

## **Раздел III.**

# **Доклады, представленные участниками тестирования систем синтаксического анализа**

В данном разделе публикуется итоговая статья организаторов тестирования систем синтаксического анализа и отдельные статьи участников тестирования. Полностью с комментирующими сообщениями участников можно ознакомиться на сайте конференции «Диалог»

# ОЦЕНКА МЕТОДОВ АВТОМАТИЧЕСКОГО АНАЛИЗА ТЕКСТА 2011–2012: СИНТАКСИЧЕСКИЕ ПАРСЕРЫ РУССКОГО ЯЗЫКА<sup>1</sup>

**Толдова С. Ю.** (toldova@yandex.ru),  
**Соколова Е. Г.** (minegot@rambler.ru)  
РГГУ, Москва, Россия

**Астафьева И.** (astafir@gmail.com),  
**Гарейшина А.** (a.r.gare@gmail.com),  
**Королева А.** (tresh\_miralissa@mail.ru),  
**Привознов Д.** (dprivoznov@gmail.com),  
**Сидорова Е.** (begushchaya.po.volnam@gmail.com),  
**Тупикина Л.** (lyubov98@gmail.com)  
МГУ им. М. В. Ломоносова, Москва, Россия

**Ляшевская О. Н.** (olesar@gmail.com)  
НИУ ВШЭ, Москва, Россия

Второй раунд форума «Оценка методов автоматического анализа текста» в 2011–2012 гг. был посвящен синтаксическим анализаторам русскоязычных текстов. В статье описываются принципы и процедура проведения дорожек форума, состав участников, тестовая коллекция и Золотой Стандарт, на основе которого осуществлялась оценка, принципы сопоставления ответов систем, сложные для оценки случаи, а также некоторые проблемные точки в работе синтаксических парсеров, которые выявила экспертиза результатов.

**Ключевые слова:** синтаксический анализ, автоматическая обработка текста, парсеры, русский язык, соревнования парсеров, оценка

---

<sup>1</sup> Работа выполнена при частичной финансовой поддержке Программы фундаментальных исследований Президиума РАН «Корпусная лингвистика». Мы выражаем благодарность А. Бонч-Осмоловской, С. Ковалю, Ю. Гришиной, М. Ионову, А. Литягиной, Н. Меньшиковой, А. Семеновской, принимавшим вместе с нами участие в организации соревнований и экспертизе результатов. Отдельно мы выражаем благодарность организаторам конференции Диалог за помощь в проведении Форума и поддержку сайта (<http://dialog-21.ru/Default.aspx?DN=605c0b84-a3bd-42d4-8225-b70760f65c1d&l=Russian>). Также мы благодарим участников форума за сотрудничество.

## NLP EVALUATION 2011–2012: RUSSIAN SYNTACTIC PARSERS

**Toldova S. Ju.** (toldova@yandex.ru),  
**Sokolova E. G.** (minegot@rambler.ru),  
Russian State University for the Humanities, Moscow, Russia

**Astaf'eva I.** (astafir@gmail.com),  
**Gareyshina A.** (a.r.gare@gmail.com),  
**Koroleva A.** (tresh\_miralissa@mail.ru),  
**Privoznov D.** (dprivoznov@gmail.com),  
**Sidorova E.** (begushchaya.po.volnam@gmail.com),  
**Tupikina L.** (lyubov98@gmail.com),  
Lomonosov Moscow State University, Moscow, Russia

**Lyashevskaya O. N.** (olesar@gmail.com),  
National Research University Higher School of Economics,  
Moscow, Russia

NLP Evaluation forum RU-EVAL started in 2010 as a new initiative aimed at independent evaluation of the methods used in Russian language resources and linguistic tools. The second evaluation campaign (2011–2012) is focused on syntactic parsing. It is open both to academic institutions and industrial companies and its general objective is to access the current state-of-the-art in the field and promote the development of syntactic technologies. The paper presents the principles and design of two tracks, which were organized thematically, namely, Main track and News. There were seven participants who follow either rule-based or statistical approach; all of them submitted runs to both tracks. The training set consisted in 100 sentences, the dataset for annotation included ca. one million words, and the test set was composed by ca. 800 sentences (500+ sentences for the Main track and 300+ sentences for the News track). The test set was annotated manually as a Golden Standard by two annotators. We describe how the outputs were compared and discuss common pitfalls for evaluation as well as some cases that are still problematic for parsing. As a side effect of the evaluation campaign and benchmarks for future development, the test data including Golden Standard and three automatically annotated answers are available to the NLP community at <http://testsynt.soiza.com>.

**Key words:** syntax, natural language processing, parsers, Russian, parsing tracks, evaluation

## 1. Введение

В начале 2011 года был объявлен второй цикл форума «Оценка методов автоматического анализа текста». Темой форума стал автоматический синтаксический анализ русского языка. При организации форума 2011–2012 года использовался опыт Семинара по оценке методов информационного поиска РОМИП (ROMIP 2009) и форума 2010 года.

Целью Форума является создание независимой площадки, где представители научных, образовательных, коммерческих и т. п. организаций обсуждают состояние и перспективы развития алгоритмов и методов автоматической обработки текста (прежде всего, для русского языка), а также проводится экспертиза лингвистических компьютерных разработок. Независимые соревнования лингвистических систем проводятся в разных странах мира (ср. проекты CLEF, Morpho Challenge, AMALGAM, GRACE, EVALITA, PASSAGE, SEMEVAL, российский семинар оценки методов информационного поиска РОМИП и др.). В 2010 году состоялся первый цикл Форума, в котором приняло участие 15 команд разработчиков из Москвы, Санкт-Петербурга, Екатеринбурга, Украины, Беларуси и Великобритании. Форум–2010 был посвящен системам морфологического анализа русского языка (см. подробнее о принципах и результатах на сайте <http://ru-eval.ru>, а также Lyashevskaya et al. 2010). В рамках второго цикла оценивалось состояние лингвистических технологий в области автоматического синтаксического анализа. В России такое сравнение происходило впервые. На конференции «Диалог 2011» был проведен круглый стол с участием ведущих разработчиков синтаксических парсеров; осенью 2011 года состоялись дорожки форума. Как и в 2010 г., второй цикл форума имел и образовательную составляющую. В экспертной группе работали студенты, связывающие свое будущее с прикладной лингвистикой. Экспертиза результатов проводилась автоматически с последующей двойной ручной перепроверкой. Рейтинг ответов систем будет объявлен в мае 2012 г. на круглом столе конференции «Диалог».

Следует сразу отметить, что сравнение работы синтаксических анализаторов на порядок сложнее, чем оценка автоматического морфологического анализа. В области морфологии существует значительная зона пересечения: существуют общие представления о морфологической норме, отраженной в словарях и грамматиках, для большинства морфологических параметров есть устоявшаяся традиция «ярлыков», а значительная часть случаев, когда имеет место варьирование, поддается унификации простым переименованием тегов. В синтаксическом анализе могут быть использованы разные формализмы и принципы представления синтаксической структуры. В этом году сравнивались результаты работы систем, представленные в виде дерева зависимостей. Но и в этом случае результаты разбора сильно отличались друг от друга и в значительной степени зависели от того, каковы конечные задачи системы, в которую данный модуль встраивается. В связи с этим проведению самого соревнования предшествовал длительный этап подготовки, в том числе и обсуждения формата проведения форума в рамках конференции «Диалог 2011». Высокая активность как академических коллективов, так и промышленных

разработчиков в процессе обсуждений показала, что данное направление автоматической обработки текста чрезвычайно востребовано на современном этапе. Помимо традиционной области применения результатов синтаксического анализа, такой как машинный перевод, данный модуль активно используется в системах автоматического анализа контента, например, в извлечении именованных сущностей или фактов из текста, при мониторинге блогов и новостей и др.

«Синтаксический» цикл проходил следующим образом. Участники форума получили специально отобранную и подготовленную коллекцию текстов, обработали их в своих системах и представили результат синтаксического анализа в некотором унифицированном формате. Правильность разбора оценивалась при сравнении с эталоном, размеченным вручную.

По результатам проведенного соревнования можно сказать, что, не смотря на различные трудности, с которыми организаторы столкнулись при проверке результатов, удалось выработать некоторый формат и принципы, позволяющие производить такое сравнение.

Форум не только позволил оценить работу синтаксических парсеров, но и дал целый ряд общезначимых в области синтаксического анализа результатов. Был получен корпус вручную размеченных и выверенных текстов, который можно использовать в научно-исследовательских целях (он представлен в свободном доступе на сайте [testsynt.soiza.com](http://testsynt.soiza.com)). В подготовке и проведении дорожек форума 2011–2012 года и в формировании финального отчета активное участие принимали студенты Отделения теоретической и прикладной лингвистики филологического факультета МГУ им. М. В. Ломоносова, которые получили возможность «пощупать руками», как работают парсеры, увидеть, в чем их сильные и слабые стороны и т. д.

Как и при проведении форума 2010 года, в основу принципов проведения дорожек 2011–2012 гг. легло следующее положение: не бывает единственно правильного решения спорных вопросов и единственно правильного алгоритма синтаксического анализа. По возможности, ошибочными считались только разборы, не мотивированные теоретическими или практическими установками авторов системы. Можно указать множество примеров того, как оптимальный выбор того или иного решения зависит от цели, для которой проводится анализ. Также существует целый ряд проблемных случаев, не имеющих единственного решения. При сравнении работы разных парсеров был уточнен список проблемных зон в области синтаксического анализа, а также множество возможных подходов к их обработке.

Таким образом, в процессе проведения форума удалось получить некоторую оценку состояния автоматического синтаксического анализа русского языка, выявить проблемные и дискуссионные места синтаксического анализа, в которых при разных подходах принимаются принципиально разные решения, оценить варьирование в базовых подходах к типизации синтаксической реальности. Также результаты форума показали, что в области автоматического синтаксического анализа русского языка разработчикам удалось достичь достаточно высокого уровня.

## 2. Подходы и проблемы, связанные с оценкой автоматического синтаксического анализа

Предварительная оценка состояния автоматического синтаксического анализа для русского языка показала, что большинство систем используют формализм зависимостей. Таким образом, при проведении конкурса рассматривались результаты, представленные в виде деревьев зависимостей, независимо от тех формализмов, которые использовали разработчики в своих системах.

При организации синтаксического цикла мы опирались на мировой опыт проведения соревнований подобного типа, некоторые из которых упомянуты во Введении, в частности, на опыт проведения аналогичной оценки систем для итальянского языка EVALITA. Для дорожки по деревьям зависимостей участники получают на вход корпус текстов, разбитых на предложения и токены. Задача заключается в том, чтобы для каждой словоформы в предложении указать ее синтаксическую вершину, а также тип синтаксической связи.

Как правило, при проведении соревнований множеством эталонного набора типов связей (имен связей и набор устанавливаемых синтаксических отношений), используемые в оценке работы систем, служат данные уже готовых синтаксически размеченных корпусов. Тем более, многие разработчики используют эти наборы при создании систем, особенно если система строится на машинном обучении. Так, например, для итальянского языка используется Turin University Treebank (TUT), размеченный в обоих формализма (и в терминах непосредственных составляющих, и в терминах деревьев зависимостей)<sup>2</sup>. Также предложения из таких трибанков часто служат тестовым корпусом, что позволяет обеспечить процедуру автоматической проверки.

Анализ пробного разбора 100 предложений, представленного разработчиками–потенциальными участниками Форума 2011–2012, показал, что в России системы синтаксического анализа развивались автономно, без использования какого бы то ни было корпуса в качестве эталона. В результате, расхождения между системами по составу тегов и по принципам установления связей оказались настолько значительными, что в целом ряде вопросов не удалось предложить единого решения для представления выходных данных. Было принято решение о том, что на данном этапе оцениваться должно только правильное определение системами синтаксически связанных пар словоформ и установление «главного» элемента в паре. При этом при оценке не должны оцениваться теоретические расхождения в трактовке тех или иных синтаксических явлений.

---

<sup>2</sup> Широко известны и активно используются в обучении анализаторов и оценке их работы для английского языка Penn Treebank, размеченный по непосредственным составляющим, The Prague Dependency Treebank для чешского языка, основанный на деревьях зависимостей.

### 3. Форум 2011–2012: синтаксические парсеры

#### 3.1. Дорожки

На форуме 2011–2012 по синтаксическому анализу текстов оценивание алгоритмов систем–участников прошло независимо по следующим отдельным дисциплинам (дорожкам):

- «общая»; в этой дорожке рассматривались различные типы текстов и синтаксический разбор всех представленных в них предложений;
- «новостная»; задача этой дорожки состояла в синтаксическом разборе предложений узкой тематики, а именно — новостного блока.

В процессе подготовки форума также затрагивались вопросы о дальнейшей разработке дополнительных дорожек — по разбору сложных предложений целиком vs. отдельному разбору простых предложений в составе сложного, выделению проективных vs. непроективных предложений и др. Однако проведение таких дорожек сильно бы усложнило и без того достаточно трудоемкую процедуру проверки.

#### 3.2. Участники

На конкурс были поданы заявки от 11 различных групп разработчиков из Москвы, Санкт–Петербурга, Нижнего Новгорода (Россия), Донецка (Украина). Одна из этих групп участвовала в проекте вне конкурса, поэтому её результаты не включались в общее соревнование. Конечные результаты, и по основной, и по новостной дорожкам, были получены от 8 из 10 участников форума: SynAutom, DictaScope Syntax, SemSin, ЭТАП–3, синтактико–семантический анализатор русского языка группы SemanticAnalyzer Group, проект AotSoft, ABBYY Syntactic and Semantic Parser (ASSP), Парсер грамматики связей. Среди них системы, использующие различные методы синтаксического разбора: грамматику зависимостей, грамматику составляющих, грамматику связей (Link grammar parser). Один из восьми разработчиков впоследствии был вынужден отозвать своё участие в конкурсе из–за проблем с конвертированием данных. Таким образом, в окончательной оценке участвовало 7 различных систем обработки текстов.

#### 3.3. Тестовая коллекция и задания

Для соревнования была подготовлена общая коллекция неразмеченных текстов. В коллекцию для «основной» дорожки вошли тексты разных жанров, включая художественную литературу, публицистику, а также 5% текстов из социальных сетей. В коллекции были представлены как отдельные предложения

(200 тыс. словоупотреблений из Национального корпуса русского языка, предоставленные для свободного скачивания), так и фрагменты связанных текстов. В новостную коллекцию вошли фрагменты текстов из новостной коллекции семинара РОМИП. В эту коллекцию попали последовательности из трех предложений, выбранные случайным образом. Все тексты были заранее разбиты на предложения и токены и проиндексированы.

Участники конкурса должны были приписать каждому токenu номер его вершины. При проверке не оценивалась правильность разбора всего предложения, оценивалась правильность приписывания вершины зависимой словоформе. Сравнение результатов по всем дорожкам проводилось на основе выборочной проверки ответов систем–участников. Для этого был подготовлен «Золотой Стандарт» — множество случайно выбранных предложений из Основной коллекции, объемом около 800 предложений (500 для основной коллекции и 300 для новостной). В ходе экспертизы ответы систем сравнивались с произведенной экспертами ручной разметкой Золотого Стандарта, см. п. 3.7–3.8.

### 3.4. Соглашения по унификации входного формата

Для унификации результатов, получаемых от разных систем, был разработан специальный входной формат представления текстовой коллекции. Исходный корпус предоставлялся участникам в двух форматах: исходный текст без разметки и xml–формат с разбивкой на предложения и токены. Были приняты некоторые соглашения относительно правил токенизации. Отдельными токенами считались словоформы, входящие в одну сложную единицу, например, в сложный союз или предлог. На отдельные токены также разбивались слова с дефисом, за исключением некоторых заданных списком слов, а также местоимений с частицей *–то*, наречий с *по–* и т. п., отдельными токенами считались знаки препинания.

Предварительная токенизация и нумерация токенов нужна была для того, чтобы минимизировать долю ручной проверки. Благодаря такой унификации можно было автоматически определять фрагменты, в которых разметка участников совпадает с Золотым Стандартом, что минимизировало долю ручной проверки. В первую очередь «вручную» просматривались места несовпадений. Многие участники игнорировали нумерацию в процессе работы собственного анализатора, но потом приводили ID токенов в соответствии с ID токенов в тестовой коллекции.

### 3.5. Соглашения об унификации выходного формата

Результат работы систем должен был быть представлен также в специальном формате. В выходном файле нумерация предложений и токенов должна была соответствовать нумерации в тестовом корпусе. Участники должны были указать для каждой словоформы номер «хозяина» (главного слова в словосочетании) и тип связи (указывался тип синтаксической связи, принятый



у разработчика), также указывалась морфологическая информация: лемма и набор морфологических характеристик. Тип связи и морфологическая информация указывалась на усмотрение разработчиков и нужна была для облегчения ручной проверки, чтобы эксперту было легче понять, в чем причина расхождения ответа системы с Золотым Стандартом.

### 3.6. Соглашения по унификации направлений связей

Подготовительный этап потребовал определенных решений, направленных на унификацию структуры синтаксических отношений в ответах, ожидаемых от парсеров. Существует достаточно много ситуаций, когда системы по-разному решают вопрос о направлении синтаксической зависимости между двумя словоформами, находящимися в отношении синтаксической связи (подробнее о расхождениях см. п. 4). Эти случаи обусловлены не ошибками при анализе, а принципиальными решениями при создании конкретных систем. В таких случаях расхождения системы с эталоном не «штрафовалось». Однако для того, чтобы не пришлось просматривать каждый подобный случай вручную, некоторые системы, по крайней мере, в части случаев согласились изменить направления связей там, где это было возможно сделать автоматически. Это касалось следующих типов связей:

- 1) предлог — существительное;
- 2) вспомогательный глагол — смысловой глагол;
- 3) связи в сочинительных конструкциях.

### 3.7. Подготовка Золотого Стандарта

Разметка Золотого Стандарта, предшествовавшая экспертизе результатов, проводилась вручную с помощью инструмента для разметки, подготовленного М. Ионовым. Каждое предложение первоначально размечалось двумя экспертами, после чего места расхождений обсуждались. На основании обсуждений принималось единое решение. Далее окончательный вариант проверялся третьим экспертом. Такая разметка позволяла достичь нескольких целей. Во-первых, это позволило автоматизировать процедуру разметки. Во-вторых, организаторы хотели по возможности избежать влияния результатов, предоставленных системой, на интуицию экспертов, и пропусков ошибок по невнимательности. В-третьих, разметка Стандарта должна была сформировать у экспертов представление о том, какие сложные случаи их ожидают, выработать критерии для оценки расхождений.

При разметке аннотаторы пользовались специальной инструкцией, обеспечивающей «устойчивость» аннотации, т. е. согласованность в принятии решений разными аннотаторами в одинаковых ситуациях.

### 3.8. Принципы разметки Золотого Стандарта

Для разметки Золотого Стандарта требовалась такая инструкция, которая бы обеспечила не столько теоретическую (абстрактную) «правильность» разметки, сколько единообразие разметки разными аннотаторами, четкую обоснованность принимаемых при разметке решений.

Мы основывались на принципах и средствах синтаксической разметки, сформулированных в (Sokolova 2011; ср. также Novy and Lavid 2010) и опробованных на занятиях по синтаксической разметке текстов студентов 4-го курса РГГУ в течение нескольких лет. Одним из важных принципов, который лег в основу принимаемых решений, является принцип «естественности» разметки: разметка должна соответствовать правильной семантической интерпретации предложения (в инструкции он формулируется следующим образом: «Синтаксическая структура языкового произведения осмысленна и единственна»). Из возможных вариантов отбирались наиболее простые и понятные решения, которые максимально согласовались с интуицией разметчика.

Форма структуры — дерево зависимостей, узлами которого являются словоформы (а не наборы морфологических интерпретаций словоформ). При этом может сохраняться некоторая неоднозначность ее морфологической и семантической интерпретации, не противоречащая структуре дерева, например, в предложении *Свидания разрешить не могу* словоформа *свидания* зависит от словоформы *разрешить* и имеет синтаксическую функцию “obj”. В структуру может входить любая из двух ее морфологических интерпретаций: — *свидание* — ед. ч., род. п. в контексте отрицания (т. е. «не разрешаю конкретное свидание»); — мн. ч., вин. п. (т. е. «вообще никакие свидания не разрешаются»). В отличие от «классической» Грамматики Зависимостей интерпретация отношений заменена на Синтаксические Функции словоформ — формализм ГЗиСФ, при котором синтаксическая функция приписывается всем словоформам — и подчиненным, и вершинам. Принцип единственности структуры заимствован из опыта разметки целых текстов студентами, при которой предложение рассматривается внутри конкретного текста. Мы сохранили этот принцип для разметки отдельных предложений в Золотом Стандарте. При этом редкие (относительно Золотого Стандарта) варианты синтаксической интерпретации предложения не считаются ошибкой.


### 3.9. Экспертиза ответов систем

Процедура экспертизы ответов синтаксических анализаторов предусматривала сравнение номеров вершин, указанных системами для каждой словоформы, с ее номером в Золотом Стандарте. Совпадение номеров автоматически

получало оценку 0. Случаи расхождений просматривались экспертами, которые должны были оценить их по следующей шкале:

- 1 — ошибка системы;
- 2 — ошибка ЗС;
- 3 — допустимое расхождение (расхождения объясняются расхождением в теоретических решениях системы и ЗС);
- 4 — допустимое расхождение (случай допустимой омонимии);
- 5 — ответ системы совпадает с ЗС, но оба неправы;
- 6 — для данного токена «хозяин» не указан, а должен быть указан;
- 7 — для данного токена «хозяин» не указан и может быть не указан;
- 8 — затрудняюсь определить (эксперт не может принять однозначное решение);
- 9 — другое.

Фрагмент проверочной таблицы указан на рисунке 1<sup>3</sup>.

Sentence 1819  

Золотой стандарт				Вариант участника			
id	token	type	head a	id	token	type	head mark
1	Каких ← результатов	amod	3	1	Каких ← результатов	Какой	3 0
2	именно ← Каких	spec	1	2	именно ← результатов	Частица	3 4
3	результатов ← ждать	obj	5	3	результатов ← ждать	Род	5 0
4	можно	pred		4	можно		
5	ждать ← можно	comp	4	5	ждать ← можно	Сост_сказ	4 0
6	от ← ждать	comp	5	6	от ← ждать	Откуда_От	5 0
7	совместных ← усилий	amod	8	7	совместных ← усилий	Какой	8 0
8	усилий ← от	rcomp	6	8	усилий ← от	Род	6 0
9	членов ← усилий	mod	8	9	членов ← группы	Род	10 1
10	группы ← членов	mod	9	10	группы ← ждать	Вин	5 1
11	.			11	.	группа (но,мн,с,жр,вн)	

**Рис. 1.** Пример разметки Золотого Стандарта и ответа одной из систем. В графе “mark” указана оценка за решение

Сравнение ответов систем с Золотым Стандартом позволило выделить наиболее распространенные отклонения от разборов, признанных эталонными (см. п. 4.1).

Проверка также показала, что не всегда удастся оценить, в какой степени тот результат, который представлен в ответе системы, определяется принципиальными решениями, принятыми в системе, проблемами «пересчета» направлений связи в соответствии с Золотым Стандартом или же ошибкой в разборе.

<sup>3</sup> Организаторы выражают благодарность Горшкову Д. В. за компьютерную поддержку в проведении конкурса, в частности за разработку базы данных и визуализации деревьев для обеспечения автоматического и ручного этапа проверки, а также повторной перепроверки.

К сожалению, таких случаев оказалось значительное количество. Они потребовали дополнительной выверки результатов. Значительную помощь в улучшении системы оценки оказали комментарии разработчиков, присланные ими после того, как они получили доступ к промежуточным оценкам. Однако даже при дополнительном пересмотре не удалось избежать ситуаций, когда «штраф» системе приписан ошибочно.

В следующем разделе остановимся более подробно на отдельных вопросах выработки ряда решений при организации Форума 2011–2012, а также на сложных моментах, с которыми нам пришлось столкнуться.

## 4. Трудные случаи и расхождения

### 4.1. Допустимая вариативность разборов

Расхождения между системами по составу тегов и по принципам установления связей оказались настолько значительными, что в целом ряде вопросов не удалось предложить единого решения для представления выходных данных.

Во-первых, разные системы не только используют разные названия для одних и тех же синтаксических отношений, но существуют значительные расхождения в самой классификации типов связи. Так, в одних системах разграничение типов связей опирается на морфологическую разметку, в других, наоборот, учитывается самая общая синтаксическая функция словоформы. Например, в одних системах отдельно выделяется тип связи “card” для связи числительного с существительным (ср. *тысячи* ← *педагогов* (card)), в других этот случай относится к общему случаю несогласованного определения. В силу этого обстоятельства решено было при сравнении результатов не учитывать имена связей.

Во-вторых, помимо конструкций, не вызывающих вопросов и размечаемых всеми одинаково (согласованное определение), существуют конструкции, относительно которых не существует единого теоретического решения. В частности, в целом ряде конструкций невозможно однозначно установить, какой из синтаксически связанных элементов является главным, а какой зависимым (подробно о таких конструкциях см., например, Iomdin 1990, Gladkij 1973): это случаи, когда либо разные критерии выделения вершин дают разные результаты (см., например, Testelets 2001), либо ни один критерий не применим. Примером может служить сочинение: при наличии союза между сочиненными элементами количество различных разборов становится немалым, потому что этому союзу можно приписать несколько разных вершин (а также считать вершиной сам союз). Однако до тех пор, пока все сочиненные члены с союзом или союзами соединяются в одну группу, нет причин считать такой разбор ошибкой. Несколько вариантов разбора допустимы также в случае становления связи между клаузами в сложноподчиненных предложениях. В ряде

систем клаузы соединяются между собой через глаголы, в других — через подчинительные союзы.

В третьих, вариативность в разборах обусловлена разными практическими задачами, решаемыми системами. Так, например, в соответствии с критериями выделения вершин главным в словосочетании ‘вспомогательный глагол + смысловой’, как в *станет писать*, является вспомогательный глагол, однако многие системы последовательно устанавливают направление связи в данном случае ‘вспомогательный глагол ← смысловой глагол’.

В результате анализа разборов Е. Г. Соколовой была составлена таблица возможных расхождений по отдельным типам связей (см. <http://testsynt.soiza.com/files/var-synt.htm>).

В дальнейшем целесообразно добиваться того, чтобы ответы систем одинаково представляли наиболее частотные случаи, в которых сейчас наблюдаются расхождения: неодносложные союзы и предлоги, сложные слова с дефисным написанием; связь между однородными членами, между главной и подчиненной клаузой, между сочиненными клаузами (включая интерпретацию союзов), союз в начале главной клаузы; глагол-связку с инфинитивами, именами, прилагательными, причастиями; группы с количественными и порядковыми числительными (включая предложные и с модификаторами типа *более*, *минимум*); связь подлежащего с именным сказуемым; связь в группах вида ‘прилагательное + прилагательное + существительное’ и нек. др.

## 4.2. Анализ ответов систем: проблемные точки

В целом, приятным итогом анализа ответов стал вывод, что в пределах простого предложения/клаузы нет «больных мест», общих для всех участников. Среди частных проблем можно назвать свободно присоединяемые предложные зависимые (или те, что отсутствуют в актантном словаре или не выучены системой). Если в предложении находится несколько потенциальных хозяев, то системы выбирают либо линейно предшествующее существительное, либо вершинный глагол, либо ближайший финитный глагол в дереве, однако не все такие варианты будут семантически оправданы, ср. допустимые (1А–В), (2А–Б) и недопустимые (1Г), (2В):

- (1) *Компания Google продолжает укреплять свои позиции на рынке приложений для совместной работы.*  
 А. <sup>ок</sup>позиции → на рынке  
 Б. <sup>ок</sup>укреплять → на рынке  
 В. <sup>ок</sup>приложений → для совместной работы  
 Г. \*укреплять → для совместной работы.

(2) ... что может добиться своей цели лишь при одном условии...

А. <sup>OK</sup> добиться → при условии

Б. <sup>OK</sup> может → при условии

В. \*цели → при условии

Большинство систем не смогло справиться с примером, в котором присутствуют три однородных определения вида X, Y и Z к существительному:

(3) В качестве пилотных субъектов РФ признаны Челябинская, Томская и Архангельская области.

Системы могут ошибочно считать, что первые определения зависят от РФ, или не найти связи с несогласованным по числу существительным.

Многие системы ошибаются при обработке идиоматических конструкций «малого синтаксиса», если срабатывают альтернативные характерные для русского языка шаблоны, ср. неверно приписанную атрибутивную связь в паре обучение → такое (4):

(4) Что такое обучение?

В сложных предложениях, безусловно, ошибок больше. Часто наблюдаются проблемы с нахождением вершины в предшествующей клаузе. Например, в (5) хозяевами вершины клаузы чтобы... двигалась... называются возьмем, образуем, но не элементы в составе деепричастного оборота. Аналогично, могут оставаться незамеченными вершины–существительные или связки типа есть.

(5) Если мы возьмем какую-то замкнутую фигуру и образуем твердое тело, вращая эту фигуру в пространстве так, чтобы каждая точка двигалась перпендикулярно к плоскости фигуры...

Наконец, во многих случаях наблюдается ложное срабатывание систем, когда дистантно расположенный зависимый выхватывается через границу клаузы, а также ненахождение связей для несловарных слов (например, ОС, Intel и др.)

## 5. Заключение. Итоги форума

На наш взгляд, несмотря на большую вариативность в теоретических подходах, практических решениях, качестве работы систем, несмотря на то, что не удалось конвертировать ответы систем в единый формат, который можно было бы автоматически сопоставить с Золотым Стандартом, проведение форума синтаксических парсеров дало много полезных результатов:

- был создан вручную размеченный эталон объемом в 800 предложений, а также передана в общее пользование инструкция, эксплицитно поясняющая те или иные решения
- для систем, представивших результаты, был создан реестр расхождений, который может быть обобщен до реестра допустимых общетеоретических решений и таблицы их «эквивалентности»
- была осознана необходимость «публичности» эталонного общезначимого трибанка с параллельной разметкой разными системами, аккумулирующая множества тегов и принципов разбора; особенно ценен такой ресурс для разработчиков, чьи знания о синтаксисе не выходят за рамки школьной программы, а также для развития систем, «варящихся в собственном соку»
- разработчики систем получили открытый доступ к своим промежуточным оценкам; по сравнению с форумом 2010 года, удалось добиться большего взаимодействия организаторов и разработчиков при подготовке дорожек и обсуждении результатов
- главный итог: до начала объявления соревнований трудно было оценить общую ситуацию с состоянием автоматического синтаксического анализа в России: какие системы представлены, какие формализмы используются, какие принципы установления синтаксической зависимости между единицами предложения положены в основу, каково множество синтаксических отношений, которые система различает. Проведенное соревнование позволило, в определенной степени, такую картину составить.

Перспективы продолжения форума нам видятся в дальнейшей автоматизации экспертизы и еще большей ее открытости; в том, чтобы повысить гибкость оценок с учетом комментариев по принципиальным решениям, высказанных разработчиками; в содержательном ключе хотелось бы большее внимание уделить типам синтаксических связей и обработке сложных предложений. А у разработчиков, в свою очередь, будет возможность улучшить результаты с учетом накопленного соревновательного опыта.

## References

1. *Gladkij A. V.* (1973), *Formal'nye grammatiki i jazyki* [Formal Grammars and Languages], Moscow, Nauka.
2. *Hovy E., Lavid Ju.* (2010), Towards a 'science' of corpus annotation: a new methodological challenge for corpus linguistics, *International journal of translation*, Vol. 22, no. 1, pp. 1–25.
3. *Iomdin L. L.* (1990), *Avtomaticeskaja obrabotka teksta na estestvennom jazyke: model' soglasovanija* [Natural Language Processing: a Model of Agreement], Moscow, Nauka.
4. *Lyashevskaya O., Astaf'eva I., Bonch–Osmolovskaya A., Garejshina A., Grishina Ju., D'yachkov V., Ionov M., Koroleva A., Kudrinsky M., Lityagina A., Luchina E., Sidorova E., Toldova S., Savchuk S., Koval' S.* (2010), *Ocenka metodov*

avtomatičeskogo analiza teksta: morfoložičeskije parsery russkogo jazyka [NLP evaluation: Russian morphological parsers], in Computational linguistics and intellectual technologies. Proceedings of International Workshop Dialogue'2010, Vol. 9 (16), Moscow, pp. 318–326.

5. *ROMIP* (2009): Rossijskij seminar po ocenke metodov informacionnogo poiska. Trudy ROMIP 2009, Petrozavodsk, 16 sentjabrja 2009 [Russian Information Retrieval Evaluation Seminar. Proceedings of ROMIP 2009, Petrozavodsk, September 16, 2009]. Saint–Petersburg, NU CSI.
6. *Sokolova E. G.* (2011), Syntactic annotation in terms of dependency grammar and syntactic functions [Sintaksičeskaja razmetka v terminax grammatiki zavisimostej i sintaksičeskix funkcij], Moscow, RGGU, available at: <http://elib.lib.rsuh.ru/elib/000003603.pdf>
7. *Testelets Ja. G.* (2001), Vvedenie v obščij sintaksis [Introduction to general syntax], Moscow, RGGU.