коммуникативных задачах // Понимание в коммуникации: человек в информационном пространстве. Т.1. Ярославль: ЯГПУ, 2012. С. 244-261.
- ⁸ О прямых и косвенных вариантах реализации структурных валентностей см.: *Шаляпина З.М.* Трехмерная стратификационная модель ..., Гл. 4. С. 109-111; *Шаляпина З.М.* Аналитические показатели управления в синтаксической структуре // Вестник РГГУ, Серия «Языкознание. Московский лингвистический журнал» 2008. № 10. М.: РГГУ. С. 11-57.
- ⁹ Целесообразность обобщений такого «внутрисловарного» типа отмечают, в частности, разработчики некоторых информационных систем. См., например: *Леонтьева Н.Н.* Об информационной системе словарей Машинного фонда русского языка // Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. С. 109-125.

- ¹⁰ Чтобы отделить обобщающие словарные статьи в лингвистическом обеспечении модуля RussLan от статей конкретных лексем, заглавия первых начинаются с символов, не используемых для обычной лексики, например, *bl, v, b*, знака подчеркивания, заглавных букв (как ниже для статьи *Вскинуть*) и т.п.
- ¹¹ Обозначения вида «X RD_i[Y]» и «X RG_i[Y]» задают синтаксическую роль единицы X по отношению к Y в структуре предложения: D_i – роль синтаксического «слуги», заполняющего валентность i единицы Y, G_i – роль синтаксического «хозяина», чью валентность i заполняет единица Y. Подробнее см.: *Канович М.И., Шаляпина З.М.* Аппарат R-отсылок ...
- ¹² *Модина Л.С., Шаляпина З.М.* Лексико-грамматическая классификация японских лингвистических сущностей и ее функции в модели японско-русского автоматического перевода ЯРАП // IV Международная конференция по языкам Дальнего Востока, Юго-Восточной Азии и Западной Африки. Часть I. М., 1997. С. 142-150.
- ¹³ *Модина Л.С., Шаляпина З.М.* Принципы организации лингвистических знаний в объектно-ориентированной модели лексико-морфологической системы японского языка // Труды Международного семинара по компьютерной лингвистике и ее приложениям «Диалог 1995» (31 мая – 4 июня 1995, г. Казань), Казань, 1995. С. 198-205.
- ¹⁴ Ср. понятие «правдоподобия» результатов лингвистического анализа в работе: *Цейтин Г.С.* Система обработки текста с процедурным представлением словарной информации // Вопросы разработки прикладных систем. Новосибирск, 1979. С. 147-152.
- ¹⁵ *Костыркин А.В.* Применение технологий ограничений в синтаксическом анализе предложения // Труды международного семинара «Диалог 2000» по компьютерной лингвистике и ее приложениям. Т. 2. Протвино, 2000. С. 199-208; *Он же.* Семантико-синтаксический анализ естественно языковых текстов как задача удовлетворения ограничений // Материалы конференции «Понимание в коммуникации – 4». М.: МГПУ, 2009. С. 33.

Историография лингвистики

Н.В. Гурьян

Об особенностях организации и подачи материала в древнекитайском словаре «Фанъянь»

Статья посвящена анализу и описанию древнейшего китайского словаря диалектной лексики. Особое внимание уделяется особенностям организации и подачи материала в словаре. Автор также затрагивает вопрос преемственности между «Фанъянем» и его предшественником «Эръя».

Ключевые слова: «Фанъянь», «Эръя», структура словаря, семантические категории, формат словарных статей (глосс).

Как известно, обычай собирать диалектизмы в Китае сложился еще в глубокой древности и нашел отражение в различных памятниках¹. Знаменитый словарь «Фанъянь» 方言 («Местные слова»), составленный на рубеже новой эры Ян Сюном 揚雄 (53 г. до н.э. – 18 г. н.э.), государственным деятелем, литератором и филологом эпохи Хань, является первой специальной диалектной работой в истории китайского языкознания. Более того, по признанию многих современных исследователей², «Фанъянь» считается первым словарем диалектной лексики в истории мировой лексикографии. Благодаря тому, что этот памятник сохранился до наших дней, он позволяет судить о лингвистических воззрениях своего составителя.

История китайских словарей ведет свой отсчет с первого тезауруса «Эръя» 爾雅 (III в. до н.э.), технические достижения которого в том или ином виде были восприняты лексикографами последующих поколений, в том числе и Ян Сюном. В этой связи анализ и описание структурных особенностей «Фанъяня» в настоящей статье ведется в сравнении с «Эръя».

1. Структура и содержание «Фанъяня» (в сравнении с «Эръя»)

Из переписки Ян Сюна с Лю Синем и предисловия Го Пу (IV в.) к его комментарию на «Фанъянь» следует, что первоначальная версия словаря состояла из 15 цзюаней. Вариант, дошедший до нашего времени, представлен в 13 цзюанях. В предисловии ханьского филолога Ин Шао 應劭 к «Фэнсу тунъи» 風俗通義 (175 г.) сказано, что «Фанъянь» содержит 9000 иероглифов, в современном же его варианте насчитывается 11900 с лишним иероглифов³. Следовательно, около 2900 знаков являются результатом более поздних дополнений⁴. Кан Цзяньчан 康建常 полагает, что эти дополнения могли быть сделаны после Ин Шао (II в.), но до Го Пу 郭璞 (IV в.)⁵. Количество глосс в «Фанъяне» составляет 675⁶ (по другим данным, 669⁷, 658⁸).

Если сравнить эти показатели с количественными характеристиками первого китайского тезауруса «Эръя», то по общему иероглифическому объему словаря примерно равноценны⁹, а вот по числу глосс расхождение существенны: в «Эръя» это число составляет чуть более двух тысяч (по разным оценкам, цифры колеблются: 2019, 2091, 2047), что примерно в три раза превышает количество глосс в словаре Ян Сюна. Последний факт говорит о том, что в «Фанъяне» на отдельную глоссу приходится больше информации, чем в «Эръя».

В литературе прочно утвердилось мнение о том, что Ян Сюн составил свой словарь по образцу «Эръя», переняв его систему аранжировки материала. Как пишет Ян Шуда¹⁰, прообраз диалектных толкований Ян Сюн мог увидеть на примере следующих глосс «Эръя» из главы «Птицы»:

«爾雅·釋鳥»:

伊洛而南，素質五采皆備成章曰翬。К югу от [рек] И и Ло [фазанов] с цветными перьями в белом оперении называют хуэй [hui].

江淮而南，青質五采皆備成章曰鷩。К югу от [рек] Янцзы¹¹ и Хуай [фазанов] с цветными перьями в синем оперении называются яо [yao].

南方曰鷩，東方曰鷩，北方曰鷩，西方曰鷩。На юге [фазанов] называют чоу [chou], на востоке называют цзы [zi], на севере называют си [xi], на западе называются цзунь [zun].

К слову, в «Эръя» встречаются и другие единичные примеры указаний на диалектные варианты лексики, в частности, в главе «Звери»:

«爾雅·釋獸»:

鼠，鼠身長須而賊，秦人謂之小驢。Цзюй [крот]: тело [как у] крысы, длинные усы, вредит, циньцы называют его слолюй.

Пу Чжичжэнь 濮之珍 считает, что Ян Сюань использовал «Эръя» не только в качестве образца для подражания, но и как исходный список единиц описания¹². Другими словами, «Эръя» послужил эмпирической базой для «Фанъяня».

Однако, как показал проведенный анализ, между словарями есть и заметные различия. Так, по сравнению с «Эръя», в «Фанъяне» сокращен круг и объем специальной (терминологической) лексики:

1) из числа прежних (тех, что представлены в «Эръя») не обнаруживаются такие семантические категории, как родственные отношения, астрономия и календарь («Небо»), музыка и музыкальные инструменты (см. таблицу 1);

2) жилище, география, ландшафт («Земля», «Холмы», «Воды»), флора («Травы», «Деревья») представлены единичными глоссами в составе разных глав;

3) сократился охват представителей животного мира – нет категории «рыбы»; животные не подразделяются на диких и домашних; звери и птицы помещены в одну главу; так же, как и в «Эръя», отдельная глава посвящена насекомым.

Неспециальная лексика преобладает в количественном отношении, что подтверждается данными китайских исследователей¹³.

Сравнение архитектоники двух словарей показывает, что, если текст «Эръя» отчетливо распадается на две части: в первой (начальные три главы) толкуется абстрактная лексика, во второй (остальные главы) толкуются термины, то в «Фанъяне» не прослеживается такого деления – «абстрактные» и «терминологические» главы чередуются в тексте словаря.

Наиболее очевидна тематика пяти «терминологических» глав (см. таблицу 2)¹⁴, в остальных восьми главах помещены обычные слова, хотя, как станет понятно ниже, это не всегда абстрактная лексика.

Обратимся к содержанию разделов «Фанъяня».

Цзюяни 1, 2. Собрана абстрактная лексика без «терминологических вкраплений». Оформление глосс практически единообразно (за небольшим исключением): на входе цепочка-перечисление лексических единиц (далее – ЛЕ) – синонимический ряд, который, как правило, толкуется посредством общепонятного слова; далее следуют пояснения относительно территориального распространения синонимов из списка.

Таблица 1. Семантическая структура «Эръя» и «Фанъянь»

«ЭРЪЯ»		Глава, №	«ФАНЪЯНЬ»
Абстрактная лексика	Толкования (zu)	I.	Обычные слова
	Речи (янъ)	II.	Обычные слова
	Почтения (сюнь)	III.	Обычные слова (частично в отношении к человеку)
Терминологическая лексика	Родство	IV.	Одежда
	Жилище	V.	Утварь, предметы обихода, с.-х. инвентарь
	Утварь	VI.	Обычные слова
	Музыка	VII.	Обычные слова
	Небо	VIII.	Животные и птицы
	Земля	IX.	Оружие, транспортные средства (повозки, лодки)
	Холмы	X.	Обычные слова
	Горы	XI.	Насекомые
	Вода	XII.	Обычные слова
	Травы	XIII.	Обычные слова
	Деревья	XIV.	
	Насекомые, гады	XV.	
	Рыбы	XVI.	
	Птицы	XVII.	
Звери	XVIII.		
Домашние животные	XIX.		

Таблица 2. Тематика глав «Фанъянь»

Семантические категории		Цзюань, №
普通詞語 (“釋話”、 “釋言”) Обычные слова (архаизмы, диалектизмы)		1, 2, 3, 6, 7, 10, 12, 13
專有名詞 (釋名物) Термины	釋服制 Одежда	4
	釋器物 Утварь и приспособления	5
	釋獸 Животные	8
	釋器 Оружие, повозки, лодки	9
	釋蟲 Насекомые	11

卷一: 烈、桢, 餘也。陳鄭之間曰桢, 晉衛之間曰烈, 秦晉之間曰肄, 或曰烈。Ле, э [означають] юй [остаток, излишек]. В [царствах] Чэнь и Чжэн говорят э, в Цзинь и Вэй говорят ле, в Цинь и Цзинь говорят и или ле.

Иногда общепонятность тех или иных слов или их распространенность на определенной территории оговаривается с использованием специальной терминологии:

卷一: 悼、怒、悴、愁, 傷也。自¹⁵關而東汝潁陳楚之間通語也。汝謂之怒, 秦謂之悼, 宋謂之悴, 楚潁之間謂之愁。Дао, ни, цуй, инь [означають] шан [убиваться, сокрушаться, горевать, печалиться, скорбеть]. Слова, распространенные на территории к востоку от заставы [Ханьгу], в районе [рек] Жу и Ин, в [царствах] Чэнь и Чу. В районе [реки] Жу говорят ни, в Цинь говорят дао, в Сун говорят цуй, в Чу и в районе [реки] Ин говорят инь.

卷一: 憊、憊、憐、牟, 愛也。韓鄭曰憊, 晉衛曰憊, 汝潁之間曰憐, 宋魯之間曰牟, 或曰憐。憐, 通語也。У, янь, лян, моу [означають] ай [любить]. В [царствах] Хань и Чжэн говорят у, в Цзинь и Вэй говорят янь, в районе [рек] Жу и Ин говорят лян. В Сун и Лу говорят моу или лян. Лян – общепонятное слово.

卷一: 嫁、逝、徂、適, 往也。自家而出謂之嫁, 由女而出為嫁也。逝, 秦晉語也。徂, 齊語也。適, 宋魯語也。往, 凡語也。Цзя, ши, цу, ши [означають] ван [уходить]. Уходит из семьи называется цзя. Подобно тому, как женщина выходит замуж, называют это цзя. Ши – циньское и цзиньское слово. Цу – циское слово. Ши – сунское и луское слово. Ван – общераспространённое (обычное) выражение.

卷二: 鈔、嫖, 好也。青徐海岱之間曰鈔, 或謂之嫖。好, 凡通語也。Цяо, ляо [означають] хао [хороший]. В Цин[чжоу] и Сюй[чжоу], в Хайдай говорят цяо, иногда говорят ляо. Хао – общепринятое выражение.

Зачастую Ян Сюн указывает на ареальную принадлежность единиц, не включенных в предшествующий синонимический ряд, к примеру:

卷一: 度、儼, 慧也。秦謂之謾, 晉謂之懸, 宋楚之間謂之健, 楚或謂之謫。自關而東趙魏之間謂之黠, 或謂之鬼。Цянь, сюань [означають] хуэй [умный, сообразительный,мышленный]. В Цинь говорят мань, в Цзинь говорят май, в [царствах] Сун и Чу говорят цзе, в Чу также говорят то. К востоку от заставы [Ханьгу], в [царствах] Чжао и Вэй говорят ся или гуй.

Некоторые глоссы содержат пояснения относительно семантических различий между синонимами. Например:

卷一：鬱悠、懷、怒、惟、慮、願、念、靖、慎、思也。晉宋衛魯之間謂之鬱悠。惟，凡思也；慮，謀思也；願，欲思也；念，常思也。東齊海岱之間曰靖；秦晉或曰慎，凡思之貌亦曰慎，或曰怒。Юйю, хуай, ни, вэй, люй, юань, нянь, цзин, шэнь [означают] сы [думать; вспоминать]. В [царствах] Цзинь, Сун, Вэй и Лу говорят юй-ю. Вэй – думать вообще, люй – обдумывать / рассчитывать; юань – желать / чаять; нянь – постоянно думать / вспоминать. В Восточном Ци и Хайдай говорят цзин, в [царствах] Цинь и Цзинь говорят шэнь, задумчивый / размышляющий вид также называют шэнь или ни.

Можно найти и более детализированные толкования к отдельным словам:

卷一：啜、啼、杓、怛、痛也。凡哀泣而不止曰啜，哀而不泣曰啼。於方：則楚言哀曰啼，燕之外鄙，朝鮮洌水之間，少兒泣而不止曰啜。自關而西秦晉之間，凡大人少兒泣而不止謂之啜，哭極音絕亦謂之啜。平原謂啼極無聲謂之啜啉。楚謂之噉咷，齊宋之間謂之啉，或謂之怒。 Сюань, си, чжо, да [означают] тун [страдать, мучиться; сожалеть, скорбеть]. Непрекращающийся горестный плач называют сюань. Горевать, но без слез, называют си [всхлипывать; тяжело вздыхать; скорбеть; сетовать, горевать]. В местностях: так, в Чу скорбь / горе называют си. На окраинах [царства] Янь, в Корее, в районе [реки] Лешуй о детях, которые беспрестанно плачут, говорят сюань [кричать, плакать]. К западу от заставы [Ханьгу], в [царствах] Цинь и Цзинь всякий непрерывающийся плач взрослых или детей называют цянь [безудержные рыдания]. Надрывный плач, когда прерывается голос, тоже называют цянь. В [округе] Пинъюань сильные и беззвучные рыдания называют цяньлан. В [царстве] Чу называют цяютяо, в [царствах] Ци и Сун называют инь или ни.

卷一：慎、濟、瞻、怒、溼、桓，憂也。宋衛或謂之慎，或曰瞻。陳楚或曰溼，或曰濟。自關而西秦晉之間或曰怒，或曰溼。自關而西秦晉之間，凡志而不得，欲而不獲，高而有墜，得而中亡，謂之溼，或謂之怒。 Шэнь, цзи, цянь, ни, ши, хуань [означают] ю [печалиться, грустить; заботиться, беспокоиться]. В [царствах] Сун и Вэй говорят шэнь или цянь, в Чэнь и Чу говорят ши или цзи. К западу от заставы [Ханьгу], в [царствах] Цинь и Цзинь говорят ни или ши. К западу от заставы [Ханьгу], в [царствах] Цинь и Цзинь [всякий раз, когда] стремишься и не получаешь, жаждешь и не выходит, [находишься] высоко и падаешь, обретаешь и затем лишаешься, говорят ши или ни.

В ряде случаев глоссы содержат ссылки на другие источники и цитаты из них в качестве иллюстраций к толкованиям:

卷二：滕、厯，豐也。自關而西秦晉之間，凡大貌謂之滕，或謂之厯；豐，其通語也。趙魏之郊燕之北鄙，凡大人謂之豐人。《燕記》曰：豐人杼首。杼首，長首也。楚謂之仔，燕謂之杼。燕趙之間言圍天謂之豐。 Мэн, ман/пан [означают] фэн [большой, крупный; огромный]. К западу от заставы [Ханьгу], в Цинь и Цзинь всё крупное называют мэн или ман/пан; фэн – общепонятное слово. В предместьях [царств] Чжао и Вэй, на северной окраине Янь всякого крупного человека называют фэнжэнь. В «Янь цзи» сказано: фэнжэнь чжэу шоу. Чжэу шоу – [это] чан шоу [удлиненная голова]. В Чу называют юй, в Янь называют чжэу, в Янь и Чжао большое в объёме называют фэн.

Единично встречаются глоссы иной формы, в частности, когда дефиниция (толкование) представляет собой не эквивалент, а объяснение в нескольких словах:

卷一：墳，地大也。青幽之間凡土而高且大者謂之墳。Фэн [возвышенность, плато] – [это] большая земля. В [областях] Цин и Ю любая большая возвышенность называется фэн.

卷二：恒慨、蓼綏、羞繹、紛母，言既廣又大也。荊揚之間凡言廣大者謂之恒慨，東甌之間謂之蓼綏，或謂之羞繹、紛母。Хэнкай, саньсуй, сюи, фэньму – говорится [в отношении одновременно] широкого и большого. В Цзин и Ян всё обширное/крупное называют хэнкай, в Восточном Оу называют саньсуй, могут также называть сюи или фэньму.

При этом могут быть использованы и иные формулы построения глосс:

卷一：張小使大謂之廓，陳楚之間謂之摸。Расширять малое, [чтобы] сделать его большим называется ко. Ко [расширять, увеличивать] в [царствах] Чэнь и Чу называют мо.

Цзюань 3.

Материал данной главы неоднороден: толкуются как абстрактные, так и конкретные понятия. Отчасти применительно к человеку – его физиологическому состоянию (病 болеть, 愈 выздороавливать), эмоциональному (快 радоваться, 仇 ненавидеть, 怒 гневаться), по месту обитания (寄 жить/находить пристанище, 厝 (居) жить (где-л.), останавливаться, располагаться), поведению (詐 обманывать; лживый; 罪 преступление/злодеяние) и др. Здесь также толкуются некоторые наименования лица по признакам: отношения родства, социального положения и рода занятий, в том числе и стилистически маркированные (с отрицательной оценкой), а также собирательное понятие 民 простой народ/люди. Ниже приведем иллюстрации (порядок следования глосс не всегда отражает их расположение в тексте главы):

卷三: 陳楚之間凡人畱乳而雙產謂之釐孳, 秦晉之間謂之健子, 自關而東趙魏之間謂之學生。女謂之嫁子。 В [царствах] Чэнь и Чу всякая двойня (близнецы), родившаяся у человека или животного, называется лицзы; в [царствах] Цинь и Цзинь [двойню] называют лянцзы; к востоку от заставы [Ханьгу], в Чжао и Вэй называют луаньшэн. Девочек называют цзяцзы [сестры-двойняшки / близнецы].

卷三: 東齊之間登謂之倩。 В Восточном Ци сюй [зятя] называют цин.

卷三: 燕齊之間養馬者謂之娠。官婢女廝謂之娠。 В Янь и Ци того, кто ухаживает за лошадьми, называют чжэнь [конюх]. Гуаньби [казенные служанки, рабыни (преданные в рабство)] и нюйсы [служанки] называются чжэнь [прислуга].

卷三: 楚東海之間亭父謂之亭公。卒謂之弩父, 或謂之褚。 В районе [царства] Чу и [округа] Дунхай тинфу [начальника тина (станции); смотрителя] называют тингун. Цзу [солдат, воин] называется нуфу [переносчики самострелов] или чжэ.

卷三: 儻, 𧯂, 農夫之醜稱也。南楚凡罵庸賤謂之田儻。或謂之辟。辟, 商人醜稱也。 Тай, бо – непристойные (некрасивое) обозначения крестьянина. В Южном Чу, браня простолюдина и бедняка, называют [его] тяньтай [холоп, мужик], иногда говорят бо или би / пи. Би / пи – [это также] непристойное (некрасивое) обозначение торговца.

卷三: 褸裂、須捷、挾斯, 敗也。南楚凡人貧衣被醜弊謂之須捷。或謂之褸裂, 或謂之褸褸, 故《左傳》曰: 「葦路褸褸以扃山林」, 殆謂此也。或謂之挾斯。器物弊亦謂之挾斯。 Люйле [тряпье, лохмотья, рублище], суйцзе [изорванная / потрепанная одежда], сесы [рваный, обтрёпанный; подержанный, ветхий] – это бай [старый, испорченный]. В Южном Чу всякого нищего в лохмотьях называют суйцзе. В «Цзо чжуань» говорится: «В лохмотьях ехать на арбе поднимать целину», так говорят о дай [быть в опасности; измотаться, выбиться из сил]. Иногда называют сесы. Ветхую утварь (скарб) также называют сесы.

卷三: 逞、曉、佼、苦, 快也。自關而東或曰曉, 或曰逞。江淮陳楚之間曰逞, 宋鄭周洛韓魏之間曰苦, 東齊海岱之間曰佼, 自關而西曰快。 Чэн, сяо (曉), сяо (佼), ку – это куай [радостный; быть довольным]. К востоку от заставы [Ханьгу] говорят сяо (佼) или чэн. В районе [рек] Янцзы и Хуай, в [царствах] Чэнь и Чу говорят чэн. На территории [царств] Сун, Чжэнь, Чжоу, [реки] Ло, [царств] Хань и Вэй говорят сяо (佼). К западу от заставы [Ханьгу] говорят куай.

卷三: 凡飲藥傳藥而毒, 南楚之外謂之痢, 北燕朝鮮之間謂之癆, 東齊海岱之間謂之眠, 或謂之眩。自關而西謂之毒。痢, 痛也。

Отравиться, приняв яд или намазав его, за пределами [царства] Чу называют ла. В Северном Янь и в Корее говорят лао, в Восточном Ци и в Хайдай говорят мянь или сюань.

卷三: 瘼、癘, 病也。東齊海岱之間曰瘼, 或曰癘, 秦曰瘥。莫, фу – это бин [болезнь, недуг]. На территории Восточного Ци и Хайдай говорят мо, фу или чэнь.

卷三: 差、間、知, 愈也。南楚病愈者謂之差, 或謂之間, 或謂之知。知, 通語也。或謂之慧, 或謂之憐, 或謂之瘳, 或謂之蠲, 或謂之除。Чай, цзянь, чжи [означают] юй [выздороветь; пройти (о болезни)]. В Южном Чу когда выздоравливают, называют это чай, цзянь или чжи. Чжи – общепонятное слово. Иногда говорят хуэй, ляо, чоу, цзюань или чу.

卷三: 氓, 民也。Мэн – это минь [простой народ, люди].

Однако, наряду с глоссами, описывающими человека, в третьей главе обнаруживаются и другие семантические категории: несколько глосс посвящены растительному миру, имеется также по одной глоссе с толкованиями загона для свиней и водоемов (пример последнего приводится далее по тексту):

卷三: 蘇、芥, 草也。江淮南楚之間曰蘇, 自關而西或曰草, 或曰芥。南楚江湘之間謂之莽。蘇亦荏也。關之東西或謂之蘇, 或謂之荏。周鄭之間謂之公蕢。沅湘之南或謂之薺。其小者謂之薺菜。Су, цзе – это цао [трава]. В районе [рек] Янцзы и Хуай, к западу от заставы [Ханьгу] говорят или цао, или цзе. В южном Чу, в районе [рек] Янцзы и Хуай говорят ман. Су [по-другому] также [называют] жэнь [перилла]. К востоку и западу от заставы [Ханьгу] называют или су, или жэнь. В [царствах] Чжоу и Чжэн называют гунфэнь. К югу от [рек] Юань и Сян говорят ся, мелкую называют нянжоу.

卷三: 薑、蕘, 蕘菁也。陳楚[宋]之郊謂之薑, 魯齊之郊謂之蕘, 關之東西謂之蕘菁, 趙魏之郊謂之大芥, 其小者謂之辛芥, 或謂之幽芥; 其紫華者謂之蘆菔。東魯謂之菘。Фэн, жао – [это] уцзин [репа, турнепс]. В предместьях (на землях) [царств] Чэнь и Чу [Сун] называют фэн, на землях Ци и Лу называют уцзин. На землях Чжао и Вэй называют дацзе, мелкую называют синьцзе или юцзе; фиолетового / пурпурного цвета называют луфу [редька, редис].

卷三: 葎, 芘, 雞頭也。北燕謂之葎, 青徐淮泗之間謂之芘。南楚江湘之間謂之雞頭, 或謂之膈頭, 或謂之烏頭。И, цянь – это цзитоу [эвриала устрашающая]. В Северном Янь говорят и, в [областях] Цин и Суй, в районе [рек] Хуай и Сы говорят цянь. В Южном Чу, в районе [рек] Янцзы и Сян говорят цзитоу или яньтоу, иногда говорят утоу.

卷三: 凡草木刺人, 北燕朝鮮之間謂之策, 或謂之壯。自關而東或謂之梗, 或謂之劇。自關而西謂之刺。江湘之間謂之棘。Всякое колючее/ колющее растение в Северном Янь и в Корее называют це, иногда называют чжуан. К востоку от заставы [Ханьгу] называют гэн или гуй. К западу от заставы [Ханьгу] называют цы. В районе [рек] Янцзы и Сян называют цзи.

卷三: 豎, 圉也。Ли [хлев (для свиней), свинарник] – это хунь [свинарник; хлев].

Часть глосс содержит толкования абстрактной лексики (действия, признаки):

卷三: 擻、擻、拂、戎, 拔也。自關而西或曰拔, 或曰擻。自關而東江淮南楚之間或曰戎。東齊海岱之間曰擻。Я, чжо, фу, жун [означает] ба [вытаскивать, выдергивать]. К западу от заставы [Ханьгу] говорят ба или чжо. К востоку от заставы [Ханьгу], в районе [рек] Янцзы и Хуай, в Южном Чу говорят жун. На территории Восточного Ци и Хайдай говорят я.

卷三: 露, 敗也。Лу [означает] бай [разрушаться, приходит в упадок].

卷三: 別, 治也。Бе [означает] чжи [приводить в порядок].

卷三: 銛, 取也。Тянь [означает] цюй [выбирать].

卷三: 廋, 隱也。Соу [означает] инь [прятать, скрывать].

卷三: 屑, 潔也。Се [означает] цзе [чистый].

Из приведенных примеров видно также, что в главе собраны глоссы разные по своему оформлению.

Цзюань 4. Четвертая глава тематически однородна: сюда вошла лексика, номинирующая одежду, обувь, головные уборы, пряжу.

Цзюань 5. Материал пятой главы можно подразделить на несколько основных подгрупп: 1) посуда, сосуды, емкости, 2) предметы обихода, приспособления различного назначения, 3) орудия, 4) предметы и части мебели, предметы интерьера.

Цзюань 6, 7. Собрана абстрактная лексика.

Цзюань 8. Собрана лексика, номинирующая представителей фауны: животных и птиц.

Цзюань 9. В девятой главе толкуются термины из сфер «оружие», а также «транспорт» (включая повозки и лодки).

Цзюань 10, 12, 13. Смешанный лексический состав.

Цзюань 11. Данная глава полностью посвящена насекомым.

Ян Дуаньчжи 楊端誌 называет 16 семантических категорий, представленных в «Фангыане», определяя их по аналогии с заглавиями разделов «Эрья» «釋 + X» («Объяснение X»)¹⁸. Рассмотрим эти категории с иллюстрациями глосс:

(1) Толкования 釋詁

Под этой категорией Ян Дуаньчжи¹⁷ подразумевает глоссы, в которых архаизмы (существительные, глаголы, прилагательные) толкуются посредством современных Ян Сюю слов:

卷三: 氓, 民也。Мэн – [это] минь [народ].

卷三: 露, 敗也。Лу [означает] бай [разрушиться, прийти в упадок].

卷十二: 築婁, 匹也。Чжули [невестки, жёны братьев] – [это] ли [достойная пара; компаньон].

卷十二: 澈, 激, 清也。Пи, чэн [означают] цин [чистый].

(2) Речи 釋言

Сюда Ян Дуаньчжи отнёс толкования редких иероглифов и диалектных вариантов посредством общеизвестных знаков. Например:

卷一: 烈、枏、餘也。陳鄭之間曰枏, 晉衛之間曰烈, 秦晉之間曰隸, 或曰烈。Ле, не [означают] юй [остаток, излишек]. В [царствах] Чэнь и Чжэн говорят не, в Цзинь и Вэй – ле, в Цинь и Цзинь – и или ле.

卷三: 慰、塵、度, 尻也。江淮青徐之間曰慰, 東齊海岱之間或曰度, 或曰塵, 或曰躡。Вэй, чань, ду – [это] цзю [жить]. В районах [рек] Янцзы и Хуай, [областях] Цинь и Сюй говорят вэй; в районе Восточного Ци и Хайдай говорят ду или чань, могут говорить цзянь.

(3) Человек 釋人

Лексика, описывающая облик человека, его состояние, человеческие отношения:

卷十二: 娃、媼、窈、豔, 美也。吳楚衡淮之間曰娃, 南楚之外曰媼, 宋衛晉鄭之間曰豔, 陳楚周南之間曰窈。自關而西秦晉之間, 凡美色或謂之好, 或謂之窈。故吳有館娃之宮, 秦有 棹 娥之臺。秦晉之間貌謂之娥, 美狀為窈, 美色為豔, 美心為窈。Ва, то, тяо, янь [означают] мэй [красивый, прекрасный]. На территории [царств] У, Чу, [гор] Хэн, [реки] Хуай говорят ва. За пределами Южного Чу говорят то; в [царствах] Сун, Вэй, Цинь, Чжэн говорят янь; в Чэнь, У и Чжоунань говорят тяо. К западу от заставы [Ханьгу], в [царствах] Цинь и Цзинь всякую красоту называют хао или тяо. В давние времена в [царстве] У был дворец Гуаньва, в [царстве] Цинь была башня Циз. В Цинь и Цзинь о красивом лице говорят э, о красивой внешности/наружности говорят тяо, о красивом цвете (цвете лица; женской красоте) говорят янь; о прекрасной душе говорят яо.

(4) Одежда 釋衣

卷四: 幫, 陳魏之間謂之帔, 自關而東或謂之襪。Цюнь [юбка] – в

[царствах] Чэнь и Вэй называется пэй, к востоку от заставы [Ханьгу] называется бай.

卷四: 袴, 齊魯之間謂之襜, 或謂之襜。關西謂之袴。Ку [штаны, брюки] – в [царствах] Ци и Лу называют цянь или лун. К западу от заставы [Ханьгу] называются ку.

(5) Пицца 釋食

卷七: 脯、飪、亨、爛、糲、酋、酷, 熟也。自關而西秦晉之郊曰脯, 徐揚之間曰飪, 嵩嶽以南陳潁之間曰亨。自河以北趙魏之間火熟曰爛, 氣熟曰精, 久熟曰酋, 穀熟曰酷。熟, 其通語也。Эр, жэнь, пэн, лань, си, цю, ку [означают] шу [обработанный на огне; состряпанный, готовый; доваренный, дошедший]. К западу от заставы [Ханьгу], на землях Цинь и Цзинь говорят эр, в Сюй[чжоу] и Ян[чжоу] говорят жэнь, к югу от горы Сун, в районе [царства] Чэнь и [реки] Ин говорят пэн. К северу от [Хуан]хэ, в Северном Чжао и Вэй готовить на огне называется лань, готовить на пару называется си. Долго готовить (выдержанный) называется цю. Зерновые/хлеба готовить называют ку. Шу – общепонятное выражение.

卷十三: 餅謂之飪, 或謂之餛飩。Бин [лепешка, блин] называется то или чжанхунь.

卷十三: 餽謂之餼。餼謂之餼。餽謂之餼。凡餼謂之餼。自關而東陳楚宋衛之通語也。Син [солодовый сахар; сладости] называется чжанхуан [хворост (из клейкого риса и муки)]. И [патока] называется гай [лакомство]. Юй [сладости из бобовых] называют суй. Син называется тан [сахар; сладости]. Общепринятые слова к востоку от заставы [Ханьгу], в [царствах] Чэнь, Чу, Сун и Вэй.

(6) Жилище 釋宮

卷十三: 屋相謂之樞。Улюй [навес; карниз крыши, стреха] называется линь.

卷十三: 甌謂之甌。Мэн называется лю [копьевая балка крыши].

(7) Утварь 釋器

卷五: 盂, 宋楚魏之間或謂之盂。盂謂之盂, 或謂之鉢。盂謂之權, 孟謂之柯。海岱東齊北燕之間或謂之蓋。Юй [большая чашка] – в [царствах] Сун, Чу, Вэй называется вань [чашка; пиала]. Вань называется юй, может также называться дяожуй. Вань называется чжо, юй называется кэ. В Хайдай, Восточном Ци, Северном Лян может называться цюань.

卷五: 牀, 齊魯之間謂之簣, 陳楚之間或謂之第。其杠, 北燕朝鮮之間謂之樹, 自關而西秦晉之間謂之杠, 南楚之間謂之趙 (兆), 東齊海岱之間謂之梓。其上板, 衛之北郊趙魏之間謂之牒, 或曰牒。Чуан [кровать] – в Ци и Лу называется цзэ, в Чэнь и Чу – цзы. Её [передняя]

поперечина в Северном Янь и Корее называется шу, к востоку от заставы [Ханьгу], в Цинь и Цзинь – ган, в Южном Чу – чжао, в Восточном Ци и Хайдай называется шэнь. Её основание в северном предместье Вэй, в Чжао и Вэй называют де или бянь/мянь.

(8) Оружие 釋兵

卷九: 箭, 自關而東謂之矢, 江淮之間謂之鏃, 關西曰箭。Цзянь [стрела] – к востоку от заставы [Ханьгу] называется ши, в районе [рек] Янцзы и Хуай называется хоу, к западу от заставы [Ханьгу] называется цзянь.

卷九: 劍削, 自河而北燕趙之間謂之室, 自關而東或謂之廊, 或謂之削, 自關而西謂之鞞。Цзяньсяо [ножны меча] – к северу от [Хуан]хэ, в Янь, Чжао называется ши, к востоку от заставы [Ханьгу] также называется лан или сяо, к западу от заставы [Ханьгу] называется бин.

(9) Повозки 釋車

卷九: 輪, 韓楚之間謂之軹, 或謂之軹。關西謂之輓。Лунь [коле со] – в [царствах] Хань и Чу называется дай или ци. К западу от заставы [Ханьгу] называется цзун.

卷九: 輓, 楚衛之間謂之輓。Юань [оглобля] в [царствах] Чу и Вэй называется чжоу.

卷九: 箱謂之軹。Сян [кузов повозки] называется пай.

(10) Лодки 釋舟

卷九: 舟, 自關而西謂之船, 自關而東或謂之舟, 或謂之航。南楚江湘凡船大者謂之舸, 小舸謂之舫, 舫謂之舫, 小舫謂之舫, 舫長而薄者謂之舫, 短而深者謂之舫, 小而深者謂之櫓。東南丹陽會稽之間謂舫為櫓。Чжоу [лодка; судно; корабль] – к западу от заставы [Ханьгу] называется чуань, к востоку от заставы [Ханьгу] называется чжоу или хан. В Южном Чу, в [районе рек] Янцзы и Сян все крупные лодки/суда называются гэ; маленький гэ называется ча [ялик, челнок; лодочка]; ча называется маосу; маленький маосу называется тин; длинный и мелкий тин называется дай; короткий и глубокий называется бу; маленький и глубокий – цюн. На юго-востоке, в [округах] Даньян и Куайцзи ча называют ли.

(11) Воды 釋水

卷三: 汜、浼、濶、注、洿也。自關而東或曰注, 或曰汜。東齊海岱之間或曰浼, 或曰濶。Фань, мэи, янь, ва [означают] у [стоячие водоёмы; пруды и озёра]. К востоку от заставы [Ханьгу] говорят ва или фань. В Восточном Ци и Хайдай говорят мэи или янь.

卷十二: 水中可居為洲。三輔謂之淤, 蜀漢謂之巽。Обитаемая [суша] на воде (реке) называется чжоу [остров; большая отмель]. В [округе] Саньфу называется юй, в Шу и Хань называется би.

(12) Земля 釋土

卷一：墳，地大也。青幽之間凡土而高且大者謂之墳。Фэнь [возвышенность, плато] – [это] большая земля. В [областях] Цин и Ю любая большая возвышенность называется фэнь.

卷十三：冢，秦晉之間謂之墳，或謂之培，或謂之埡，或謂之采，或謂之垠，或謂之壠。自關而東謂之丘，小者謂之墳，大者謂之丘，凡葬而無墳謂之墓，所以墓謂之壠。Чжун [могильный холм, курган], в Цинь и Цзинь называется фэнь или лоу, также может называться юй, цай, лан или лун. К востоку от заставы [Ханьгу] называется цю; маленький называется лоу, большой называется цю. Всякое захоронение без холма/кургана называется му, при помощи му (墓) называют мо/му (壠).

(13) Травы 釋草

卷三：蘇、芥，草也。江淮南楚之間曰蘇，自關而西或曰草，或曰芥。南楚江湘之間謂之莽。蘇亦荏也。關之東西或謂之蘇，或謂之荏。周鄭之間謂之公蕒。沅湘之南或謂之善。其小者謂之藟菜。Су [перилла нанкинская], цзе [горчица пониклая] – это цао [травы]. В районах [рек] Янцзы и Хуай, в Южном Чу говорят су, к западу от заставы [Ханьгу] говорят цао или цзе. В Южном Чу, в районе [рек] Янцзы и Хуай называют ман [заросли трав]. Су – [это] также жэнь [перилла базиликковая]. К западу и востоку от заставы [Ханьгу] называют су или жэнь. В Чжоу и Чжэн называют гунфэнь. К югу от [рек] Юань и Сян говорят ся. Маленькую называют нянжоу.

卷三：薑、蕘，蕘菁也。陳楚之郊謂之薑，魯齊之郊謂之蕘，關之東西謂之蕘菁，趙魏之郊謂之大芥，其小者謂之辛芥，或謂之幽芥；其紫華者謂之蘆菔。東魯謂之菘蕘。Фэн/сун, жао – это уззин [репа, турнепс]. В Чу и Чэнь называют фэн, в Лу и Ци называют жао, к востоку и западу от заставы [Ханьгу] называют уззин, в Чжао и Вэй называют дацзе, мелкую называют синьцзе или юцзе; пурпурную/фиолетовую называют луфу [редька; редис]. В Восточном Лу называют лата.

卷三：凡草木刺人，北燕朝鮮之間謂之棗，或謂之壯。自關而東或謂之梗，或謂之劇。自關而西謂之刺。江湘之間謂之棘。Всякое колючее (колючее, колкое) растение в Северном Янь и в Корее называют цэ [шиш; колючка] или чжуан. К востоку от заставы [Ханьгу] называют гэн или гуй. К западу от заставы [Ханьгу] называют цы. В районе [рек] Янцзы и Сян называют цзи.

(14) Звери 釋獸

卷八：貔，陳楚江淮之間謂之豸，北燕朝鮮之間謂之豸，關西狸。П и [? лису; дикую кошку] в [царствах] Чэнь, Вэй и в районе [рек] Янцзы и Хуай называют лай, в Северной Янь и Корее называют пэй. К западу от заставы [Ханьгу] называют ли.

(15) Птицы 釋鳥

卷八：布穀，自關東西梁楚之間謂之結誥，周魏之間謂之擊穀，自關而西或謂之布穀。Бугу [жукушка] к востоку от заставы [Ханьгу], в [царствах] Лян и Чу называется цзеггао; в Чжоу и Вэй называется цзигу; к западу от заставы [Ханьгу] называется бугу.

卷八：野鳧，其小而好沒水中者，南楚之外謂之鷺鷥，大者謂之鶻蹶。[Что касается] ефу [дикой утки], маленькую и хорошо ныряющую в Южном Чу называют пити, крупную называют худи.

(16) Насекомые 釋蟲

卷十一：螳螂謂之髦，或謂之虻，或謂之蟀蟀。Танлан [богомолы] называют мао, также называют дин или мими.

卷十一：蠅，東齊謂之羊。陳楚之間謂之蠅。自關而西秦晉之間謂之蠅。Ин [муха] в Восточном Ци называется ян. В Чэнь и Чу называется ин. К западу от заставы [Ханьгу], в Цинь и Цзинь называется ин.

卷十一：馬蚻，北燕謂之蛆蝶。其大者謂之馬蚻。Масянь [кивсак] в Северной Янь называется цюйцюй. Крупный [кивсак] называется маю.

Анализируя приведенные выше семантические категории и иллюстрации к ним, можно заключить следующее. Во-первых, данный список не совпадает с количеством цзюаней и не отражает структурное деление материала на главы. Во-вторых, упомянутые семантические категории не равноценны по объему: некоторым из них в словаре посвящены более или менее обширные участки словаря вплоть до отдельных цзюаней (насекомые, звери и птицы, утварь, оружие и транспорт), другие же представлены минимально – единичными глоссами (вода, земля, травы).

Более наглядную картину дают конкретные цифры, отражающие количество слов той или иной семантической категории с указанием их ареальной принадлежности (см. таблицу 3)¹⁶:

Представленные данные позволяют судить, во-первых, о том, лексика каких ареалов в словаре доминирует количественно (это явно диалекты царств Цинь, Цзинь-Ань и царства Чу-Вэй), во-вторых, о количественном соотношении глосс той или иной семантической категории (в целом преобладает абстрактная лексика, на которую приходится чуть менее половины – 704 слова, далее по убыванию следуют категории «утварь» – 245, «человек» – 190, «одежда» – 75, «насекомые» – 72 и т. д.

Таблица 3. Соотношение диалектизмов различных семантических категорий¹⁹

№	Сем. категории Ареал	Слова	Человек	Одежда	Пища	Утварь	Оружие	Повозки	Лодки	Воды	Земля	Травы	Звери	Птицы	Насекомые	Итого
		言	人	衣	食	器	兵	車	舟	水	土	草	獸	鳥	蟲	
1	秦晉	129	31	16	6	40	9	6	2	2	6	5	5	15	21	293
2	周韓鄭	45	11	9	3	19	5	6	2	2	1	5	3	15	11	137
3	趙魏	26	9	12	6	44	1	3	0	0	0	4	1	6	5	117
4	衛宋	65	14	4	4	40	0	4	0	0	0	0	1	5	4	141
5	齊魯	64	10	5	4	11	0	3	0	0	1	3	0	3	3	107
6	東齊海	54	19	1	1	9	2	2	0	2	1	2	0	5	3	103
7	燕代	9	10	0	0	2	1	2	0	0	1	0	0	0	2	27
8	北燕朝	26	7	1	1	9	0	0	0	0	0	3	2	5	5	60
9	楚	139	29	15	7	49	5	9	1	0	0	2	4	5	8	274
10	南楚	89	32	12	0	10	4	1	1	0	0	8	5	0	6	168
11	南越	12	5	0	2	3	0	2	0	0	0	0	0	6	4	34
12	吳越	45	11	0	2	8	4	0	1	0	0	1	1	1	0	74
	Итого	704	190	75	36	245	31	38	7	6	10	33	22	66	72	1535

2. Формат глосс

Если говорить о формате глосс (словарных статей) «Фанъяня», можно назвать три основные разновидности, которые в своей типичной форме имеют следующую структуру:

1. На входе находится список синонимов, которые семантизируются посредством [общепонятного] эквивалента, далее к синонимам из перечня (а зачастую и к единицам сверх перечня) даются пояснения относительно их ареальной принадлежности, в некоторых случаях приводятся дополнительные толкования (примеры таких толкований приводились выше). Это наиболее распространенная разновидность глосс.

卷二：剽、蹶，猶也。秦晉之間曰猶，楚謂之剽，或曰蹶；楚鄭曰葛，或曰姑。Цзяо/чао, цзюэ [означают] куай [хитрый, лукавый; коварный]. В Цинь и Цзинь говорят куай, в Чу говорят цзяо/чао, иногда говорят цзюэ; в Чу и Чжэн говорят вэй, иногда говорят хуа.

2. На входе находится общепонятное слово, далее указывается территория и соответствующий диалектизм:

卷八：虎，陳魏宋楚之間或謂之李父，江淮南楚之間謂之李耳，或謂之於旄。自關東西或謂之伯都。Ху [тигра] в [царствах] Чэнь, Вэй, Сун и Чу называют лифу, в районе [рек] Янцизы и Хуай, Южном Чу называют лиэр и уту. К западу и востоку от заставы [Ханьгу] называют боду.

В расширенном формате эта разновидность глосс, кроме диалектных соответствий заглавного слова, может включать толкования некоторых смежных понятий:

卷八：豬，北燕朝鮮之間謂之豨，關東西或謂之彘，或謂之豕。南楚謂之豨。其子或謂之豚，或謂之豨，吳揚之間謂之豬子。其檻及磨曰槽。Чжу [свинью, кабана] в Северной Янь и Корее называют цзя. К востоку, западу от заставы [Ханьгу] иногда называют чжи или ши. В Южном Чу называют си. Её детеныша (поросенка) называют тунь или си, в У и Ян называют чжуцзы. Загон и [соломенную] подстилку для свиней называют цзэн.

Иногда на входе может находиться не слово, а описательное словосочетание:

卷五：飲馬橐，自關而西謂之掩囊，或謂之掩箠，或謂之樓箠。燕齊之間謂之帳。Мешок для кормления лошадей к западу от заставы [Ханьгу] называют яньнан, иногда называют яньдоу или лоудоу. В Янь и Ци называют чжэнь.

3. Глосса состоит только из поясняемого (неизвестного) и поясняющего элементов, без прочих толкований. В зависимости от формулы описания на входе может находиться как поясняемая единица (для формулы “А, В也” А – это [есть; то же самое, что и...; по-другому называется...] В), так и толкование (для формул с операторами 謂之 значит, имеет значение, 曰 говорят, называют, 為 есть [является], называется), например:

卷十三：珎，好也。珎，美也。Цзу [означает] хао [хороший]. Цзу [означает] мэи [красивый].

卷十三：芋，大也。Сюй [означает] да [большой].

卷十三：匕謂之匙。Би [большую ложку, черпак] называют чи.

卷十三：甍謂之甍。Мэн [коньковую балку крыши] называют лю.

3. Заключение

Итак, резюмируя сказанное в предыдущих разделах настоящей статьи, обозначим основные различия двух словарей на уровне макро- и микроструктуры:

1. Корпус «Фанъяня» не имеет такой строгой разбивки на семантические категории и четкой рубрикации, какую мы видим в «Эръя». В последнем она задается в заголовках разделов: «Родство», «Дворец», «Музыка», «Утварь», «Небо», «Земли», «Воды», «Деревья», «Травы», «Рыбы», «Звери», «Домашние животные» и др., главы же «Фанъяня» просто пронумерованы. В «Фанъяне» сокращен набор специальной (терминологической) лексики, соответственно, меньше и количество глав (13 против 19-ти в «Эръя»): отсутствуют отдельные главы, посвященные терминам родства, астрономии, географии, флоре. Таким образом, Ян Сюн отошел от той более или менее стройной идеологической системы, которая была продемонстрирована в первом словаре «Эръя». И, хотя считается, что расположение материала в «Фанъяне», как и в «Эръя», подчинено семантическому критерию, однако, очевиден он не для всех глав, поэтому принципы организации лексики еще требуют уточнения.

2. В сравнении с «Эръя», на уровне глосс толкования «Фанъяня» в целом более информативны, т.к. помимо объяснения значений, содержат информацию о диалектной принадлежности поясняемых единиц, часть материала снабжена дополнительными толкованиями, касающимися особенностей семантики лексических единиц. Правда, на этом фоне явно выделяются две последние главы «Фанъяня», в которых такая информация практически отсутствует. Объясняют это тем, что Ян Сюн не успел завершить свой труд, и часть материала осталась без диалектных и прочих пояснений.

Примечания

- ¹ Яхонтов С.Е. Древнекитайский язык. М.: Наука, 1965; Завьялова О.И. Большой мир китайского языка. М.: Вост. лит., 2010.
- ² См., напр., *Serruys P. L.-M. The Chinese Dialects of Han Time According to Fang Yen.* Berkeley and Los Angeles: University of California Press, 1959; *Huang Jianhua. Chinese and Western Metalexigraphy // Lexicography: Critical Concepts.* Vol. III. Lexicography, Metalexigraphy and Reference Science / Ed. R.R.K. Hartman. N.Y.: Routledge, 2003; *Yong H., Peng J. Chinese Lexicography: A History from 1046 BC to AD 1911;* 柳玉宏. 论《方言》的学术价值: 硕士学间论文. 宁夏大学, 2005.
- ³ 王力. 中国语言学史. 太原: 山西人民出版社, 1981. С. 21.
- ⁴ 康建常. 扬雄的语言观及其《方言》的价值 // 殷都学刊. 1991. № 1. С. 92 – 97; 黄志明. 《方言》母题研究: 硕士学位论文. 华中科技大学, 2004.
- ⁵ 康建常. 扬雄的语言观及其《方言》的价值. С. 92.
- ⁶ 濮之珍. 《方言》与《尔雅》的关系 // 学术月刊. 1957. № 12. С. 88.
- ⁷ 杨端志. 训诂学 / 殿焕先校订. 济南: 山东文艺出版社, 1985. С. 478.

- ⁸ *Yong H., Peng J. Chinese Lexicography: A History from 1046 BC to AD 1911. Oxford, N.Y.: Oxford University Press, 2008. P. 84.*
- ⁹ Относительно общего иероглифического объема «Эрья» данные источников расходятся в пределах от 10900 до 13113 иероглифов, подробнее см. *Гурьян Н.В. Первый китайский толковый словарь «Эрья»: историко-типологическая характеристика // Вестн. Моск. ун-та. Сер. 13. Востоковедение. 2011. № 1. С. 89.*
- ¹⁰ *杨树达. 积微居小学述林. 北京: 中华书局, 1983. С. 271–272.*
- ¹¹ Здесь и далее топоним 江 *Цзян [Река]* переводится как *[река] Янцзы.*
- ¹² *濮之珍. 《方言》与《尔雅》的关系; 濮之珍. 中国语言学史. 上海: 上海古籍出版社, 1987.*
- ¹³ *黄志明. 《方言》母题研究. С. 27–29.*
- ¹⁴ *Там же; 窦秀艳. 中国雅学史. 济南: 齐鲁书社, 2004. С. 72.*
- ¹⁵ Возможно, в начале этой фразы перед иероглифом 自 пропущен знак 傷, см. *方言校笺: 附索引 / 周祖谟校笺. 北京: 中华书局, 2004. С. 3.*
- ¹⁶ *杨端志. 训诂学. С. 482–486.*
- ¹⁷ *Там же.*
- ¹⁸ *Статистические данные взяты из работы Ван Чжицюнь см. 王智群. 《方言》与扬雄词汇学. 北京: 高等教育出版社, 2011. С. 96.*
- ¹⁹ *Из семантических категорий в таблице не представлены 释诂 толкования и 释官 жилище, т.к. для этой лексики словарь не дает указаний относительно их ареальной принадлежности, см. 王智群. 《方言》与扬雄词汇学. С. 97.*

Проблемы лингвистического образования

С.Ю. Семенова

О спецкурсе по теоретическим и прикладным вопросам изучения русской параметрической лексики

Обсуждаются лингвистические и методические основы спецкурса для студентов-лингвистов, посвященного свойствам русских параметрических слов, их структурированному лексикографическому представлению, а также алгоритмизации извлечения параметрической информации из текстов. Основное внимание в спецкурсе уделяется параметрическим существительным: их синтаксическим свойствам, словообразовательным моделям, тематической принадлежности, лексической сочетаемости, полисемии.

Ключевые слова: лингвистическое образование, лексическая семантика, параметрическое существительное, параметрическое прилагательное, параметрический глагол, извлечение информации из текста, автоматический семантический словарь.

1. Предмет и задачи спецкурса «Русская параметрическая лексика: теоретические и прикладные вопросы»

Автором настоящей статьи разработан и читается для студентов Института лингвистики РГГУ (специальность «Теоретическая и прикладная лингвистика» и направление «Фундаментальная и прикладная лингвистика») специальный курс, посвященный теоретическим и прикладным вопросам изучения тех русских слов, за которыми в лексической семантике закрепилось название «параметрические»: *высота, количество, весить, возрастая, длинный, скоро, цвет* и др. Для теории языка данный лексический пласт интересен целым рядом своих семантических, синтаксических, морфоло-

гических особенностей. Для приложений лингвистики параметрическая лексика важна тем, что с ее помощью в естественном языке выражается информация о величинах и о качественных признаках субстанций, которая часто служит объектом автоматического поиска по тексту и предметом пользовательских запросов к информационным системам. Спецкурс охватывает свойства параметрических слов, алгоритмизацию поиска параметрической информации и необходимое для реализации поиска формализованное словарное представление параметрической лексики.

Основное внимание в спецкурсе уделяется параметрическим существительным (параметрическим именам): их синтаксическим свойствам, словообразовательным моделям, тематической принадлежности, лексической сочетаемости, полисемии. На обсуждение свойств параметрических существительных отводится примерно половина аудиторного времени. Более кратко рассматриваются параметрические слова других частей речи: глаголы, прилагательные, наречия. Для несубстантивной параметрической лексики на рассмотрение, в частности, выносятся аспекты, получившие подробное освещение в семантической литературе: поведение параметрических глаголов в контексте отрицания, а также сочетаемость параметрического прилагательного с предметным именем, изучение которой во второй половине XX века стало объектом одного из направлений психолингвистических исследований.

Прикладной раздел спецкурса связан с задачей извлечения информации. Рассматриваются фрагменты алгоритма, который позволяет распознавать в тексте информацию о значениях количественных параметров, выраженную с помощью типовых лексических и синтаксических средств. Алгоритм в существенной мере опирается на словарные данные. Поэтому в спецкурсе рассматриваются основные приемы описания параметрической лексики в формализованном семантическом словаре; при этом используются форматы словаря РУСЛАН, структуру и метаязык которого разработала Н.Н. Леонтьева¹. Словарные статьи параметрических слов могут использоваться не только в задачах извлечения информации, но и при других видах автоматического текстового анализа.

Спецкурс базируется, главным образом, на результатах научных изысканий в соответствующих разделах лексической семантики и компьютерной лингвистики, проводившихся отечественными специалистами с 1970-х гг. по настоящее время. В определенной мере спецкурс отражает и опыт исследовательской работы автора.

2. О лингвистических основах спецкурса

Параметрическая лексика ряда однозначных частей речи, отражающая фундаментальные механизмы языковой номинации в абстрактной сфере, давно уже обращает на себя внимание исследователей. Еще Г. Пауль привлекал параметрическую лексику для рассмотрения аналогии между протяжением в пространстве и протяжением во времени²; О. Есперсен ввел релевантное для параметрических имен понятие некусного существительного, изучал лексико-грамматические средства выражения меры и сравнения количеств³; Е. Курилович на примере слов *высокий* и *высота* пояснял процессы синтаксической и лексической деривации⁴; Э. Сепир в книге «Язык» оперировал именами *height*, *depth*, *length* и некоторыми другими при объяснении явления фузии, а в статье «Градуирование. Семантическое исследование» предложил рассматривать форму компаратива как исходную по отношению к позитиву: *большой* и *близкий* трактовались им при помощи *больше* и *ближе*⁵; В.В. Виноградов с помощью русских параметрических прилагательных иллюстрировал широкое развитие антонимии в рамках грамматической категории качественных прилагательных⁶.

В последние десятилетия в работах по лексической семантике (включая ее логическое направление), психолингвистике, автоматической обработке текста и созданию лингвистических информационных ресурсов наблюдается ощутимый рост и углубление внимания к параметрической лексике. Это обусловлено целым рядом ее особенностей, в числе которых:

- значимость параметрической лексики, в первую очередь пространственной, для когнитивных исследований, в том числе для изучения закономерностей языкового кодирования визуальной информации;

- специфическая логико-семантическая природа параметрических имен, в частности, их близость к математическим функциям (способность принимать разные значения при варьировании характеризуемого объекта, который выступает в роли аргумента функции);

- необходимость релевантного лексикографического представления параметрической лексики в прикладных системах, вытекающая из актуальности параметрической информации для задач извлечения знаний;

- релевантность параметрической лексики для сферы научно-

технической терминологии, активно включаемой в отраслевые лексикографические информационные системы и в онтологии;

– развитая полисемия параметрической лексики, имеющая, в основном, регулярный характер, а также склонность такой лексики к дальнейшей семантической деривации.

Значительной вехой в развитии знаний о параметрических словах стали работы Ю.Д. Апресяна⁷. В понимании особой логической природы параметрических существительных важную роль сыграли работы Е.В. Падучевой, в частности статья «Об атрибутивном стяжении подчиненной предикации» 1980 г., а также работы рубежа 1990 – 2000 гг. о параметрах лексического значения и о параметрической диатезе глагольных лексем⁸. Параметрическим прилагательным, объяснению их сочетаемости с предметными именами посвящены работы А.Н. Журина⁹ и Е.В. Рахилиной¹⁰, давшие существенные импульсы развитию психолингвистического направления.

Начиная с 1990-х гг. в зону внимания исследователей входят такие вопросы, как, например, семантика отдельных параметрических слов и конструкций¹¹, метафоризация параметрических слов¹², их усвоение в онтогенезе¹³, типологическое изучение параметрической лексики¹⁴, различные аспекты извлечения параметрических данных¹⁵, а также общие грамматические и философско-лингвистические проблемы, связанные с параметрической лексикой¹⁶.

В области изучения параметрической лексики (эту область можно, на наш взгляд, назвать «лингвистической параметрикой») остаются невыясненные вопросы. В частности, нет полной ясности относительно состава, границ, внутренней семантики класса нечислительных параметрических имен, не устоялось и само название данного класса. Требуется дополнительного (и разноаспектного) изучения класс параметрических глаголов. Имеются сложности с объяснением различных сочетаемостных ограничений, связанных с параметрической лексикой. Далеко от полноты описание полисемии на всем множестве параметрических слов. Остаются неясности с интерпретацией частеречных преобразований параметрических лексем. Интересным представляется отслеживание элементов математического и физического мышления при образовании названий параметров. Требуются исследования в области параметрических терминологических словосочетаний. Лишь в начальной стадии пребывает изучение параметрической лексики в типологическом аспекте. Недостаточно изученными остаются возможности представления содержащейся в текстах информации при помощи параметрических фреймовых структур.

Ознакомление студентов – слушателей описываемого спецкурса – с накопленными теоретическими знаниями, с достижениями в прикладной области и с нерешенными задачами могут в дальнейшем способствовать заполнению указанных лакун.

3. К методическому обоснованию спецкурса

В настоящее время спецкурс читается студентам 4 курса специальности «Теоретическая и прикладная лингвистика». Их предварительным багажом, необходимым для усвоения учебного материала, являются знания из прослушанных ранее учебных курсов «Морфология»¹⁷, «Синтаксис», «Семантика», «Лексикография»¹⁸. Параллельно со спецкурсом студенты слушают обязательную дисциплину «Автоматическая обработка текста», обеспечивающую опорные знания для обсуждения прикладных вопросов спецкурса.

Изучение свойств параметрической лексики в спецкурсе в значительной мере происходит в форме диалога преподавателя с аудиторией. Диалог предполагает создание проблемных ситуаций на занятиях. Кроме того, участие в учебном процессе подразумевает выполнение студентами определенных домашних работ. Это способствует выработке у студентов таких навыков, как умение обосновать принадлежность конкретных слов к изучаемым семантическим классам, построить модельные примеры и контрпримеры. Студенты учатся отличать лексику смежных классов (прежде всего имен единиц измерения), применять лексические функции Жолковского-Мельчука для отражения сочетаемости и частеречных словообразовательных связей, устанавливать сочетаемостные свойства параметрических лексем, использовать фоновые знания о величинах из математики, физики, химии, истории, искать по корпусу дополнительные по отношению к материалам курса синтаксические конструкции и т.п. Определенная совокупность навыков складывается при совместном с преподавателем составлении статей параметрических лексем для прикладного семантического словаря: выявление актантной структуры лексем, описание таксономических категорий измеряемых объектов, указание словообразовательных и тезаурусных связей, отбор английских эквивалентов и т.п.

Материал спецкурса создает солидное пространство для самостоятельной научной работы студентов, в том числе для курсовых и дипломных работ. Кроме объемных научных задач, которые обозначены в предыдущем разделе, можно назвать, например, такие «локальные» задачи, как историко-лингвистические обзоры языко-

ведческих работ, затрагивающих параметрическую проблематику, или сопоставление свойств отдельных параметрических слов в русском и в изучаемых студентами иностранных языках.

После завершения перехода на двухступенчатую систему высшего образования спецкурс о параметрической лексике с соответствующими модификациями может быть встроен в учебные планы бакалавриата и магистратуры по направлениям фундаментальной, теоретической и прикладной лингвистики.

4. Общая структура и содержание спецкурса

Спецкурс состоит из двух частей. Первая посвящена параметрическим существительным, вторая — другим частям речи и извлечению параметрической информации. Ниже раскрываются основные темы обеих частей.

Тема 1. Определение параметрического существительного

Понятие параметрического слова. Параметрические слова разных полнозначных частей речи: существительные, глаголы, прилагательные, наречия. Словообразовательные связи между параметрическими словами разных частей речи.

Определение параметрического существительного / параметрического имени (по Ю.Д. Апресяну). Параметрические существительные — наименования количественных и неколичественных параметров. Параметрическое существительное как обозначение функции (в математическом смысле). Варьируемость значений параметров либо при варьировании объектов, характеризующихся по параметрам, либо при варьировании условий характеристики.

Диагностические контексты параметрических существительных. Конструкция с предикатами получения/передачи информации $V_{\text{inf}} \rightarrow N_{\text{par}}$ (*вычислить скорость; указать фамилию* и т.п.). Сочетаемость критерии отнесения существительного к классу количественных параметров.

Тема 2. Синтаксические конструкции с параметрическими существительными

Стандартная актантная структура параметрического существительного: валентность на имя характеризующегося объекта, валентность на значение параметра.

Атрибутивные конструкции, связывающие имя параметра с именем характеризующегося объекта. Ограничения на употребление параметрических имен в атрибутивных конструкциях. Атрибутивные

конструкции общего вида $N_{obj} \rightarrow N_{par}$ (*дом большой высоты; дом высотой 10 м* и нек. др.), обязательность зависимого при атрибуте. Сравнение сочетаемостных ограничений на подчиненное существительное в конструкциях $V_{inf} \rightarrow N_{par}$ и $N_{obj} \rightarrow N_{par}$.

Конструкции (лексико-синтаксические модели), выражающие значения количественных параметров. Языковые средства выражения математического равенства (точного, приближенного), неравенств (сверху, снизу, двойного). Описательные способы выражения значений параметров. Предложные атрибутивные конструкции, выражающие значения параметров (*трава по пояс* и т.п.).

Тема 3. Параметрические существительные с нестандартной актантной структурой

Имена двухобъектных количественных параметров (*расстояние, сумма* и др.). Синтаксическая однородность / неоднородность членов объектной пары. Сочинительное сокращение в конструкциях с двухобъектными параметрическими именами. Возможность / невозможность коммутирования членов объектной пары.

Имена параметров с интервальным значением. Способность ряда параметров выступать либо как двухобъектные, либо как интервальные. Параметрические имена с семантикой изменения количества (*приращение, перепад* и др.). Представление о диатезах параметрических имен.

Имена числовых пропорций. Значимость обоих членов пропорции vs формальный характер одного из них.

Тема 4. Количественный параметр и единица измерения

Наименования единиц измерения, некоторые общенаучные и этимологические сведения о них. Основные и производные единицы измерения, окказиональные единицы измерения. Мотивировка некоторых темпоральных и энергетических единиц измерения. Антропоцентризм ряда единиц измерения. Представление о грамматике описания единиц измерения в словарях и онтологиях.

Тема 5. Модели вопросов о значениях параметров

Представление о вопросно-ответном соответствии и о вопросно-ответных системах на естественном языке. Диагностический контекст параметрического существительного ($V_{inf} \rightarrow N_{par}$) как сжатый косвенный вопрос (по Е.В. Падучевой). Близость вопроса с вопросительным словом *каков* (*Каков + N_{par}?*) к диагностическому контексту параметрических существительных.

Модели вопросов о значениях параметров, включающие N_{par} . Ограничения на употребление конкретных N_{par} в моделях. Модели

вопросов о значениях параметров, не включающие $N_{\text{пар}}$. Восстановление по контексту наименования параметра, значение которого запрашивается. Средства замещения $N_{\text{пар}}$ в вопросах о значениях параметров: наименования единиц измерения; несубстантивная параметрическая лексика.

Модели вопросов о результатах математических операций: о сумме, разности, кратности величин. Соотнесение вопросительных местоимений и наречий с наименованиями количественных и не-количественных параметров.

Тема 6. Внутренняя структура подкласса имен количественных параметров

Имена линейных измерений и их семантико-синтаксические свойства (по Ю.Д. Апресяну). Проявление полисемии «параметр – большое значение параметра» в данной семантической группе.

Тематические группы параметрических существительных: имена пространственных, темпоральных, скоростных, фактурных, энергетических, экономических, вероятностных, общеколичественных параметров.

Различия параметрических имен по степени выраженности количественного компонента в их семантической структуре. «Сильные» и «средние» количественные параметры. Периферия класса имен количественных параметров, «квазипараметры».

Тема 7. Регулярные механизмы образования параметрических существительных

Субстантивация мотивирующего прилагательного как механизм синтаксической деривации параметрических имен, лексическая деривация параметрического значения (по Е. Куриловичу). Имена параметров и имена свойств, образованные от прилагательных (*высота* и *белизна*). Представление о нексусных существительных (по О. Есперсену). Образование параметрического имени от мотивирующего глагола. Лексические функции, связывающие исходный глагол и параметрическое имя. Образование параметрических имен по метонимии от имен пространственных сущностей и физических явлений.

Образование параметрических имен от названий единиц измерения, параметрические имена французского происхождения на (-аж). Составные параметрические имена. Дериваты от обозначения меры: *диаметр*, *периметр*, *параметр*, *размер*.

Тема 8. Таксономия измеряемого объекта и вторичная семантическая деривация у параметрических существительных

Представление о теории лексического значения предикатных лексем Е.В. Падучевой. Исходные таксономические категории измеряемых объектов для параметрических имен разных тематических групп. Генитивная конструкция $N_{\text{par}} - > (N_{\text{obj}})_{\text{gen}}$ (*глубина колодца*) как опорный контекст для изучения семантической деривации параметрических имен.

Метафорические значения параметрических имен, «сильная» и «слабая» метафоризация. Индивидуальное развитие метафорических значений у конкретных параметрических имен. Образование параметрических терминов. Метонимический сдвиг таксономической категории измеряемого объекта. Десемантизация параметрических имен. Эллиптическое опущение наименования измеряемого объекта. Образование предметных значений у параметрических имен.

Регулярная полисемия «параметр – большое значение параметра». Основные типы контекстов для лексем, обозначающих большое значение параметра.

Тема 9. Имена неколичественных параметров

Состав подкласса имен неколичественных параметров. Периферия подкласса. Имена неколичественных параметров как наименования атрибутов в формальных информационных структурах. Основные способы выражения значений неколичественных параметров: указание гипонима, аксиологическая оценка, заполнение фрейма, перечисление пунктов, связный текст.

Тема 10. Параметрический глагол

Понятие параметрического глагола. Ядерная группа параметрических глаголов, имеющих обязательную валентность на значение параметра. Параметрические глаголы под оператором отрицания (по Ю.Д. Апресяну).

Параметрические глаголы как средства замещения параметрического имени в высказываниях, выражающих значения параметров. Приставочные параметрические глаголы. Глаголы изменения количества, их модели управления и лексическая сочетаемость. Преобразование глагольных конструкций в номинативные структуры при автоматическом извлечении информации.

Глаголы, составляющие типовые контексты параметрического имени.

Тема 11. Параметрическое прилагательное

Основные типы прилагательных, выражающих значения параметров: качественные прилагательные приблизительной коли-

чественной оценки (*высокий, дешевый*), составные относительные прилагательные точной и приблизительное количественной оценки (*двухлетний, многочасовой, крупногабаритный*).

Конструкция $\text{Adj}_{\text{dim par}} + \text{N}_{\text{obj}}$, где $\text{Adj}_{\text{dim par}}$ – качественное параметрическое пространственное прилагательное типа *высокий*, а N_{obj} – имя определяемого объекта (*высокое дерево*). Некоторые закономерности сочетаемости пространственного параметрического прилагательного и предметного существительного (по Ю.Д. Апресяну, А.Н. Журиンスкому и Е.В. Рахилиной). Полисемия качественных параметрических прилагательных.

Конструкция $\text{Adj} + \text{N}_{\text{par}}$. Прилагательные, выражающие количественное значение N_{par} . Сочетаемость метафорических значений качественных пространственных прилагательных с N_{par} (*высокая температура, широкий диапазон*).

Прилагательные, характеризующие обстоятельства измерения величин. Относительные прилагательные, образующие названия параметрических терминов (некоторые примеры реляций).

Тема 12. Параметрическое наречие

Качественные параметрические наречия приблизительной оценки, образованные от прилагательных. Параметрическое наречие как участник синонимических преобразований и деривационных процессов (*быстро бежать – быстрый бег; глубокий водоем, глубоко погружаться, глубокое погружение* и т.п.). Параметрические наречия других морфологических типов (*вдали, налегке* и т.п.). Образование предикативных значений у параметрических наречий.

Тема 13. Общенаучное понятие параметра

Параметр как аргумент некоторой функции. Определения слова «параметр» в толковых и энциклопедических словарях. Параметр в математике, технике, физико-химическом анализе. Понятие параметра в общенаучном дискурсе.

Параметр как функция (как функциональная зависимость). Терминологическая традиция в лексической семантике, связанная с понятием параметра. Параметрическое слово как слово, соотношенное с величиной (параметрические существительные и глаголы), и как обозначение количественного значения величины (параметрические прилагательные и наречия). Изменчивость значения как основное свойство параметра. Имена, обозначающие неизменяемые величины.

Тема 14. Извлечение (поиск) параметрической информации

Задачи автоматического извлечения информации и знаний в

компьютерной лингвистике. Понятие фактографического поиска. Параметрическая информация. Параметрическая триада. Задача извлечения параметрической информации. Формы представления параметрической информации: объектно-характеристическая таблица, фактографическая база данных, база текстовых фактов, предикатно-аргументная структура, фрейм.

Алгоритмизация поиска параметрической информации. Этапы автоматической обработки текста, предшествующие поиску: морфологический анализ, синтаксический анализ (общее представление об этих этапах). Автоматическое распознавание полного названия параметра. Автоматическое распознавание имени измеряемого объекта. Автоматическое распознавание количественного значения параметра.

Дальнейшие преобразования параметрической информации в информационных системах.

Тема 15. Словарные ресурсы для извлечения параметрической информации

Круг словарных сведений, необходимых для извлечения параметрической информации. Представление об автоматическом семантическом словаре. Словарь РУСЛАН (НИВЦ МГУ): назначение, структура. Общее представление о метаязыке словаря РУСЛАН.

Описание параметрической лексики в словаре РУСЛАН: построение статей параметрических существительных, глаголов, прилагательных, наречий и предикативов.

Примечания

- ¹ *Леонтьева Н.Н.* Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие для студ. лингв. фак. вузов. М.: Академия, 2006. В настоящее время словарь РУСЛАН базируется в НИВЦ МГУ.
- ² *Пауль Г.* Принципы истории языка. М.: Изд. иностр. лит., 1960. С. 116.
- ³ *Есперсен О.* Философия грамматики. М.: Изд. иностр. лит., 1958. Гл. XX и XVIII.
- ⁴ *Курилович Е.* Очерки по лингвистике. М.: Изд. иностр. лит., 1962. С. 64–65.
- ⁵ *Сепир Э.* Язык // *Сепир Э.* Избранные труды по языкознанию и культурологии. 2-е изд. М: Прогресс, 2001. С. 124. *Он же.* Градуирование. Семантическое исследование // Новое в зарубежной лингвистике. Вып. 16. М.: Прогресс, 1985.
- ⁶ *Виноградов В.В.* Русский язык (Грамматическое учение о слове). 3-е изд., испр. М.: Высш. школа, 1986. С. 177.
- ⁷ *Апресян Ю.Д.* Лексическая семантика. Синонимические средства языка. М.: Наука, 1974. *Он же.* Синтаксические признаки в модели языка (на

- материале одного класса атрибутивных конструкций) // Вопросы кибернетики. Прикладные аспекты лингвистической теории. М., 1987. *Он же*. Лексикографические портреты (на примере глагола *быть*) // Научно-техническая информация. Серия 2. 1992, № 3. *Он же*. Фундаментальная классификация предикатов и системная лексикография // Грамматические категории: иерархии, связи, взаимодействие. Материалы международной научной конференции. СПб., 2003.
- ⁸ *Падучева Е.В.* Об атрибутивном стяжении подчиненной предикации // Машинный перевод и прикладная лингвистика. М., 1980. Вып. 20. *Она же*. Семантика глаголов выбора // Известия РАН. Серия лит. и яз. 1999, Т. 58, № 5–6. *Она же*. О параметрах лексического значения глагола: таксономический класс участника // Русский язык в научном освещении. 2002. № 1 (3).
- ⁹ *Журицкий А.Н.* О семантической структуре пространственных прилагательных // Семантическая структура слова. Психолингвистические исследования. М.: Наука, 1971.
- ¹⁰ *Рахилина Е.В.* Семантика размера // Семиотика и информатика. М., 1995. Вып. 34. *Она же*. Атрибутивные конструкции с русскими прилагательными размера: частный случай // *Nazwy barw i wymiarow*. Stockholm.: Stockholms universitet. Slaviska institutionen, 1997.
- ¹¹ См., например, *Богуславский И.М., Иомдин Л.Л.* О валентных свойствах одного широкого класса существительных // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». Вып. 9 (16). М., 2010; *Кюсева М.В., Рыжова Д.А., Холкина Л.С.* Прилагательные *тяжелый* и *легкий* в типологической перспективе // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодн. Междунар. конференции «Диалог». Вып. 11 (18). Т.1. М., 2012; *Семенова С.Ю.* Параметрические имена с нестандартной актантной структурой. // Московский лингвистический журнал. 1996. Т.2; *Она же*. О некоторых свойствах имен пространственных параметров // Логический анализ языка: Языки пространств. М.: Языки русс. культуры, 2000; *Урысон Е.В.* *Большой и маленький*: шкала размера в русском языке // Логический анализ языка: Квантификативный аспект языка. М.: Индрик, 2005; *Шеманаева О.Ю.* Точные и приближительные оценки размеров предметов в русском языке // Компьютерная лингвистика и интеллектуальные технологии: Труды Междунар. конференции «Диалог 2006». М., 2006.
- ¹² *Семенова С.Ю.* О таксономии актантов параметрических имен // Динамические модели : Слово. Предложение. Текст. Сб. статей в честь Е.В. Падучевой. М., 2008; *Она же*. Русское имя параметра: метафорические и метонимические процессы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодн. Междунар. конференции «Диалог». Вып. 11 (18). Т.1. М., 2012.
- ¹³ *Пузанова Ю.С.* Параметрические прилагательные русского языка в онтогенезе. Автореф. дисс.... канд. филол. наук. СПб., 2012.
- ¹⁴ *Гилярова К.А.* Языковая концептуализация формы физических объектов. Автореф. дисс. ... канд. филол. наук. М., 2002; *Кюсева М.В., Рыжова Д.А., Холкина Л.С.* Указ. соч.

- ¹⁵ *Семенова С.Ю.* Алгоритм извлечения информации о параметрах из текстов рефератов и первичных документов // Научно-техническая информация. Серия 2. 1991, № 6; *Она же.* Поиск параметрической информации в тексте: алгоритмический и лексикографический аспекты // Труды Международного семинара Диалог'96 по компьютерной лингвистике и ее приложениям. М., 1996; *Она же.* Прилагательные в семантическом словаре одной прикладной системы // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. М., 1998; *Она же.* Наречия и предикативы в прикладном семантическом словаре // Труды Международного семинара Диалог'99 по компьютерной лингвистике и ее приложениям. Таруса, 1999. Т.1.
- ¹⁶ *Колесов В.В.* Философия русского слова. СПб.: Юна, 2002. С. 210–212. *Крылов С.А.* Количество как понятийная категория // Логический анализ языка: Квантификативный аспект языка. М.: Индрик, 2005; *Семенова С.Ю.* Параметризация как метод познания и как языковой механизм // Логический анализ языка: Квантификативный аспект языка. М.: Индрик, 2005.
- ¹⁷ *Устинова Е.Г.* О построении курса «Морфология» по специальности «Теоретическая и прикладная лингвистика». Программа и формы контроля знаний // Вестник РГГУ. Серия «Языкознание/МЛЖ». 2007. № 8. С. 145–167.
- ¹⁸ Синтаксис. Семантика. Лексикография: Программы курсов / Отв. ред. С.И. Гиндин. М.: Изд. центр РГГУ, 2000.

Abstracts

Nina N. Leontieva

To Our Colleagues Whose Anniversaries We Celebrate This Year

The author tries to portray three outstanding contemporary Russian linguists: Irina Sevbo, Anatoly Shaikevich and Alexey Gladkij

Keywords: Irina Sevbo, text structure, text generation, stylistical diagnostics, Anatoly Shaikevich, statistical linguistics, decoding procedures, Alexey Gladkij, semantical analysis, syntactic structures, linguistic education, researcher's modus vivendi.

N. A. Korotaev

Analyzing Russian Coordinative Constructions in Terms of Syntactic Groups by A. V. Gladkij

Representation of syntactic structure in terms of subordination trees is highly problematic when dealing with coordinate constructions. Several adaptations of subordination trees have been proposed so far, but none of them allows for an intuitively adequate analysis. The paper presents the basic ideas and formal methods used in the Syntactic Groups model introduced by A.V. Gladkij. It shows that most Russian coordinate constructions are successfully analyzed — in a way that combines the advantages of both subordination trees and immediate constituents approaches. However, at least one type of coordination, the one based on lexical and/or semantic (and not functional) homogeneity, may not be properly represented neither in immediate constituents nor in the Syntactic Groups model. In the paper, modifications of the original model are proposed that use the co-subordination principle to provide intuitively adequate structures for both functionally and lexically/semantically supported coordinate constructions.

Keywords: Russian syntax, formal models, syntactic groups, coordination, homogeneity.

Marina R. Kaul

“If You are Interested in Statistics, the Only Way for You is to Shaikevich”

The author recollects her first meetings with Anatolij Ja. Shaikevich during undergraduate and postgraduate years, the impressions of his scientific works and of his public activities.

Keywords: Anatolij Ja. Shaikevich, pedagogical style, distributial statistical analysis of texts, struggle against sham in science, civic fortitude, modesty of behaviour scientific.

Valentina I. Perebejnos

How I Started to Study Statistical Linguistics

The author tells how Anatolij Ja. Shaikevich's constructive criticism helped her to understand the problems and methods of statistical linguistics

Keywords: Anatolij Ja. Shaikevich, statistical methods, the style of scientific relations.

Marina Ja. Glovinskaja

'Harmony' in the "Production Dictionary of Russian"

The paper contains two entries for the verbs *garmonirovat'* and *garmonia*.

Keywords. Production dictionary, , types of lexicographic information.

Ju.D. Apresjan

Begat' and Bezhat' – Two Entries for the "Production Dictionary of Russian"

The paper contains entries for the verbs *begat'* and *bezhat'* from the "Production dictionary of Russian" which is now being prepared for publication. Both entries are supplied with detailed linguistic comments. Since the dictionary is designed for purposes of speech production it has to contain such information on words in their current senses as is necessary and sufficient not only for their understanding in an arbitrary text but also for their proper use by the speakers. The implementation of this principle results in a considerable expansion of the customary lexicographic information and in the inclusion of totally new types of information in dictionary entries. Among them one should list analytical definitions of lexical meanings and their modifications in various textual conditions, a description of complicated interactions of lexical and grammatical meanings, detailed government patterns and their transformations, non-valent syntactic constructions characteristic of the word at issue, a semantically ordered description of its co-occurrence constraints, information on its synonyms, antonyms, derivatives and other words semantically related to the given item, and the like.

Keywords. Production dictionary, types of lexicographic information, analytical sense definitions, semantic modifications, interaction of lexis and grammar, government patterns, non-valent syntactic constructions, co-occurrence constraints.

L.P. Krysin

Scientific and Technical Terms in the Russian Urban Non-codified Speech

The paper contains an analysis of the usage of special terms in the contemporary Russian urban un-normative speech. This usage is characterized by formal and semantic transformations of special terms.

Keywords: national language, subsystem of language, special terminology, urban un-normative speech, adaptation of words.

Maria V. Akhmetova

Variation of Names of Inhabitants of the Cities (Based on Electronic Database "Integrum")

The paper is based on statistical analysis of usage of names of citizens in local newspapers (cities Irkutsk, Kirov, Penza, Petrozavodsk, Pyatigorsk, Stavropol, Tambov and Tver) on the materials of electronic database "Integrum". Research that has been undertaken, reveals local preference in use of variants or their rivalry; the latter can be caused both with editorial practice and actual tendencies of language.

Keywords: variation, katoikonyms, usage, language norm, "Integrum"

Ekaterina L. Brouko

Constructing and Imitation as Strategies of Syntax Acquisition

This article considers the characteristics of child speech in the period of transition from one- to two-word utterances. Two mechanisms have been identified that allow combination of words into utterances: the first mechanism is which imitation requires systematic reproduction by the child of the speech fragments already available from the adults' or their own speech. The second mechanism is construction which presupposes independent creation by the child of utterances that have no examples in the speech of the adults. A correlation has been traced between the dominant strategy and several aspects of language acquisition, such as the contents of the early vocabulary and child's level of activity in a dialogue. Furthermore, the article describes changes at different linguistic levels that accompany transition from one strategy to the other.

Keywords: language acquisition, syntax, individual strategies

Tatiana Yu. Kobzareva

Syntactical sentence structure encoded by punctuation marks

In this work, we investigate which aspects of Russian sentence structure are encoded by intra-sentential punctuation marks. We solve this problem via defining an ideal sentence structure based on dependency syntax. Using this ideal structure, we analyze the semantic spectrum of intra-sentential punctuation marks and build their functional trichotomy. This allows us to expose the functions of punctuation marks in complex sentences in Russian and other European languages.

Keywords: syntax, punctuation, semantics of Russian punctuation marks, syntactic structure of sentences

Anna R. Pestova, Olga A. Sharykina

Russian Idiomatic Expression *navesti shoroh (-a / -u)* in Various Spheres of Speech Communication

The present paper deals with functioning of the phraseological unit *navesti shoroh (-a / -u)* in various spheres of speech communication. The problem we are interested in is which semantic components of the unit come into focus depending on the sphere of speech communication and whether its government is influenced. The study showed that depending on the context such potential semantic components of the idiomatic expression as 'an unexpected success', 'reprimand', 'start a row', 'force', 'establish order', 'surprise' and 'frighten' may come into focus. The government does not depend on the sphere of communication in which this phraseological unit is used.

Keywords: phraseology, speech communication, government.

Anastasia A. Saltykova

Some Cases of Implicated Axiological Modality Indication in Mass-media Texts

Present article contains brief results of analysis of implicated axiological modality in informational-analytical television texts. The objects of research were the texts by Marianna Maximovskaya ("Nedelya", TV REN) Michael

Leontiev (“Odnako”, The First Channel) and Victor Shenderovich (“Itogo”, NTV Channel). The main aim of investigation was to reveal and classify indicators allowing to discern appraisals implicated in utterances and to bring out reasons for the axiological modality to appear. In this article there are exposed six general types of indicators and three contributing factors of implicatures in an utterance.

Keywords: pragmatics, implicature, modality, mass-media, implicated appraisal indicators, Cooperative Principle, polarization.

A.Sh. Kaganov

Instrumental Investigations of the Spectral Characteristics of Speech for the Sake of Forensic Identification of the Speaker: Theoretical Grounds and Technology of Research

The theme of this article is research’s technology of the spectral characteristics of the speaker in the task of forensic identification. The article also reviews theoretical foundations that establish a connection between the spectral characteristics of speech and biometrics of human’s vocal apparatus. This article is based on literature data and on the material of forensic investigations.

Keywords: oral speech, criminalistics, forensic examination, identification, formant, spectral analysis, biometry.

Zoya M. Shalyapina

Some Prospects of Automatic Optimization of Lingware for Text Processing within the Entity-Based Linguistic Framework

The paper discusses some possibilities opened up by the entity-based linguistic framework for simulating human language acquisition as automatic optimization and self-organization of lingware designed for text processing. The discussion is based on the JaRAP experimental complex for Japanese-Russian automatic translation (Institute of Oriental Studies of the Russian Academy of Sciences).

Keywords: automatic language processing, lingware, optimization, entity-based linguistic framework

Natalia V. Gurian

A Note on Some Features of the Lexical Material Organization and Presentation in the Ancient Chinese Dictionary “Fanyan”.

The article is devoted to the analysis and description of the first Chinese dictionary of local words. Special attention is given to the ways of lexical items organization and description used in “Fangyan”. The author also touches upon the question of the continuity between Fangyan and its predecessor Erya.

Keywords: Fangyan, Erya, dictionary structure, semantic domains, format of the dictionary entries (glosses).

Sophia Yu. Semenova

On the Special Course “The Russian Parametric Words: Theory and Applications”

The linguistic and methodic foundations of the special course on the Russian

parametric words features, the parametric words representation in the machine semantic dictionary, and parametric data extraction are under consideration. The principle attention in the special course is paid to the parametric nouns. The semantic, syntactic, and derivative features of the parametric nouns are discussed as well as the inner structure of the lexical class above.

Keywords: education of linguists, lexical semantics, parametric noun, parametric adjective, parametric verb, information extraction, machine semantic dictionary.

Об авторах

Юрий Дереникович Апресян – доктор филологических наук, академик РАН, профессор, зав. Сектором теоретической семантики Институт русского языка РАН; juri.apresjan@gmail.com

Мария Вячеславовна Ахметова – канд. филол. наук, докторант Центра типологии и семиотики фольклора РГГУ; malinxi@rambler.ru

Екатерина Леонидовна Бровко – преподаватель Каф. теоретической и прикладной лингвистики Института лингвистики РГГУ; katia.brovko@gmail.com

Марина Яковлевна Гловинская – докт. филол. наук, ведущий науч. сотрудник Института русского языка им. В.В. Виноградова РАН; marinaglov@yandex.ru

Наталья Викторовна Гурьян – канд. филол. наук, докторант Каф. китайской филологии Института стран Азии и Африки МГУ им. М.В. Ломоносова; guryan-natalia@yandex.ru.

Александр Шлемович Каганов – канд. техн. наук, старший науч. сотрудник Российского центра судебно-медицинской экспертизы Минздрава России, эксперт Главного центра судебно-медицинских и криминалистических экспертиз Министерства обороны Российской Федерации; ash_kaganov@mail.ru.

Марина Рафаиловна Кауль – канд. филол. наук, профессор Каф. английского языка Института филологии и истории РГГУ, лексиколог, лексикограф. Организатор Российско-американского учебно-научного центра РГГУ; kaulmar@mail.ru.

Татьяна Юрьевна Кобзарева – канд. филол. наук, доцент Каф. теоретической и прикладной лингвистики Института лингвистики РГГУ; tipling@gmail.com.

Николай Алексеевич Коротаев – канд. филол. наук, доцент Учебно-научного центра лингвистической типологии РГГУ; n_korotaev@hotmail.com.

Леонид Петрович Крысин – доктор филол. наук, профессор, зав. Отделом современного русского языка Института русского языка им. В.В. Виноградова РАН; leonid-krysin-mail.ru

Нина Николаевна Леонтьева – доктор техн. наук, профессор, ведущий науч. сотрудник Научно-исследовательского вычислительного центра МГУ им. М.В. Ломоносова; leont-nn@yandex.ru.

Валентина Исидоровна Перебейнос – профессор Киевского национального университета иностранных языков, зав. Лабораторией компьютерной лексикографии.

Анна Разифовна Пестова – младший науч. сотрудник Отдела современного русского языка Института русского языка им. В.В. Виноградова РАН; pestova2012@gmail.com.

Анастасия Александровна Салтыкова – преподаватель Каф. русского языка Московского государственного горного университета; aasaltykova@mail.ru

Софья Юльевна Семенова – канд. филол. наук, старший науч. сотрудник Института научной информации по общественным наукам РАН, доцент Каф. теоретической и прикладной лингвистики Института лингвистики РГГУ; sonya_sem@mail.ru

Зоя Михайловна Шалапина – канд. филол. наук, зав. Отделом языков народов Азии и Африки Института востоковедения РАН; languages@ivran.ru.

Ольга Александровна Шарыкина – канд. филол. наук, зав. Отделом лингвистической типологии редакции журнала «Вопросы языкознания»; oka_2311@mail.ru.

About the Authors

Maria V. Akhmetova – Cand. of Phil., Doctoral Student of the Center for Typology and Semiotics of Folklore, RSUH; malinxi@rambler.ru

Ju. D. Apresjan – Doctor Hab. in Phil., Member of the Russian Academy of Sciences, Main Scientific Researcher in the A.Kharkevich Institute for the Problems of Information Transfer, Chair of the Section for Theoretical Semantics in the Viktor Vinogradov Institute of the Russian Language; juri.apresjan@gmail.com.

Ekaterina L. Brovko – Lecturer at the Department of the Theoretical and Applied Linguistics, RSUH; katia.brovko@gmail.com.

Marina Ja. Glovinskaja – Doctor Hab. in Phil., Leading Scientific Researcher in the Viktor Vinogradov Institute of the Russian Language (Russian Academia of Science); marinaglov@yandex.ru

Natalia V. Gurian – Cand. of Phil., Doctoral Student of the Chinese Philology Department of the Institute of Asian and African Studies in the Lomonosov Moscow State University; guryan-natalia@yandex.ru.

A. Sh. Kaganov – Cand. of Techn., Senior Researcher of Russian Center for Forensic-Medical's Examination of Ministry of Public Health of RF, expert of Main Center Forensic-Medical and Forensics' Examinations of Ministry of Defence of RF; ash_kaganov@mail.ru.

Marina R. Kaul – Cand. of Phil., Professor at the Department of Modern English at the School of Philology and History, RSUH. Lexicologist and Lexicographer. M. Kaul is organized Russian-American Academic Center at the RSUH in 2004; kaulmar@mail.ru.

Tatiana Yu. Kobzareva – Cand. of Phil., Assistant Professor of the RSUH Department of Theoretical and Applied Linguistics of Institute of Linguistics; tipling@gmail.com.

N. A. Korotaev – Cand. of Phil., Associate Professor at the Research Centre for Linguistic Typology, RSUH; n_korotaev@hotmail.com.

L. P. Krysin – Doctor Hab. in Phil., Chief of the Department of the Contemporary Russian Language in the Vinogradov's Institute of Russian Language (Russian Academia of Science); leonid-krysin-mail.ru

Nina N. Leontyeva – Doctor Hab. in Techn., Leading Researcher of the Scientific Computational Center, the Moscow University; leont-nn@yandex.ru.

Valentina I. Perebejnos – Professor of the Kiev National University for the Foreign Languages, the Chair of the Laboratory of Computational Lexicography.

Anna R. Pestova – Junior Research Worker of the Modern Russian Language Department of Viktor Vingoradov Russian Language Institute (Russian Academy of Sciences); pestova2012@gmail.com.

Anastasia A. Saltykova – Lecturer of the Moscow State Mining University, Russian language department; aasaltykova@mail.ru

Sophia Yu. Semenova – Cand. of Phil., Senior Research Worker of the Institute of Scientific Information on Social Sciences (Russian Academy of Sciences), Lecturer at the RSUH Department of Theoretical and Applied Linguistics of Institute of linguistics; sonya_sem@mail.ru

Zoya M. Shalyapina – Cand. of Phil., Head of the Department of Asian and African Languages of the Institute of Oriental Studies (Russian Academy of Sciences); languages@ivran.ru.

Olga A. Sharykina – Cand. of Phil., Head of Linguistic Typology Department of Linguistic Typology of “The Problems of Linguistics” Editorial Office; oka_2311@mail.ru.

Заведующая редакцией: *И.В. Лебедева*

Художник: *В.В. Сурков*

Художник номера: *В.Н. Хотеев*

Компьютерная верстка: *Н.Г. Семенова*

Формат 60x90 ¹/₁₆

Усл. печ.л. 14,4 Уч.-изд. л. 15,1

Тираж 1050 экз. Заказ № 97

Издательский центр
Российского государственного
гуманитарного университета
125993 Москва, Миусская пл., 6
www.rgggu.ru
www.knigirgggu.ru