

И. А. Пильщиков,
А. С. Старостин

*Основные проблемы
автоматизации
базовых процедур
ритмико-синтаксиче-
ского анализа
силлабо-тонических текстов¹*

1.



ундаментальная теоретическая проблема, в рамках которой ставятся и решаются вопросы, затронутые в статье, — это проблема формализации стиховедческого знания и автоматизации стиховедческого труда (ср. [Вигурский, Пильщиков 2003]). Конкретная задача — это решение проблемы алгоритмизации и автоматизации основных процедур лингвостиховедческого анализа. Речь пойдет преимущественно об анализе силлабо-тонических поэтических текстов на русском языке, но многое из сказанного может быть распространено на силлаботонику немецкую или английскую.

В статье излагается частично реализованный план теоретического исследования и серии практических работ, нацеленных на развитие методов автоматизированной акцентологической

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 08-06-00192).

и морфолого-синтаксической разметки силлабо-тонических текстов и адаптацию этих методов к нуждам стиховедческих штудий (в круг которых входит анализ метрики, ритмики, строфики, рифмы, морфологии и синтаксиса стиха). На начальном этапе исследования на первый план выходят проблемы, связанные с ритмикой и синтаксисом стиха. В коллективном исследовании принимают участие несколько человек, которых можно считать соавторами настоящей статьи².

Общий ход исследования предусматривает постановку и решение ряда взаимосвязанных теоретических задач и получение в качестве основного практического результата нескольких программно-инструментальных и программно-информационных продуктов. Важным побочным результатом (by-product) должны стать новые способы учета и поиска результатов лингвостиховедческого анализа — то есть новые способы аккумуляции стиховедческого знания.

2. Начнем с *общих* задач алгоритмизации и автоматизации лингвостиховедческого анализа. Таких задач, на наш взгляд, как минимум четыре.

2.1. Необходимо оптимизировать морфолого-акцентологическую разметку с учетом

- 1) особенностей русского стиха (слова, обычно безударные в прозе, могут принимать на себя ударение в контексте стихотворного метра),
- 2) особенностей русского синтаксиса (слова могут быть ударными либо безударными в зависимости от синтаксической позиции),
- 3) особенностей исторической акцентологии русского языка (орфоэпические нормы XVIII и XIX веков значительно отличаются от сегодняшних).

Отмечая расхождения между современными ударениями в словоформах, которые предлагает программа-анализатор, и фактическими ударениями, которые подсказывает метрическая схема

² Это А. С. Белоусова, В. С. Белоусова, К. А. Головастикова и С. Е. Ляпин.

стиха, мы можем продолжить сбор материалов к дифференциальному словарю ударений русского языка XVIII и XIX веков, начатый Н. А. Еськовой, — фиксируя (в отличие от нее) не только расхождения, но и совпадения между современной и старинной нормой, а также частотность вариантов ([Еськова 2008]; ср. [Перцов 2006]). В результатах такой работы будут заинтересованы не только стиховеды, но и специалисты по исторической орфоэпии и лексикологии.

2.2. Необходимо поставить и разрешить практически задачу автоматического определения стихотворного метра конкретных текстов. Для этого должны быть

- 1) разработаны программы, соотносящие распределение фактических ударений слов в стихотворных строках с метрическими схемами, актуальными для данной культурной традиции;
- 2) найден теоретический и практический подход к проблеме аномалий в рамках метра, затрудняющих распознавание стихотворного размера, но не выводящих текст за его пределы.

Чтобы решить эти проблемы, необходимо получить не *дескриптивные*, а *конструктивные* определения основных понятий классической русской метрики. Напомним, что дескриптивным определением в методологии науки называют дефиницию, задающую объект перечислением требуемых его свойств или функций; а конструктивное определение — это явное описание строения соответствующего объекта [Яглом 1980]. (Компьютер понимает только конструктивные определения.) Как заметил М. И. Шапир, развивая мысли И. М. Яглома (1980), можно сказать, что прикладные науки занимаются переводом дескриптивных определений в конструктивные, а теоретические — переводом конструктивных определений в дескриптивные.

Конструктивное определение стиха как такового науке пока не доступно, однако системы стихосложения определять мы можем (этому вопросу был посвящен доклад М. И. Шапира, состоявшийся летом 2005 г. на семинаре по лингвостиховедческому анализу поэтических текстов, первоначально проходившем под руководством С. А. Старостина). Основная мысль доклада такова: если при чисто теоретическом подходе системы стихосложения следует (по Шапиру) определять прежде всего в зависимости от наименьшей парадигма-

тической константы [Шапир 1995; 2000, с. 36–75]³, то конструктивное определение, полученное путем конвертации из чисто теоретического, может быть основано только на вероятностном подходе. Скажем, силлабика — это система стихосложения, в которой число слогов в строке неслучайно. Соответственно, в тонике неслучайно число тактов, а в силлаботонике неслучайно распределение ударных и безударных слогов (при этом неслучайно не значит постоянно). Чтобы это определение стало практическим, надо вычислить вероятности распределения ритмических формант в строке и соотнести их с реальными показателями интересующего нас текста. Речь идет не о сравнении реальных показателей с так называемой «языковой» или «прозаической» моделью размера, а о более высокой ударности маркированных элементов (например, иктов) по сравнению с немаркированными (в данном случае — с метрически безударными слогами). Также вероятно (через решение проблемы порогов) должен решаться вопрос о логаядах и переходных метрических формах⁴.

Из того, что метр рассматривается как инвариант, следует (по мысли Шапира) что он, метр, есть, прежде всего, характеристика не строки, а совокупности строк: фрагмента, произведения и т. д. Отсюда первостепенная роль контекста при диагностировании метра. Прямое следствие такого взгляда — это конструктивная дефиниция концепта «определение метра»: установить (диагностировать) метр — значит сначала соотнести фактическое звучание строк между собой в поисках общего и различного в их строении (то есть попытаться определить инвариант, модификациями которого являются соответствующие строки), а потом найти для этого инварианта ближайшее соответствие в метрическом репертуаре данной культурной традиции (ср. [Шапир 2005]). Для наших дел важно, что

³ Для силлаботоники такой константой является стопа — простейшая метрическая константа, образованная устойчивой комбинацией ударных и безударных слогов и могущая выступать в качестве конструктивной единицы стихотворной строки. Важно добавить, что под «ударными и безударными» в этом определении понимаются слоги, имеющие разное акцентологическое качество: обязательно ударные, преимущественно ударные, произвольно ударные, преимущественно безударные и обязательно безударные (из доклада М. И. Шапира).

⁴ Этим вопросом сейчас занимается А. С. Белоусова.

эту процедуру можно в значительной степени автоматизировать. Преобразование дескриптивных определений основных понятий классической русской метрики в конструктивные представляет важный шаг на пути к формализации стиховедческого знания и позволяет перевести важнейшие стиховедческие концепты на язык компьютерной системы.

2.3. Результаты автоматического акцентологического анализа должны быть подвергнуты филологическому редактированию. На основании полученной метаинформации, зафиксированной с помощью языка разметки, становится возможной автоматическая квалификация ритмических форм в рамках установленного метрического инварианта. На этом этапе следует приступить к созданию корпуса русских поэтических текстов XVIII–XXI веков с разметкой метрических и ритмических форм, регистрацией их в каталоге (указателе) и обеспечением поиска по ним (с возможностью перехода от результатов поиска к полным текстам). Затем должна быть добавлена автоматическая каталогизация словораздельных вариаций, а также диагностирование строфических и твердых форм (с обеспечением возможности поиска по всем этим параметрам).

2.4. Дальнейшим шагом должна стать более «глубокая» разметка сформированного русского поэтического корпуса. Внутри корпуса необходимо снимать морфологическую, синтаксическую и акцентную омонимию. При этом омонимия на одном уровне часто разрешается за счет другого уровня. Так, например, при анализе русского ямба акцентная информация позволяет избавиться от большого количества морфологических омонимов. Оптимизация автоматической морфологической разметки с учетом особенностей русского синтаксиса позволит до определенной степени минимизировать «ручное» редактирование (но, конечно, не даст возможности полностью от него отказаться). Филологическая редакция предполагает снятие (disambiguation) морфологической и синтаксической неоднозначности в разметке — или (NB!) фиксацию объективной неоднозначности, которая в стихе играет бóльшую роль, чем в прозе: для поэтического текста неоднозначность (ambiguity) является

не приводящим, а конститутивным свойством⁵. Отсюда возникает вопрос о способах фиксации и учета альтернативных интерпретаций текста при поисках и подсчетах.

При разработке алгоритмов автоматического лингвостиховедческого анализа необходимо поставить следующую сверхзадачу — совместить восходящий и нисходящий принципы (движение от текста к метру и от метра к тексту). На первом этапе морфо-акцентологические характеристики текста позволяют выдвинуть гипотезу о его стихотворном размере, которая может быть использована для дальнейшего анализа текста — прежде всего для снятия акцентной, морфологической и других видов омонимии. В идеальном анализаторе различные гипотезы должны оцениваться и ранжироваться по вероятности.

Следующий шаг — обеспечение возможностей полнотекстового поиска по морфолого-синтаксическим параметрам. Необходимо предоставить исследователю возможность соотносить эти параметры с метрическими и ритмическими характеристиками стиха (то есть нужно организовать комбинированный поиск с заранее определенными значениями атрибутов). Отдельная подзадача — как при формировании такого рода запросов учитывать слоги- и стоподеление. Параллельно должна быть решена проблема автоматизированной фонетической, грамматической и лексикологической квалификации рифм.

3. Для выполнения описанных выше работ целесообразно было создать информационную модель метаданных, которыми снабжаются тексты в результате анализа. Дескриптивный потенциал модели должен быть сопоставим с современным уровнем лингвистической и стиховедческой науки. Важным фактором является возможность последующего дополнения информационной модели без ущерба для данных, сформированных в рамках предыдущих версий модели.

В текущей версии модели используется несколько базовых понятий:

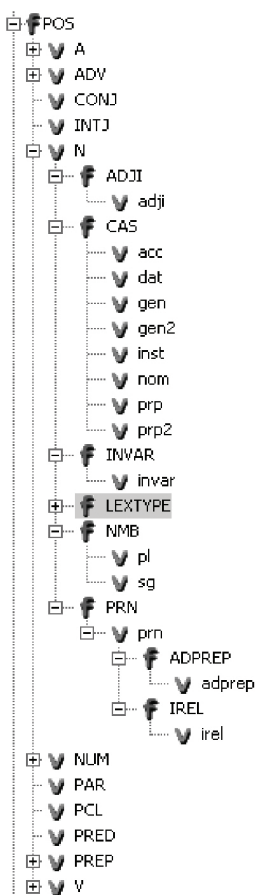
⁵ Актуальность проблеме поэтической неоднозначности, поставленной еще в 1930 г. английским поэтом и теоретиком литературы Уильямом Эмпсоном (см. [Emrson 1930]), в последние годы вернули М. И. Шапир [Шапир 1999b] и Н. В. Перцов [Перцов 2000].

- *атрибут* (именованная категория, которая может иметь заранее определенный или неопределенный конечный или бесконечный набор значений);
- *значение атрибута* (один из элементов указанного набора);
- *аннотация* (набор пар вида <атрибут, значение>, привязанный к конкретному месту текста; каждая аннотация включает служебный атрибут, называемый типом аннотации, значением которого может быть строка из заранее определенного набора строк);
- *тринотация* (расширение понятия аннотации: кроме типа и набора атрибутов тринотация может содержать в себе информацию о связях между входящими в нее аннотациями и другими тринотациями; таким образом, аннотация является частным случаем тринотации);
- *хранилище тринотаций* (множество тринотаций, допускающее такие операции, как удаление, добавление и модификация тринотаций, а также разнообразные средства поиска и выборки тринотаций по различным критериям);
- *иерархия атрибутов* (дерево, в котором отражаются допустимые наборы атрибутов и значений для того или иного типа аннотаций).

В модели иерархии атрибутов задаются для каждого типа аннотаций, что позволяет хорошо ее упорядочивать и структурировать.

Модель включает в себя несколько частных моделей: модель фонетики, модель морфологии, модель синтаксиса (см. о ней ниже, п. 6) и модель стиховедческих метаданных. Для всех моделей разрабатываются типы аннотаций и строится иерархия атрибутов.

Все иерархии являются пополняемыми, то есть при необходимости систему можно



дополнять новыми категориями и значениями. Фрагмент иерархии атрибутов для аннотаций морфологического уровня приводится справа. На верхнем уровне иерархии представлен атрибут POS (part of speech), отражающий категорию части речи. Набор его значений (A, ADV, N и т.д.) отражает набор частей речи, которые выделяются для русского языка в рамках модели (прилагательное, наречие, существительное и т.д.). Далее, в поддереве, отвечающем значению N, представлены те категории, которые выделяются для существительных: падеж (CAS), число (NMB) и другие.

Одному «уровню» метаданных обычно соответствует несколько типов аннотаций. Например, на фонетическом уровне выделяется как минимум три типа аннотаций: аннотации-звуки, аннотации-слоги, аннотации-слова (фонетические). Значит, в плане автоматизации разметки возникают сразу три задачи: задача автоматического транскрибирования поэтических текстов, задача автоматической разбивки текста на слоги и задача автоматического разбиения текста на фонетические слова.

Исходный набор звуков (например, выделяемый в кн. [Аванесов 1972]) присутствует в модели в качестве различных значений атрибутов для типа аннотаций, отвечающего за звуки. После автоматического транскрибирования на каждом отрезке поэтического текста появляется набор таких аннотаций-звуков, соответствующих прочтению отрезка.

В модели присутствует тип тринотаций, соответствующий слогам. Каждая тринотация-слог содержит в себе одну или более аннотаций-звуков. Среди атрибутов тринотаций-слов следует вы-



делить очень важный для дальнейшего анализа атрибут, характеризующий ударность слога. Для этого атрибута можно ввести два или три значения (безударность и ударность; безударность, слабуюдарность и ударность). Ниже приводится пример разбивки на слоги. Ударные слоги отмечены символом '.

Для редактирования и пополнения информационной модели создан компьютерный инструментарий, а также система хранения и визуализации текстов и метаинформации, работающая в соответст-

вии с информационной моделью.

Таким образом, для решения поставленных задач необходимы:

- 1) разработка специальных алгоритмов, позволяющих автоматизировать рутинные процессы стиховедческого анализа;
- 2) создание компьютерных программ автоматизированной лингвостиховедческой разметки;
- 3) разработка и создание программно-информационной среды для представления результатов автоматизированного лингвостиховедческого анализа конкретных текстов.

Конечным практическим результатом работ должно стать создание корпуса русских поэтических текстов, снабженного лингвостиховедческой метатекстовой информацией. Создаваемая компьютерная среда должна обеспечить возможность поиска необходимых текстов и сегментов текста по заданным лингвостиховедческим параметрам. Такой подход позволяет, как уже говорилось, оптимизировать процессы аккумуляции стиховедческого знания: стиховеду уже не придется продельвать одни и те же подсчеты по несколько раз, а при использовании чужих данных их проверка должна существенно упроститься.

4. Переходим теперь от *общих* проблем автоматизации стиховедческих процедур к *специальным* алгоритмам и моделям, нацеленным на лингвостиховедческий анализ силлабо-тонических (в первую очередь русскоязычных) поэтических текстов.

Исходный пункт исследования — это идентификация и учет (в целях статистического и компаративного анализа) схемных ударений (ритмика иктов).

А. С. Старостин разработал программу автоматизированного акцентологического анализа, позволяющую регистрировать и квалифицировать ритмические формы в рамках установленного метрического инварианта (например, восемь ритмических форм 4-стопного ямба). Алгоритм работы этой программы сводится к следующему:

- 1) Производится предварительная обработка. В нее входит выделение в тексте трех видов тринотаций:
 - тринотаций-слов,

- тринотаций, соответствующих строкам стихотворного текста,
 - морфологических тринотаций.
- 2) С каждым словом текста соотносится набор потенциальных акцентных схем, информация о которых берется из словаря А. А. Зализняка.
 - 3) Акцентные схемы, сгенерированные на предыдущем шаге, комбинируются в рамках отдельных строк. Получаемые комбинации динамически соотносятся с заранее заложенными в систему шаблонами форм соответствующего стихотворного размера.
 - 4) Для каждой проанализированной строки возможны следующие варианты:
 - удается однозначно определить ритмическую форму;
 - удается предположить несколько ритмических форм;
 - не удается предположить никакой ритмической формы.

Следует отметить, что в подавляющем большинстве случаев форму определить удастся. Однако даже в тех случаях, когда это не так, сама по себе информация о неудаче представляет определенную ценность. Например, таким образом можно автоматически регистрировать отклонения от акцентных норм, зафиксированных в словаре Зализняка (ср. выше, п. 2.1).

Для разбивки на слоги можно использовать алгоритм, опирающийся на принцип «волны звучности» (или сонорности): вершину (ядро) слога нормально образует гласный (V), а в сочетаниях согласных внутри слога более близкую позицию к ядру занимает сонорный (R), а более дальнюю — шумный (T). Указанную тенденцию можно сформулировать как предпочтение универсальной схемы TRVRT. Любой вкладывающийся в эту схему слог (RV, TV, TRV, TVRT и т. д.) соответствует принципу «волны звучности», следуя которому, можно с высокой точностью разбить текст на слоги, после чего небольшой процент ошибок может быть легко исправлен вручную.

В дальнейшем планируется дополнить анализ ритмики иктов, во-первых, анализом ритмики сверхсхемных ударений и, во-вторых, анализом словораздельных вариаций стихотворной строки (Г. А. Шенгели называл их «ритмическими модуляциями») [Шенгели 1923; Гаспаров 1974, с. 15, 207–219] — то есть анализом важнейших ритмических факторов силлаботоники после ритмики схемных уда-

рений. Для того, чтобы автоматизированный анализ сверхсхемной и словораздельной ритмики стал возможен, необходимо решить проблему акцентологической разметки моно- и полисиллабических клитик (то есть графически выделенных слов, не несущих самостоятельного ударения). В конечном счете это позволит решить проблему автоматического разбиения текста на «фонетические» и «метрические» слова (см. ниже).

5. Существует два упрощенных подхода к ритмике моносиллабов — условно говоря, «стиховой» и «языковой». При первом подходе односложники — вне зависимости от их грамматической характеристики — считаются ударными, если попадают на икт, и безударными, если попадают на метрически безударное место. Вторым подходом, напротив, учитывает только грамматическую характеристику односложников: им приписывается ударность или безударность в соответствии со словарем А. А. Зализняка, без учета их места в стихе. В действительности моносиллабы должны быть разбиты не на две, а, как минимум, на три категории [Гаспаров 1974, с. 131–137]:

- 1) обязательно ударные,
- 2) обязательно безударные (собственно клитики),
- 3) атонирующиеся либо неатонирующиеся на метрически безударном месте в зависимости от синтаксической конструкции, в которую они входят (это «проблемные» клитики или, как их еще называют, «полуударные слова»)⁶.

Без правильного учета клитик невозможен подсчет ритмики словоразделов и сверхсхемных ударений (эту проблему сейчас изучают К. А. Головастиков и С. Е. Ляпин). Сложность проблемы клитик определяется тремя аспектами. Во-первых, наряду с клитиками, имеющими постоянное положение относительно опорного слова, существуют такие, которые могут находиться как в препозиции, так и в постпозиции по отношению к нему (*Шалун уж заморозил пальчик* и *Онегин верно ждет уж нас*). Во-вторых, слово может принимать либо не принимать ударение в зависимости от разных факторов: от места в стихе (на метрически ударном или метрически безударном слоге), от синтаксической роли, от места во фра-

⁶ Этим кругом вопросов (правда, не на стихотворном, а на прозаическом материале) занимается С. А. Крылов [Крылов 2006а, 2006b, 2008].

зе (под фразовым ударением или без него) и даже от частеречной принадлежности и, соответственно, реализуемого значения (среди служебных слов чрезвычайно развита омонимия). Наконец, очень усложняет автоматическую разметку клитик способность самостоятельной части речи атонироваться, а служебной — принимать ударение, как в примере *Со сна́ садится в ванну со́ льдом* (где первый предлог *со* безударен, а второй ударен).

Однако даже сумев правильно разметить фонетические слова, мы не решим окончательно те задачи, которые ставит перед нами наука о стихе. Стих накладывает на текст свое собственное, специфически стиховое (а не только общеязыковое) членение. Для подсчета ритмики словоразделов стиховеды ввели единицу, которую М. Л. Гаспаров условно назвал «метрическим словом» (это группа слогов, объединенных вокруг одного метрического ударения) [Гаспаров 1974, с. 144–145, 169–173 и др.]. Без введения этой единицы вряд ли возможно эффективно обобщать результаты подсчетов: при разбиении на фонетические слова в четырехстопном ямбе существует около двухсот словораздельных вариаций строки, а подсчет по метрическим ударениям сводит это число к сорока. Выделение метрических слов, обычно осуществляемое стиховедом интуитивно, должно быть формализовано с учетом разных факторов, играющих роль при интонационном членении поэтического текста, в первую очередь с учетом синтаксиса. (Так, в строке из «Евгения Онегина» *Давал три бала ежегодно* метрическим словом является словосочетание *три бала*, а вовсе не *Давал три*, поскольку синтаксическая связь между числительным *три* и существительным *бала* сильнее, чем между глаголом *Давал* и числительным *три*.)

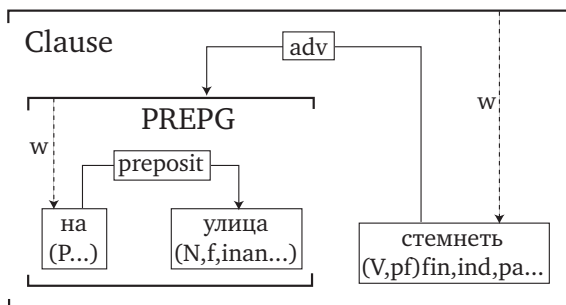
Для достижения значимых результатов в этой области необходимо в первую очередь автоматизировать процесс выделения фонетических слов, составив словарь постоянно-безударных служебных моносиллабов и изучив механизмы их присоединения к значимым словам. Затем должен быть разработан алгоритм выделения «метрических слов»: в цепочке графических слов каждые 1, 2 или более слов составляют слово фонетическое (при этом фонетические слова не пересекаются); в цепочке фонетических слов каждые 1, 2 или более слов составляют слово метрическое (при этом метрические слова не пересекаются). Таким образом, множество «метрических»

словоразделов мыслится как подмножество множества словоразделов «фонетических».

Следующим шагом в этом направлении должно стать использование компьютерной модели синтаксиса для выделения фонетических и метрических слов и анализа «глубины» словоразделов. Рассмотрение системы словоразделов значимо, так как ритм строки, помимо прочего, может зависеть (как считал, например, Р. О. Якобсон) от «метрической структуры заполняющих ее слов», то есть от «закономерностей распределения по строке словоразделов и их взаимодействия с системой оппозиций ударений» [Якобсон 1979]. Описание системы словоразделов будет неполным, если не учитывать, что словоразделы «неодинаковы по величине»: иерархия словоразделов (что показал еще Люсьен Теньер в «Основах структурного синтаксиса») «соответствует иерархии синтаксических связей» [Теньер 1988].

6. Из вышеизложенного ясно, что ключом к решению целого ряда лингвостиховедческих вопросов является автоматизированный поверхностно-синтаксический анализ стиха. Возможность автоматического построения набора синтаксических структур, реализованных в стихотворном тексте, позволит исследователям (помимо оптимизации процессов снятия омонимии) решать такие актуальные задачи современного стиховедения, как

- изучение распределения синтаксических связей внутри строки,
- автоматический подсчет силы межстрочных связей,
- изучение синтаксической организации строф и строфоидов.



Над компьютерной моделью русского синтаксиса и программой морфолого-синтаксического анализа работает А. С. Старостин. Программа оснащена графическим интерфейсом, который позволяет снабжать тексты тринотациями при помощи различных программных ресурсов (морфологического анализатора, фрагментатора и т.п.), а также редактировать отдельные тринотации вручную. Это даст возможность корректировать ошибки автоматических анализаторов. В системе предусмотрена возможность выгрузки текущего состояния разметки в текстовый файл формата XML. Благодаря этому результаты работы можно загружать в систему хранения аннотаций. В модели синтаксиса использована оригинальная формальная система, автором которой является А. С. Старостин [Мальковский, Старостин 2006] (идейно система восходит к работам А. В. Гладкого: [Гладкий 1985] и др.). Основные положения этой формальной системы хорошо выражаются в терминах тринотаций. Тринотация — это аннотация, которой приписан лес (множество корневых деревьев), в узлах которого стоят другие тринотации, а дугам приписаны синтаксические связи⁷. Дополнительно выделяется служебный тип связи, символизирующий то, что одна тринотация является составляющей другой. Для описанной структуры должен выполняться ряд аксиом, приведенных в [Мальковский, Старостин 2006]. На соседней странице приводится пример тринотации (пунктиром обозначены служебные связи).

Все допустимые имена связей перечислены в синтаксической модели. Кроме того, для спецификации сочетаемости тринотаций используется специальный формализм, представляющий собой систему декларативных правил. Каждое правило выражает один из двух смыслов: при условии, что некоторые тринотации удовлетворяют определенным ограничениям (например, согласуются по атрибутам), они могут соединяться связью либо в рамках некой уже существующей (контекстной) тринотации, либо в рамках некоторой новой тринотации (агрегат, составляющая), которая, в свою очередь, попадает внутрь контекстной тринотации. Пер-

⁷ Корневым деревом называют ориентированное дерево (связный ациклический граф), в котором выделяются узлы трех видов: *корень* (узел, в которой не заходит ни одна дуга), терминальные узлы (*листья*) и промежуточные (нетерминальные) узлы (*ветви*).

вый вариант соответствует обычному для модели дерева зависимостей выделению зависимости между единицами, а второй соответствует обычному для модели составляющих выделению агрегата, который наделяется собственными синтаксическими атрибутами.

О полнофункциональной реализации модели говорить пока рано, но первоначальное наполнение системы синтаксическими правилами уже произведено. Предстоит тщательная формализация тонких моментов русского синтаксиса, таких, как, например, поведение частиц. Особый интерес для лингвостиховедческой проблематики представляют те синтаксические явления, которые характерны для русской поэтической речи в большей степени, чем для прозаической.

7. Если взглянуть на ритмический анализ стихотворной строки с точки зрения возможностей его автоматизации, то в нем можно выделить четыре основных этапа:

- 1) расстановка ударений на основании морфологического анализа с автоматическим снятием омонимии при помощи синтаксического анализа и сопоставления ритмических гипотез с метрической схемой;
- 2) регистрация сверсхемных ударений и выделение фонетических слов на основании идентификации клитик и анализа ударности моносиллабов;
- 3) выделение метрических слов путем соотнесения фонетических словоразделов с метрической схемой;
- 4) измерение относительной глубины словоразделов на основании синтаксического анализа строки и ее контекстного окружения.

Анализ внутристрочных синтаксических связей должен идти еще в одном направлении: в сторону регистрации повторяющихся частеречевых последовательностей и других ритмико-грамматических формул и клише.

Современная терминология, описывающая эти явления, была предложена М. Л. Гаспаровым [Гаспаров 1986] и несколько модифицирована М. И. Шапиром [Шапир 2000, с. 380], который пред-

ложил различать:

- 1) *ритмические формулы* (в образовании которых участвуют ритм и лексика),
- 2) *ритмико-грамматические клише* (в образовании которых участвуют ритм и грамматика),
- 3) *ритмико-грамматические формулы* (в образовании которых участвуют ритм, грамматика и лексика),
- 4) *тавтологические стихи*.

Автоматическая регистрация и поиск перечисленных явлений позволит сопоставлять по заданным и свободным параметрам одноязычные и разноязычные тексты и наборы текстов (например, оригинал и пародии на него, подлинник и его переводы, подлинники и дубильные произведения и т.д.).

Помимо внутрисклонового синтаксиса в дальнейшем необходимо будет перейти к автоматизированному анализу силы межстрочных связей и — на основе последнего — к автоматическому анализу синтаксической организации строф и строфоидов. Эта проблема была поставлена в пионерском исследовании Г. О. Винокура [Винокур 1941], а новые подходы к ее изучению были найдены М. Л. Гаспаровым, Т. В. Скулачевой, М. Г. Тарлинской и М. И. Шапиром [Гаспаров, Скулачева 1993, 1999; Гаспаров 2001; Тарлинская 2000; Шапир 1999а, 2000, с. 163–164, 2003]. Представляется, что автоматизированная синтаксическая разметка стиха (с последующим филологическим редактированием) позволит поставить вопрос об измерении силы грамматических связей на более прочную почву и ответить на целый ряд вопросов, сформулированных учеными — первооткрывателями этого научного направления.

ЛИТЕРАТУРА

- Аванесов 1972—Аванесов Р. И. Русское литературное произношение. Изд. 5-е, перераб. и доп. М., 1972.
- Вигурский, Пильщикова 2003—Вигурский К. В., Пильщикова И. А. Филология и современные информационные технологии: (К постановке проблемы) // Известия РАН. Серия литературы и языка. 2003. Т. 62, № 2. С. 9–16.

- Винокур 1941—*Винокур Г.* Слово и стих в «Евгении Онегине» // Пушкин: Сборник статей / Под ред. А. Еголина. М., 1941. С. 155–213.
- Гаспаров 1974—*Гаспаров М. Л.* Современный русский стих: Метрика и ритмика. М., 1974.
- Гаспаров 1986—*Гаспаров М. Л.* Ритмико-синтаксическая формульность в русском 4-стопном ямбе // Проблемы структурной лингвистики 1983. М., 1986. С. 198.
- Гаспаров 2001—*Гаспаров М. Л.* Синтаксическая структура стихотворной строки // Славянский стих: Лингвистическая и прикладная поэтика. Материалы международной конференции 23–27 июня 1998 г. М., 2001. С. 121–129.
- Гаспаров, Скулачева 1993—*Гаспаров М. Л., Скулачева Т. В.* Ритм и синтаксис в свободном стихе // Очерки истории языка русской поэзии XX века: Грамматические категории; Синтаксис текста. М., 1993. С. 20–43.
- Гаспаров, Скулачева 1999—*Гаспаров М. Л., Скулачева Т. В.* Синтаксис четырехстопного полноударного ямба // Поэтика. История литературы. Лингвистика: Сборник к 70-летию Вяч. Вс. Иванова. М., 1999. С. 93–101.
- Гладкий 1985—*Гладкий А. В.* Синтаксические структуры естественного языка в автоматизированных системах общения. М., 1985.
- Крылов 2006a—*Крылов С. А.* Фонетическое слово и его корреляты в русском письменном тексте (с точки зрения корпусной лингвистики) // Корпусная лингвистика-2006. Труды международной конференции, 10–14 октября 2006 г. СПб., 2006. С. 190–200.
- Крылов 2006b—*Крылов С. А.* Фонетическое слово и его корреляты в русском письменном тексте (с точки зрения корпусной лингвистики) // Пятая выездная школа-семинар «Порождение и восприятие речи». Череповец, 2006. С. 66–96.
- Крылов 2008—*Крылов С. А.* О частотном словаре фонетических слов // Фонетика и нефонетика: К 70-летию Сандро В. Кодзасова. М., 2008. С. 387–399.
- Мальковский, Старостин 2006—*Мальковский М. Г., Старостин А. С.* Модель синтаксиса в системе морфосинтаксического анализа «Treetop» // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» (Бекасово, 31 мая — 4 июня 2006 г.). М., 2006. С. 481–492.

- Перцов 2000—*Перцов Н. В.* О неоднозначности в поэтическом языке // Вопросы языкознания. 2000. № 3. С. 55–82.
- Перцов 2006—*Перцов Н. В.* Об одном случае акцентной вариантности в русском литературном языке первой половины XIX века // Известия РАН. Серия литературы и языка. 2006. Т. 65, № 5. С. 50–51.
- Теньер 1988—*Теньер Л.* Основы структурного синтаксиса / Пер. с франц. под общ. ред. В. Г. Гака. М., 1988. С. 37–38.
- Шапир 1995—*Шапир М. И.* «Versus» vs «prosa»: пространство-время поэтического текста // Philologica. 1995. Т. 2, № 3/4. С. 7–47.
- Шапир 1999a—*Шапир М. И.* Ритм и синтаксис ломоносовской оды: (К вопросу об исторической грамматике русского стиха) // Поэтика. История литературы. Лингвистика: Сборник к 70-летию Вяч. Вс. Иванова. М., 1999. С. 55–79.
- Шапир 1999b—*Шапир М. И.* Язык поэтический // Введение в литературоведение: Литературное произведение: основные понятия и термины. М., 1999. С. 6.
- Шапир 2000—*Шапир М. И.* Universum versus: Язык — стих — смысл в русской поэзии XVIII—XX веков. М., 2000. Кн. 1.
- Шапир 2003—*Шапир М. И.* Три реформы русского стихотворного синтаксиса: (Ломоносов — Пушкин — Иосиф Бродский) // Вопросы языкознания. 2003. № 3. С. 31–78.
- Шапир 2005—*Шапир М. И.* «Тебе числа и меры нет»: О возможностях и границах «точных методов» в гуманитарных науках // Вопросы языкознания. 2005. № 1 (см. по расшир. переизд. в кн.: *Ярхо Б. И.* Методология точного литературоведения: Избранные труды по теории литературы. М., 2006. С. 883–885).
- Шенгели 1923—*Шенгели Г.* Трактат о русском стихе. Изд. 2-е, перераб. М.; Пг., 1923. Ч. I: Органическая метрика. С. 38, 57 сл., 138 сл.
- Яглом 1980—*Яглом И. М.* Математические структуры и математическое моделирование. М., 1980. С. 13–14.
- Якобсон 1979—*Якобсон Р.* Ретроспективный обзор работ по теории стиха [1979] / Пер. с англ. М. Л. Гаспарова // *Якобсон Р.* Избранные работы. М., 1985. С. 255.
- Empson 1930—*Empson W.* Seven Types of Ambiguity. London, 1930.