

Александр Костыркин

Институт Востоковедения РАН

1. Введение

Работа посвящена проблеме *синтаксического анализа* предложений на естественном языке и нацелена на построение компьютерной модели синтаксического анализа для системы Японско-Русского Автоматического Перевода ЯРАП при использовании сущностного подхода к организации лингвистических знаний в их формальном описании [Шаляпина 99; Шаляпина и др. 2000].

В центре нашего внимания будет находиться явление *структурной неоднозначности*^[2] предложения, как одно из важнейших препятствий, стоящих на пути автоматизации анализа текста.

В работе предлагается новый взгляд на эту проблему, позволяющий рассматривать ее с позиции теории *задач удовлетворения ограничений* (*constraint satisfaction problems*). Предлагается (i) использовать математический формализм этой теории для внутреннего представления лингвистических знаний и структуры предложения в процессе анализа и (ii), что составляет новизну подхода, проводить синтаксический анализ и решать задачу снятия структурной неоднозначности методом так называемых *активных возвратов*, разработанным в теории задач с ограничениями.

Работа построена по следующему плану: вводятся основные требования к процедуре синтаксического анализа (разд. 1), рассматриваются проблемы, возникающие при анализе структурно-неоднозначных предложений (разд. 2), основные понятия и методы из теории задач удовлетворения ограничений (разд. 3), переход от лингвистического описания к формализму задач удовлетворения ограничений (Разд. 4).

1. Постановка задачи

Говоря о *синтаксическом анализе предложения*, будем иметь ввиду формальную процедуру, которая в соответствии с заданной лингвистической моделью анализа строит для предложения его структурное представление в виде дерева зависимостей в традиционном понимании этого конструкта. Мы отвлекаемся от места процедуры синтаксического анализа среди прочих аналитических операций и ее связей с ними, но предъявляем к ней некоторые внешние требования: процедура анализа должна принимать на входе последовательность словоформ, каждой из которых сопоставлены все их возможные лексико-морфологические (ЛМ-) интерпретации; на выходе должно быть построено по возможности полное дерево зависимостей, в котором каждой словоформе присвоена ровно одна ЛМ-интерпретация (пример см. в разд. 4).

Ниже в обобщенном виде будет предложена модель синтаксического анализа, в которой мы стремились насколько возможно близко воспроизвести стратегию языкового поведения человека.

Особый интерес здесь представляет явление синтаксической неоднозначности. В своей модели анализа мы исходим из того предположения, достаточно широко аргументированного в области психолингвистики, что в *подавляющем большинстве случаев* человек понимает текст однозначно и с первого раза, чаще всего не замечая того, что текст допускает множество различных интерпретаций. При этом нельзя сказать, что альтернативные прочтения отвергаются им, как неподходящие по контексту: они просто не актуализируются; иначе коммуникативное использование языка стало бы просто невозможно.

При этом есть одно интересное исключение, которое заставляет нас делать оговорку "в подавляющем большинстве случаев", и оно заслуживает особого внимания. Имеется ввиду анализ так называемых "предложений заблуждения" (garden path sentences [3]).

2. Предложения заблуждения

Это структурно неоднозначные предложения, неоднозначность в которых при первом прочтении приводит читателя к ошибочной интерпретации. Ее ошибочность обнаруживается, когда в ходе чтения появляются структурно и/или семантически несовместимые с ними фрагменты, а вместе с тем и информация, необходимая для исправления интерпретации уже пройденного неоднозначного фрагмента. Например, во фразе, взятой из [Sturt et al. 99],

The photographer accepted the money could not be sent yet [4].

до появления слова *money* включительно строится вполне приемлемая интерпретация "фотограф принял деньги", но дальше обнаруживается нестыковка: глагол *could* вместе с последующей группой слов не могут быть включены в уже построенную структуру. Это сигнал того, что начальный анализ не удался и его требуется исправить, проведя *повторный анализ*. В данном случае исправление будет состоять в повторном анализе многозначного слова "ассерт". В своем втором значении - "признавать" - оно может подчинять себе придаточное предложение, которое интерпретируется как "содержание признаваемого факта". На рис.1 перечислены основные этапы анализа: на втором этапе анализ заходит в тупик, на третьем происходит возврат к неоднозначному фрагменту, на четвертом исправляется ошибка и пятый этап успешно и корректно завершает анализ.

Рис.1. Поэтапная развертка анализа показывает как происходит корректировка с изменением интерпретации второй словоформы. В последней строке, отвечающей успешному завершению анализа, подчеркнуты слова, которые при корректировке не пришлось переинтерпретировать.

В общем случае при повторном анализе "предложения заблуждения" встает несколько вопросов:

- 1) в чем причина неудачи начального анализа;
- 2) какую часть уже построенной интерпретации требуется откорректировать, а какую оставить без изменений;
- 3) какого типа лингвистическую информацию нужно привлечь для выявления ошибки и какую для ее исправления;
- 4) в чем именно должны состоять исправления;
- 5) как определить степень сложности исправления (его "стоимость") и
- 6) как удачно завершить анализ.

Существует множество синтаксических и психолингвистических теорий, которые пытаются ответить на эти вопросы (см. обзорные статьи [Секерина 96], [Altmann 89]), но по утверждению Лина Фрейзера, одного из самых авторитетных исследователей в этой области, "до сих пор не было предложено ни одной полной и эксплицитной модели повторного анализа" [Frazier & Clifford 96]. Главная причина этого — в комплексности структурных преобразований при повторном анализе и в сложности их формализации. Кроме того, "серьезная слабость многих теорий структурных изменений в том, что они не подвергались экспериментальной проверке и полагаются на одни лишь интуитивные данные" [Sturt et al. 99: 136].

Можно подытожить: предложения заблуждения создают ситуацию, где достаточно очевидным оказывается механизм анализа, которым пользуется человек, и который мы хотим смоделировать. Этот анализ заключается в *поиске* правильной интерпретации путем последовательного *перебора* вариантов в порядке их предпочтительности с точки зрения анализирующего субъекта (вопрос определения этой предпочтительности выходит за рамки настоящего исследования). Оказывается, что это очень избирательный и исключительно эффективный поиск. Относительно него можно выдвинуть следующие гипотезы:

- 1) к перебору прибегают лишь в критических случаях,

2) при поиске руководствуются принципом максимального сохранения построенной интерпретации,

3) поиск проходит не на всем пространстве предложения, а на ограниченных участках, и именно там где он наиболее перспективен.

Очевидно, что стратегия поиска при этом существенно зависит от языковой компетенции — чем хуже человек владеет языком, тем чаще он "заблуждается" и прибегает к повторному анализу.

Исходя из этих соображений, для моделирования анализа целесообразно использовать механизм, позволяющий осуществлять поиск синтаксических структур с минимальным перебором, при этом оптимальную стратегию должен подсказывать лингвистический аппарат, а не архитектурная специфика анализатора.

В качестве такого механизма мы предлагаем использовать алгоритмы поиска, разработанные в теории *задач удовлетворения ограничений*.

3. Задачи поиска в ограничениях

Задача удовлетворения ограничений (Constraint Satisfaction Problem, далее CSP) в обобщенной формулировке традиционно ставится следующим образом.

Определение 1. CSP-задачей (V, C) будем называть конечное множество *переменных* V с конечными областями определения, для которого задано множество *ограничений* C ; ограничения определяют допустимые сочетания значений переменных. Требуется найти значения всех переменных, удовлетворяющие всем ограничениям, или установить, что такого решения не существует.

Определение оперирует минимумом сущностей достаточно общего характера, и это придает ему большую выразительную силу. В терминах переменных и ограничений могут быть представлены такие разные задачи как составление расписаний, размещение объектов на плоскости, оптимизация маршрутов, машинное зрение, диагностика оборудования и пр. Цель дальнейшего изложения — показать, что к этому списку можно добавить и задачу синтаксического анализа.

В общем случае решение CSP-задач требует экспоненциального перебора, поэтому суть всех методов, разработанных для их решения, сводится к попыткам уменьшить вычислительные затраты за счет оптимальных стратегий поиска, основанных на эвристиках. Наиболее известны такие алгоритмы, как *поиск с направленными возвратами* (directed backtracking), *с прыжками назад* (backjumping), *с предварительным планированием* (lookahead), *с опережающей проверкой* (forward checking) и пр. Отдельные классы составляют метод *согласования дуг* (arc consistency) и параллельные алгоритмы. Классификацию и сравнение подходов см. в обзоре [Vipin Kumar 92].

Рассмотрим алгоритм поиска *с активными возвратами* (dynamic backtracking), предложенный Мэттью Гинсбергом [Ginsberg 93][5]. Принципы перебора, положенные в его основу, удачно согласуются с теми требованиями к синтаксическому анализу, которые мы сформулировали в предыдущем разделе.

Для иллюстрации главных идей этого алгоритма удобно использовать задачу раскраски географической карты, которая также может быть представлена как CSP-задача.

Дана карта со множеством стран на ней; для каждой пары стран известно, есть между ними граница или нет; дан также конечный набор различных красок. Требуется закрасить каждую из территорий одним цветом, соблюдая условие: каждые две соседние страны должны быть разного цвета. В этой задаче множество стран — это переменные; область определения каждой из переменных — множество цветов; запрет на присвоение двум соседним странам одного цвета является ограничением на значения переменных.

Пусть, например, дана карта как на рис.2-а и имеются три краски: синяя, красная и зеленая. Одно из возможных решений представлено на рис.2-б.

Определение 2. Для данной CSP задачи (V, C) гипотезой называется упорядоченное подмножество переменных из V , которым сопоставлены значения, не нарушающие ограничений.

Например, на рис.2-а представлена пустая гипотеза $\emptyset$, на рис.2-б - гипотеза $(A=синий, B=красный, C=синий, D=красный, E=зеленый)$, которая является решением данной задачи.

Рис.2. Карта пяти стран - A, B, C, D и E. Каждой из стран нужно присвоить цвет, отличный от цветов соседей.

Алгоритм начинает с пустой гипотезы и пытается расширить ее до полного решения, присваивая переменным разрешенные значения. Может оказаться, что из-за ограничений некоторую переменную невозможно связать. Это говорит об ошибочности построенной гипотезы и необходимости ее исправить.

Допустим, что для карты на рис.2-а мы получили гипотезу (А=синий, В=красный, С=зеленый, D=красный), см. рис.3. Ограничения не позволяют раскрасить территорию Е и завершить решение. Такая ситуация, когда вся область определения переменной попадает под ограничения, называется *конфликтом*.

Рис.3. Гипотеза, которую невозможно расширить — для Е не осталось разрешенных цветов. Это тупиковая ветвь решения.

Преимущество алгоритма Гинсберга по сравнению с другими разработками состоит в методе разрешения конфликтных ситуаций.

Для прояснения сути проблемы покажем, как бы в данном примере работал известный алгоритм поиска в глубину с возвратами (depth-first search). Его решением (см. рис.4) было бы вернуться на шаг назад — к переменной D и попробовать исправить ее значение (этап 2), получить еще одну конфликтную ситуацию, вернуться еще на шаг назад, сменить значение для С (этап 3) и только тогда завершить поиск (этап 4).

Очевидно, что реальный источник конфликта в этом примере — переменная С, так что первоначальное значение D менять не имело смысла. Простой поиск с возвратами оказывается крайне неэффективен — на исправление ошибки уходит слишком много времени, при этом правильная часть решения может быть напрасно разрушена.

Идея алгоритма Гинсберга, соответствующая естественной человеческой интуиции, состоит в следующем: в ошибочное решение целесообразно вносить локальные исправления и в первую очередь обращаться к менее благополучным фрагментам. А именно: при попадании в конфликтную ситуацию можно указать предположительных "виновников" конфликта, т.е. те переменные, значения которых привели к нарушению ограничений. Если возвращаться напрямую к этим переменным, сохраняя промежуточные удачные фрагменты решения, то можно получить двойной выигрыш: во-первых, возврат к месту ошибки не будет отягощен обработкой не причастных к конфликту переменных, во-вторых, после исправления ошибки не придется повторно выполнять уже однажды сделанное вычисление.

Согласно этим принципам исправление раскраски карты на рис.3 будет выглядеть следующим образом:

Рис.5. Причиной тупиковой ситуации (этап 1) объявляется значение переменной С. Применение направленного возврата с сохранением значения промежуточной переменной D (этап 2) дает максимально быстрое решение.

Формализация таких очевидных соображений представляет большую сложность, поэтому при интуитивной простоте главная идея алгоритма требует достаточно нетривиальной реализации. Но это компенсируется тем, что предложенный Гинсбергом механизм позволяет привлекать к вычислению самые разнообразные эвристики выявления *вероятных* виновников конфликта при условии, что эти эвристики определяют всех *действительных* виновников.

Обратимся к переменной Е на рис.3. В ее области определения {красный, синий, зеленый} нет ни одного разрешенного контекстом значения. Запрет каждого из значений можно попытаться объяснить, установив ту переменную из построенной гипотезы, со значением которой конфликтует это значение. В данном случае применима такая эвристика: обратиться к непосредственным соседям территории Е. Значение "синий" невозможно выбрать из-за переменной А, значение "красный" конфликтует с переменной В, наконец запрет на "зеленый" объясняется значением переменной С.

Для корректировки гипотезы выбирается та переменная из списка "обвиняемых", которая была связана последней — переменная С. Ее текущее значение "зеленый" отменяется и на него налагается запрет, объяснением которого будет оставшаяся часть списка обвиняемых (А, В). Логически это те переменные с которыми значение "зеленый" оказалось в опосредованном конфликте; у территорий А и С нет общей границы. После этой операции текущая гипотеза принимает вид (А=синий, В=красный, D=красный), см. рис.5-этап 2. Теперь в любом порядке можно взять оставшиеся переменные с их разрешенными значениями — "синий" для С и "зеленый" для Е — и завершить решение.

Полученный список (А, В, С) является, по сути, объяснением конфликта — это та часть гипотезы, где предположительно была допущена ошибка. Согласно данной эвристике в

список не попала переменная $D=$ красный, в другом эвристическом рассуждении этот список мог бы оказаться другим. Ясно, что если эвристика всегда в качестве объяснения конфликта выдает все переменные из текущей гипотезы, в примере это было бы (A, B, C, D) , то она совершенно неинформативна и возвращает нас к простому поиску в глубину с возвратами. Если же окажется, что объяснение не содержит ни одной переменной, это будет означать безусловный запрет и невозможность разрешить конфликт. Это происходит когда задача не имеет решения.

Для практической реализации важно, что алгоритм обладает свойствами полноты и корректности, а необходимое пространство равно $O(v^2d)$, где v - количество переменных, а d - размер наибольшей из областей определения переменных.

Скорость поиска зависит от того насколько точно эвристическая процедура объяснения конфликта определяет "виновные" переменные. Чем слабее эвристика, тем больше "шума" она дает, и находить реальных виновников в этом "шуме" приходится опять же перебором. Перебором, для сокращения которого и используются эвристики. Не исключена также и ситуация, когда повышение точности работы эвристики приводит к тому, что ее вычислительная сложность превышает сложность исходной задачи. Поэтому встает дополнительная проблема: обеспечить возможность эффективно балансировать между стоимостью короткого перебора с дорогой эвристикой и долгого перебора с дешевой эвристикой.

Архитектурно алгоритм Гинсберга состоит из двух логически связанных компонентов. Первый отвечает за полноту и корректность перебора в пространстве гипотез, второй — эвристический компонент — направляет поиск в этом пространстве по кратчайшему пути. В отсутствие второго компонента алгоритм работает как простой поиск в глубину с возвратами, т.е. как показано в примере на рис. 4.

Уже при поверхностном знакомстве с CSP-задачами и методами их решения обнаруживается сходство с проблемами из области синтаксического анализа. Аналогия становится более очевидной при обращении к методу Гинсберга — устройство его поискового механизма как нельзя лучше отвечает нашему взгляду на синтаксический анализ. Эвристический компонент соответствует нашему пониманию роли лингвистического аппарата — он призван сокращать перебор гипотез. (В нашем случае гипотезой будет синтаксическая структура для фрагмента предложения.) Без лингвистических знаний не находится сдерживающих перебор факторов. Важно также, что распределение вычислительной нагрузки на компоненты алгоритма Гинсберга отражает те роли, которые выполняют механизм перекомпоновки структуры и лингвистическая информация при корректировке предложений заблуждения — при любых обстоятельствах переборный механизм гарантирует исправление ошибки, стоимость же исправления зависит от качества языкового описания.

4. Предлагаемый подход

Будем представлять предложение последовательностью графических сегментов, выступающих при анализе в виде своих лексико-морфологических (ЛМ-) интерпретаций — ЛМ-представлений словоформ как будущих узлов структуры зависимостей. Будем считать далее, как это предполагается при сущностной организации лингвистического описания, что в структуре зависимостей, отвечающей такой последовательности сегментов, каждая связь между двумя сегментами реализует валентность какого-либо из этих сегментов, и все лингвистические сведения об этой связи задаются в описании реализуемой ею валентности.

Рассмотрим пример анализа структурно-неоднозначного предложения *Pupils of the teacher dilated* "зрачки учителя расширились". Допустим, что в ходе разбора для четырех из пяти сегментов предложения был построен фрагмент дерева зависимостей, показанный на рис.6. В целях наглядности мы опускаем некоторые несущественные здесь лингвистические детали представления узлов дерева, отождествляя их с сегментами.

Рис.6. Построена частичная структура предложения, которой соответствует интерпретация "ученики учителя".

Будем считать, что построение этого фрагмента дерева включало кроме прочих следующие аналитические операции:

- 1) для сегмента PUPILS из двух вариантов лексико-морфологической интерпретации *pupils1* "ученики" и *pupils2* "зрачки" был выбран первый вариант;
- 2) сегмент OF интерпретируется как двухвалентный предлог, первая из его валентностей заполняется PUPILS в значении "ученики", а вторая — TEACHER в значении "учитель";
- 3) артикль, сопоставленный сегменту THE, подчиняется TEACHER по своей единственной валентности.

На следующем шаге анализа в фокус попадает сегмент DILATED, которому можно сопоставить ЛМ-интерпретации *dilate1* "расширился" и *dilate2* "расширил". У непереходного и у переходного вариантов этого глагола есть обязательная валентность с семантическим ограничением на объект, способный расширяться. Так как в текущей интерпретации предложения нет единицы с требуемой семантической характеристикой, то завершить данную линию анализа не удастся. Это уже знакомая нам тупиковая ситуация, когда необходим *повторный анализ*, все сложности которого, описанные в разделе 2, остаются верны и для данного простого примера.

Предлагаемое ниже формальное построение позволит нам получить комплексное решение проблемы прямого и повторного анализа.

Будем считать, что

- 1) все сегменты предложения являются переменными. Их области определения — соответствующие ЛМ-варианты. Например, для PUPILS это $\{pupils1, pupils2\}$.
- 2) валентности каждого ЛМ-варианта являются переменными. Их области определения — множество сегментов во всех их ЛМ-интерпретациях. Иными словами, стрелка в дереве зависимости отождествляется со связанной переменной. Далее говоря "валентность" и "сегмент", будем иметь ввиду переменные соответствующих типов.
- 3) Синтаксическая структура предложения, которая строится из множества переменных двух типов — сегментов и валентностей, — соответствует понятию гипотезы в смысле определения 2 в разд.3 (см.).
- 4) Записанные в лингвистическом описании лексические, морфолого-синтаксические и семантические требования к единицам, способным заполнять валентности, являются ограничениями на значения этих валентностей.
- 5) Невозможность связать обязательную валентность некоторого ЛМ-варианта запрещает этот ЛМ-вариант в качестве значения переменной-сегмента.
- 6) Допустимые варианты взаимного линейного расположения единиц также задаются как ограничения.
- 7) Требования древесности и проективности являются ограничениями на комбинации значений переменных-валентностей.

В этих условиях задача синтаксического анализа будет звучать следующим образом: найти значения всех сегментов и значения (возможно не всех) валентностей такие, что они удовлетворяют всем ограничениям на заполнители этих валентностей, а также на их комбинации.

Первый существенный результат такого построения состоит в следующем: в теории предложений заблуждения само описание тупиковых ситуаций представляло большую сложность из-за комплексности структурных причин, препятствующих продолжению анализа. Теперь мы можем рассматривать эту ситуацию как конфликт между переменными. Конфликт, напоминаем, состоит в том, что в текущей гипотезе ограничения не позволяют связать данную переменную. Алгоритм Гинсберга обязывает объяснять каждый из обнаруживаемых запретов на некоторое значение путем указания на те переменные в уже построенной структуре, значения которых предположительно привели конфликту с данным значением (и тем самым к его запрету).

В нашей постановке задачи возможны конфликты двух типов: (1) запрет наложен на все возможные значения валентности, т.е. на всех кандидатов в ее заполнители, (2) под запретом оказались все ЛМ-интерпретации сегмента.

В обоих случаях объяснение конфликта будет представлено списком переменных — указанием на предположительно ошибочные интерпретации сегментов и установленные между ними связи. В примере, описанном выше, сегменту DILATED в интерпретации "расширился" сопоставлена единственная валентность, которую не удалось связать. Это конфликт первого типа.

В терминах переменных и ограничений при наличии разумной эвристической процедуры причиной этого конфликта может быть указание на значение самого сегмента DILATED и на сегмент PUPILS. Содержательно это значит, что либо

- (i) дана неверная интерпретация *dilated1* и следует выбрать альтернативу *dilated2*, либо

- (ii) использована неверная интерпретация первого сегмента, и необходимо заменить его на *pupils2* "зрачки", либо
- (iii) верно и то и другое.

Так или иначе, анализатору достаточно этой информации, чтобы разрешить конфликт — освободить переменную DILATED в значении *dilated1* и объяснить этот запрет указанием на PUPILS. После того, как и вторая альтернатива *dilated2* на приведет к успеху, оба значения переменной DILATED окажутся под запретом. Это уже конфликт второго типа. Его объяснением станет вычисленное выше указание на PUPILS. Значит следующим шагом нужно будет вернуться к первому сегменту, перешагнув через "OFTHE TEACHER". После выбора интерпретации *pupils2* "зрачки", алгоритм снова обратится к концу предложения и успешно завершит анализ.

Таким образом, второе важное следствие приведенного выше построения состоит в возможности разрешения конфликтов при точной модификации (в чем бы она ни состояла) только необходимой части структурной интерпретации.

5. Эксперименты и выводы

Для экспериментальной проверки данного подхода на Прологе были реализованы два алгоритма — алгоритм поиска в глубину с простыми возвратами и алгоритм Гинсберга. В таблице 1 представлены результаты работы обоих алгоритмов на разных входах. Для эксперимента подбирались простые для анализа предложения, и формально сходные с ними предложения заблуждения.[\[6\]](#)

1. *Pupils of the dean*

	сделано возвратов	проверено переменных	время, сек.
Простой поиск	3	24	4.78
Гинсберг	3	24	3.30

2. *Pupils of the dean dilate pupils*

	сделано возвратов	проверено переменных	время, сек.
Простой поиск	8	42	11.65
Гинсберг	8	42	12.75

3. *Pupils of the teacher dilated* *

	сделано возвратов	проверено переменных	время, сек.
Простой поиск	71	164	40.65
Гинсберг	12	49	12.63

Таблица 1. Знаком (*) отмечено предложение трудное для анализа, в том числе и для человека.

Первое и второе предложение анализируются за один проход. В структуре не возникает значительных конфликтов, которые потребовали бы радикального пересмотра решения. Первый и второй алгоритмы не дают значительных различий при анализе таких предложений. Фраза 3, наоборот, представляют сложность для анализа — для корректировки решения "ученики учителя" требуется возврат к самому первому узлу, что при поиске с простыми возвратами приводит к попыткам перестроить интерпретацию для середины предложения, в то время как поиск с активными возвратами сразу обращается к первому слову, оставляя промежуточную структуру как она есть.

Проведенные эксперименты говорят о работоспособности предложенной техники анализа и показывают, что на трудных для анализа предложениях она может давать значительные преимущества.

В ходе анализа могут встретиться самые различные затруднительные положения, например:

— выбраны неверные интерпретации нескольких сегментов, а вместе с ними установлены ошибочные связи в том числе и с корректными фрагментами структуры;

— для валентности выбран не тот заполнитель, и это не позволило построить дерево без нарушения проективности;

— две валентности одного слова претендуют на один и тот же заполнитель;

— наконец, просто не удастся построить связанное дерево зависимостей.

Все эти и многие другие конфликтные ситуации получают в нашем подходе точное описание и средства их эвристического объяснения.

Заметим, что в других разработанных к настоящему времени системах представления и обработки предложений заблуждения подобная диагностика невозможна. Более того, данное описание предложений является первой попыткой рассмотрения указанных проблем в рамках грамматики зависимостей; все предыдущие исследования проводились в терминах составляющих. Разумеется, решение данной проблемы потребует еще значительных усилий и решения многих сложных вопросов — как в области лингвистического описания, так и в

сфере компьютерной реализации. Думается, однако, что исследование в этом направлении достаточно перспективно.

Литература

Секерина 96 — И. А. Секерина "Американские теории синтаксического анализа предложения в процессе понимания" // *Вопросы языкознания*, №3, 1996, с. 100-138.

Шаляпина 99 — З. М. Шаляпина "Оппозиция «Часть-целое» и сущностный подход к моделированию языковой компетенции" // *Роман Якобсон: тексты, документы исследования*, М., Издательский центр РГГУ, 1999, с.541-551.

Шаляпина и др. 2000 — З. М. Шаляпина, М. И. Канович, А. В. Костыркин, Л. С. Модина "Лексико-синтаксическая информация в словаре для автоматического перевода (в рамках сущностного подхода к лингвистическому описанию)" // *Материалы 3-й международной школы-семинара "Словарь в современном мире"*, — Иваново, 2000 [в печати]

Altmann 89 — Gerry T. M. Altmann "Parsing and Interpretation: An Introduction" // *Language and Cognitive Processes* Vol.4, No.3/4, 1989, p.1.

Frazier & Clifford 96 — Frazier L., Clifford C. "Construal", 1996.

Ginsberg 93 — Ginsberg M. L. "Dynamic Backtracking" // *Journal of Artificial Intelligence Research*, '1, 1993, p. 25-46.

Kurohasi & Nagao 96 — Sadao Kurohasi and Makoto Nagao "Structural disambiguation in Japanese by case structure evaluation based on examples in a case frame dictionary" // *Recent advances in parsing technology*, Kluwer Academic Press, 1996. (Hury Bunt and Masaru Tomita, eds.)

Sturt et al. 99 — Patrick Sturt, Martin J. Pickering, Matthew W. Crocker "Structural Change and Reanalysis Difficulty in Language Comprehension" // *Journal of Memory and Language*, Vol.40, No.1, January 1999, p.136-165.

Tür & Oflazer 98 — Gökan Tür and Kemal Oflazer "Tagging English by Path Voting Constraints" // *COLING-ACL' 98, Vol.2, Montreal*, 1998, p.1277-1281.

Vipin Kumar 92 — Vipin Kumar "Algorithms for Constraint Satisfaction Problems: A Survey" // *Artificial Intelligence Magazine*, Vol.13, '1, 1992, p.32-44.

White 95 — Christopher M. White "Converting Context-free Grammars to Constraint Dependency Grammars" // A Master of Science Thesis, Purdue University, August 1995.

[1] Исследование выполняется при частичной финансовой поддержке РФФИ (проект ? 98-01-00539)

[2] Другие термины - многозначность, омонимия, полисемия - у разных авторов по-разному соотносятся друг с другом и получают разные по объему значения. В настоящей работе для

описания синтаксиса выбран термин "неоднозначность" как наиболее общий и не конфликтующий с аналогичными терминами из лексикологии и лексикографии.

[3] Этот тип предложений получил такое название из-за аналогии с тем как человек блуждает по тупиковым дорожкам парка в поисках выхода.

[4] Хотя работа нацелена на анализ японского языка, для наглядности все языковые примеры даются на английском.

[5] Для экономии места мы не приводим алгоритм в полной форме, отсылая читателя к первоисточнику.

[6] Лингвистическое описание ЛМ-вариантов входящих в эти предложения сегментов сделано З.М.Шаляпиной.