

# Модели и методы учета пунктуации при синтаксическом анализе предложения русского языка

## Models and methods of punctuation use in Russian language syntax parsing

**Окатыев В. В.** (oka@dictum.ru),  
**Ерехинская Т. Н.** (te@dictum.ru), **Скатов Д. С.** (ds@dictum.ru)

ООО «Диктум», г. Нижний Новгород

Рассматривается функциональная омонимия знаков препинания в русском языке. Предложена формальная модель обособлений и рядов. Определены отношения между ними. Предложена математическая постановка задачи анализа пунктуации в контексте задачи машинного синтаксического анализа и алгоритм ее решения.

В настоящей работе рассматривается задача учета пунктуации при синтаксическом анализе предложения. Одной из основных проблем компьютерной лингвистики является омонимия, в частности функциональная омонимия знаков препинания. Знание функций знаков препинания в анализируемом предложении помогает определить структуру подчинительных связей слов, что, в конечном счете, позволяет более качественно решать задачу синтаксического анализа предложений на естественном языке.

На наш взгляд, в настоящее время задача синтаксического анализа предложения с учетом пунктуации недостаточно формализована. В работе [1] отмечается, что знаки препинания не только организуют текст, но иногда оказываются единственным доступным средством выбора правильной интерпретации текста. К сожалению, в этой работе не предлагаются конкретные алгоритмы реализации пунктуационного разрешения неоднозначности для систем автоматической обработки текста, т.к., по мнению ее авторов, эти алгоритмы сильно зависят от общего устройства конкретной системы.

Разными исследователями информация о знаках препинания в предложении используется по-разному. При статистических подходах пунктуацию часто игнорируют. Интересен подход, основанный на использовании пунктуации — фрагментационный анализ [3-5]. При фрагментационном анализе заданное предложение разбивается по знакам препинания и союзам на отрезки, а затем с помощью специальных правил и словарей выполняются операции объединения, сочинения и подчинения этих отрезков. При построении анализатора на основе

правил знак пунктуации может считаться токеном, и включаться в состав правил, как это делается в работах [2] и [10]. Кроме того, в работе [10] критикуется разделение синтаксического анализа на два уровня: обработка фрагментов, не содержащих знаков препинания, и затем определение отношений между этими фрагментами.

Много внимания уделяется пунктуации в работах Кобзаревой Т. Ю. В работе [6] описывается вложение сочиненных групп — «матрешка» сочинительных конструкций, ограничения по размещению отдельных компонентов. В работе [8] рассматривается структура вложений придаточных предложений и обособленных оборотов. В работе [7] описана омонимия знаков препинания и их способность выполнять несколько функций одновременно.

В настоящей работе описана формальная модель пунктуации и предложена математическая постановка задачи анализа пунктуации в контексте задачи синтаксического анализа. Формальная модель синтаксической конструкции, содержащей знаки препинания, понятия композиции конструкций и покрытия знаков препинания предложены Окатыевым В. В. Им же предложена формальная постановка задачи анализа пунктуации. Алгоритм решения задачи, его эвристические улучшения и интеграция с методами синтаксического анализа разработаны Ерехинской Т. Н.

### 1. Обзор функций знаков препинания

В грамматике русского языка [9] различаются:

1. Знаки выделяющие. Сюда относятся: две запятые (как единый парный знак), два тире (то же самое), скобки, кавычки;
2. Знаки отделяющие. Сюда относятся: точка, вопросительный знак, восклицательный знак, запятая, точка с запятой, двоеточие, тире, многоточие.

Использование знаков препинания в качестве формальных признаков при автоматической обработке текста осложняется их многофункциональностью. Например, запятая может выполнять как отделяющую функцию, так и выделяющую. Более того, одна и та же запятая может выполнять эти функции одновременно. В частности, запятая может закрывать причастный оборот и одновременно разделять однородные члены.

Проведенные исследования позволяют сделать вывод о наличии весьма строгих закономерностей, которым подчиняются правила пунктуации русского языка. Обнаруженные закономерности допускают их формализацию с помощью математических моделей. Эти модели могут использоваться для разработки компьютерных программ синтаксического анализа, в частности, они используются в системе синтаксического анализа DictaScore компании «Диктум» [11, 12].

## 2. Моделирование обособлений и рядов

В данной работе из всех синтаксических конструкций, в построении которых используются знаки препинания, будем рассматривать только обособления и ряды однородных членов. При машинном синтаксическом анализе эти конструкции невозможно рассматривать отдельно, поскольку в их построении применяются одни и те же знаки пунктуации, прежде всего – запятая.

При разработке методов анализа обособлений и рядов необходимо учитывать наличие сочинительных союзов, участвующих в построении союзных рядов однородных членов. Поскольку сочинительные союзы выполняют отделяющую функцию, то их в рамках данного исследования целесообразно рассматривать наравне с запятыми.

Отдельно следует сказать о двоеточии. В рамках данной работы будем его игнорировать. Разработанный подход к анализу пунктуации допускает включение этого знака препинания в модель, однако это привело бы к усложнению описания. Такое решение не приводит к существенному ухудшению результатов, поскольку основные сложности при анализе пунктуации связаны с запятыми, встречаемость которых на порядок выше, чем у двоеточий [12].

Перейдем к формальному описанию обособлений и рядов. Для этих целей пронумеруем слова в анализируемом предложении по порядку слева направо, начиная с единицы. Каждому знаку препи-

нания припишем номер слова, после которого он непосредственно стоит, увеличенный на 0.5. Таким образом, знакам препинания будут соответствовать дробные номера. Например, если между вторым и третьим словами в предложении находится запятая, то у нее будет номер 2.5. В рамках разрабатываемой модели имеет значение линейный порядок слов и знаков препинания, а не величина их номеров.

Будем рассматривать ряд как множество пар однородных членов, между которыми можно предположить наличие сочинительной связи. Чтобы сделать такое предположение для пары слов  $b$  и  $d$ , необходимо указать их общего предка, от которого они зависят, и разделитель. Функцию разделителя может выполнять сочинительный союз, либо знак препинания: запятая или точка с запятой. Таким образом, сочинительная связь моделируется четверкой чисел:

$$f = \langle a, b, c, d \rangle,$$

где  $a$  – главное слово;  $b$ ,  $d$  – зависимые слова;  $c$  – разделитель однородных членов (сочинительный союз, запятая, точка с запятой). Для двойных сочинительных союзов будем указывать позицию второго союза.

Для  $f$  выполняется следующее соотношение:

$$b < c < d \ \& \ (a < b \vee d < a).$$

Вариант  $a < b$  соответствует препозиции главного слова,  $d < a$  – постпозиции.

Будем говорить, что сочинительная связь  $f$  содержит разделитель  $c$ , а разделитель  $c$  принадлежит сочинительной связи  $f$ .

(1) *Запрещается<sup>1</sup> повреждать<sup>2</sup> или<sup>3</sup> загрязнять<sup>4</sup> покрытие<sup>5</sup> дорог.<sup>6</sup>*

Союз *или* принадлежит сочинительной связи

$$f = \langle 1, 2, 3, 4 \rangle,$$

показанной на рис. 1. Стрелка обозначает подчинительную связь. Символ | обозначает разделитель однородных членов.

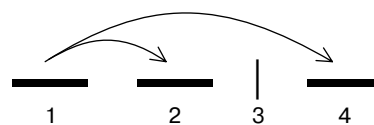


Рис. 1. Пример сочинительной связи.

Для моделирования сочинительной связи между предложениями в составе сложносочиненного предложения будем использовать виртуальное (воображаемое) главное слово с нулевым номером, у которого в подчинении находятся сказуемые  $b$  и  $d$  сочиняемых предложений:

$$f = \langle 0, b, c, d \rangle,$$

где  $c$  — союз или запятая, разделяющая предложения в составе сложноподчиненного.

(2) *Пришел, увидел, победил.*

В предложении присутствуют конструкции

$$f_1 = \langle 0, 1, 1.5, 2 \rangle \text{ и } f_2 = \langle 0, 2, 2.5, 3 \rangle.$$

Рассмотрим модель обособления. Обособление является общим понятием, обозначающим деепричастный, причастный, адъективный и другие обороты. Для целей настоящего исследования к обособлениям будем также относить и придаточные предложения. Для обособления обязательным является требование парности знаков препинания. Как правило, используются запятые, открывающая и закрывающая. В случаях, когда конец обособления совпадает с концом предложения, закрывающим знаком препинания для обособления служит знак конца предложения, например, точка.

При моделировании обособления существенными являются позиции знаков препинания — открывающего и закрывающего, а также позиции опорного и главного слов обособления, между которыми устанавливается подчинительная связь (по направлению от опорного к главному). Например, для обособляемого причастного оборота главным является причастие, а опорным — существительное в препозиции.

Таким образом, обособление также моделируется четверкой чисел:

$$g = \langle w, x, y, z \rangle,$$

где  $w$  — опорное слово обособления;  
 $y$  — главное слово обособления;  
 $x$  — открывающий разделитель (запятая);  
 $z$  — закрывающий разделитель (запятая, точка с запятой или знаки конца предложения).

Кроме того,  $x$  и  $z$  могут быть парными тире, скобками или кавычками.

Для  $g$  выполняется следующее соотношение:

$$x < y < z \ \& \ (w < x \vee z < w).$$

Вариант  $w < x$  соответствует препозиции опорного слова,  $z < w$  — постпозиции.

Будем говорить, что обособление  $g$  содержит разделители  $x$  и  $z$ , а разделители  $x$  и  $z$  принадлежат обособлению  $g$ .

(3) *Водители<sup>1</sup> транспортных<sup>2</sup> средств<sup>3</sup> с<sup>4</sup> включенным<sup>5</sup> проблесковым<sup>6</sup> маячком<sup>7</sup> синего<sup>8</sup> цвета,<sup>9</sup> выполняя<sup>10</sup> неотложное<sup>11</sup> служебное<sup>12</sup> задание,<sup>13</sup> могут<sup>14</sup> отступить<sup>15</sup> от<sup>16</sup> требований<sup>17</sup> настоящих<sup>18</sup> Правил<sup>19</sup>.*

Запятые принадлежат обособлению

$$g = \langle 14, 9.5, 10, 13.5 \rangle,$$

показанному на рис. 2. Скобки на рисунках и в примерах условно обозначают границы обособления.

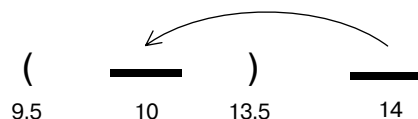


Рис. 2. Пример обособления.

Для моделирования обособления в начале предложения будем использовать виртуальную запятую с номером 0.5, расположенную между нулевым (виртуальным, воображаемым) и первым словами:

$$g = \langle w, 0.5, y, z \rangle.$$

Обособление, расположенное в конце предложения, будем моделировать четверкой чисел

$$g = \langle w, x, y, n + 0.5 \rangle,$$

где  $n$  — количество слов в предложении.

В предлагаемой модели к сочинительным союзам и ко всем знакам препинания, используемым в обособлениях и сочинительных связях, применяется общее понятие — разделитель. Будем называть *внутренним* разделитель с номером  $x$ ,  $1 < x < n$ , где  $n$  — количество слов в предложении.

### 3. Формирование множества синтаксических конструкций

Обособления и сочинительные связи для краткости будем называть конструкциями. Обозначим  $C$  множество возможных (предполагаемых) конструкций, построенных автоматически на основе заданного предложения и базы правил русского языка. Описание алгоритмов построения множества предполагаемых конструкций выходит за рамки данной работы. Ознакомиться с ними можно в отчете [12].

В силу морфологической и синтаксической омонимии множество  $C$  будет содержать как истинные (те, которые построил бы человек при правильном разборе предложения), так и ложные конструкции. Истинные конструкции  $T$ , где  $T \subseteq C$ , соответствуют смыслу предложения и являются объектом поиска для анализатора.

Для целей дальнейшего изложения представим множество конструкций  $C$  в виде  $C = R \cup I$ , где  $R$  — множество сочинительных связей,  $I$  — множество обособлений.

(4) *K<sup>1</sup> пешеходам<sup>2</sup> приравниваются<sup>3</sup> лица<sup>4</sup>, передвигающиеся<sup>5</sup> в<sup>6</sup> инвалидных<sup>7</sup> колясках<sup>8</sup> без<sup>9</sup> двигателя<sup>10</sup>, ведущие<sup>11</sup> велосипед<sup>12</sup>, мoped<sup>13</sup>, мотоцикл<sup>14</sup>, везущие<sup>15</sup> санки<sup>16</sup>, тележку<sup>17</sup>, детскую<sup>18</sup> или<sup>19</sup> инвалидную<sup>20</sup> коляску<sup>21</sup>.*

Далее приводится список предполагаемых конструкций для данного предложения. Истинные конструкции помечены в списке символом V.

<i>I</i> — множество обособлений:	<i>R</i> — множество сочинительных связей
⟨4 4.5 15 21.5⟩	⟨4 5 10.5 11⟩ V
⟨4 4.5 11 21.5⟩	⟨11 12 12.5 13⟩ V
⟨4 4.5 15 17.5⟩	⟨3 12 12.5 13⟩
⟨4 4.5 15 16.5⟩	⟨3 4 12.5 13⟩
⟨4 4.5 11 14.5⟩	⟨11 13 13.5 14⟩ V
⟨4 4.5 11 13.5⟩	⟨3 13 13.5 14⟩
⟨4 4.5 11 12.5⟩	⟨3 4 13.5 14⟩
⟨4 4.5 5 10.5⟩ V	⟨11 14 14.5 16⟩
⟨4 10.5 15 21.5⟩	⟨11 14 14.5 17⟩
⟨4 10.5 11 21.5⟩	⟨11 14 14.5 21⟩
⟨4 10.5 15 17.5⟩	⟨4 11 14.5 15⟩ V
⟨4 10.5 15 16.5⟩	⟨15 16 16.5 17⟩ V
⟨4 10.5 11 14.5⟩ V	⟨11 16 16.5 17⟩
⟨4 10.5 11 13.5⟩	⟨11 14 16.5 17⟩
⟨4 10.5 11 12.5⟩	⟨15 17 17.5 21⟩ V
⟨4 14.5 15 21.5⟩ V	⟨11 17 17.5 21⟩
⟨4 14.5 11 21.5⟩	⟨11 14 17.5 21⟩
⟨4 14.5 15 17.5⟩	⟨21 18 19 20⟩ V
⟨4 14.5 15 16.5⟩	

#### 4. Отношения между синтаксическими конструкциями

Рассмотрим отношения между конструкциями, вытекающие из их взаимного расположения в предложении. В общем случае две конструкции из *C* могут находиться в одном из трех отношений:

- быть соседними;
- быть вложенными одна в другую;
- быть несовместимыми.

Отношение вложенности конструкций можно разделить на четыре группы:

- сочинительная связь вложена в другую сочинительную связь;
- обособление вложено в сочинительную связь;
- обособление вложено в другое обособление;
- сочинительная связь вложена в обособление.

Дадим формальное определение указанных отношений.

**Определение:** Конструкции

$$f = \langle a, b, c, d \rangle \text{ и } g = \langle w, x, y, z \rangle$$

находятся в отношении соседства  $S, S \subseteq C \times C$ , если

$$\max(a, b, c, d) \leq \min(w, x, y, z) \quad \& \quad f, g \in R \rightarrow d \neq x \vee \min(a, b, c, d) \geq \max(w, x, y, z) \quad \& \quad f, g \in R \rightarrow z \neq b.$$

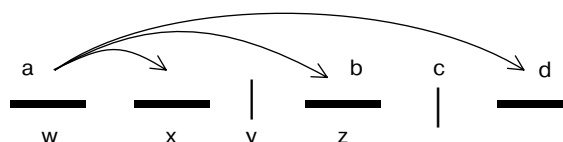
**Определение:** Конструкции *f* и *g* находятся в отношении вложенности  $M, M \subseteq C \times C$ , в следующих случаях:

а) Пусть  $f = \langle a, b, c, d \rangle$  и  $g = \langle w, x, y, z \rangle$  — сочинительные связи.

Тогда *g* вложена в *f*, если

$$a \leq w \leq b \quad \& \quad a < x \quad \& \quad z \leq b \quad \& \quad (z = b \rightarrow a = w) \vee b \leq w < c \quad \& \quad b < x \quad \& \quad z < c \vee c < w \leq d \quad \& \quad c < x \quad \& \quad z < d \vee d \leq w \leq a \quad \& \quad d \leq x \quad \& \quad z < a \quad \& \quad (d = x \rightarrow a = w).$$

Например, конструкции *f* и *g* могут располагаться так, как это показано на рис. 3.



**Рис. 3.** Сочинительная связь вложена в сочинительную связь.

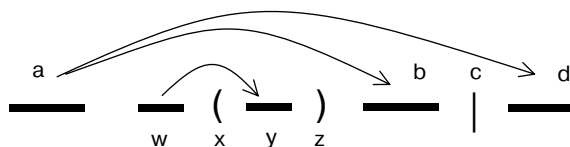
б) Пусть  $f = \langle a, b, c, d \rangle$  — сочинительная связь,

$$g = \langle w, x, y, z \rangle \text{ — обособление.}$$

Тогда *g* вложена в *f*, если

$$a \leq w \leq b \quad \& \quad a < x \quad \& \quad z < b \vee b \leq w < c \quad \& \quad b < x \quad \& \quad z \leq c \vee c < w \leq d \quad \& \quad c \leq x \quad \& \quad z < d \vee d \leq w \leq a \quad \& \quad d < x \quad \& \quad z < a.$$

Одно из возможных расположений показано на рис. 4.

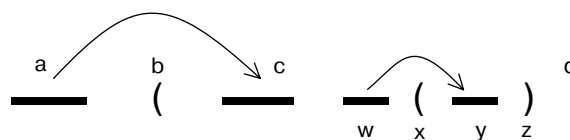


**Рис. 4.** Обособление вложено в сочинительную связь.

в) Пусть  $f = \langle a, b, c, d \rangle$  и  $g = \langle w, x, y, z \rangle$  — обособления.

Тогда *g* вложена в *f*, если

$$a \leq w < b \quad \& \quad a < x \quad \& \quad z \leq b \vee b < w \leq c \quad \& \quad b < x \quad \& \quad z < c \vee c \leq w < d \quad \& \quad c < x \quad \& \quad z \leq d \vee d < w \leq a \quad \& \quad d \leq x \quad \& \quad z < a.$$



**Рис. 5.** Обособление вложено в обособление.

д) Пусть  $f = \langle a, b, c, d \rangle$  — обособление,

$g = \langle W, X, Y, Z \rangle$  – сочинительная связь,

Тогда  $g$  вложена в  $f$ , если

$$\begin{aligned} & a \leq w < b \ \& \ a < x \ \& \ z < b \vee \\ & b < w \leq c \ \& \ b < x \ \& \ z < c \vee \\ & c \leq w < d \ \& \ c < x \ \& \ z < d \vee \\ & d < w \leq a \ \& \ d < x \ \& \ z < a \vee \\ & a = w \ \& \ b < x \ \& \ z < d. \end{aligned}$$

Наиболее характерные возможные размещения показаны на рис. 6, 7.

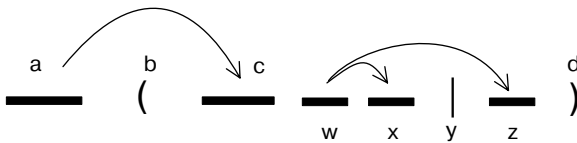


Рис. 6. Сочинительная связь вложена в обособление

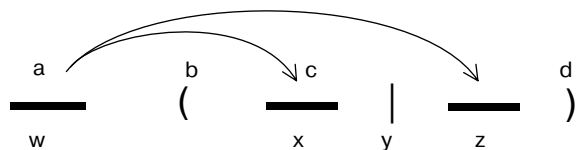


Рис. 7. Сочинительная связь вложена в обособление «Обособленный ряд».

Из определений следует, что отношения  $S$  и  $M$  не пересекаются.

Пары  $(f, g)$  из  $C \times C / (S \cup M)$  составляют несовместимые конструкции. Таким образом, любые две конструкции в предложении находятся строго в одном из трех указанных отношений. При этом любая пара истинных конструкций принадлежит строго одному из отношений  $S$  или  $M$ :  $T \times T \subseteq S \cup M$ .

### 5. Формальная постановка задачи анализа пунктуации

Определим рекурсивно понятие композиции конструкций.

**Определение.** Композицией конструкций является:

1) Любая конструкция  $f$  из  $C$ ;

Множество конструкций

$$\{f_i\} \cup f, \quad 1 \leq i \leq |\{f_i\}|, \text{ где } \{f_i\} \text{ — композиция}$$

и любая пара  $(f_i, f)$  находится в одном

из отношений:

$$\text{а) } f_i S f \quad \text{б) } f_i M f \quad \text{в) } f M f_i$$

**Определение.** Композиция конструкций содержит разделитель, если как минимум одна конструкция из композиции содержит этот разделитель.

**Определение.** Назовем покрытием такую композицию конструкций, которая содержит все внутренние разделители предложения.

Из определения следует, что композиция является покрытием независимо от того, содержит ли она виртуальный разделитель в начале предложения и знак препинания в конце предложения.

В рамках предлагаемой модели задача определения функций знаков препинания для заданного предложения русского языка сводится к построению покрытия. Следует заметить, что в общем случае покрытие строится неоднозначно.

### 6. Алгоритм построения покрытия

Рекурсивный алгоритм построения покрытия вытекает из определения композиции.

*Исходные данные:*

$C$  — множество возможных конструкций;

$|C|$  — мощность  $C$ ;

$Q$  — композиция конструкций, изначально  $Q = \emptyset$ ;

$k$  — номер конструкции в  $C$ , с которой начинается поиск, изначально  $k = 1$ ;

*Возвращаемое значение:*

*true*, если удалось построить композицию, *false* — иначе.

*Поиск покрытия*  $(C, Q, k)$

Цикл  $i = \overline{k, |C|}$  {

Если  $(\forall K \in Q : (K, C[i]) \in M \cup S)$  {

$$Q_1 = Q \cup \{C[i]\};$$

Если (Поиск покрытия  $(\tilde{N}, Q_1, i+1)$ ),  
то вернуть *true*;

}

Если ( $Q$  содержит все внутренние разделители),  
то вернуть *true* ( $Q$  является покрытием);

Иначе вернуть *false*.

Данный алгоритм является переборным и представляет лишь теоретический интерес. В работе [12] приведены эвристики, применимые на практике.

## 7. Учет пунктуации в синтаксическом анализе

Как известно, синтаксический анализ сводится к выбору подчинительных связей, образующих дерево, из числа потенциально возможных связей. Простейший способ учета пунктуации в синтаксическом анализе основан на использовании покрытия в качестве фильтра: имея покрытие, можно провести удаление значительной части ложных синтаксических связей.

Рассмотрим эту идею отдельно для обособлений и рядов, составляющих покрытие. Пусть конструкция  $f = \langle a, b, c, d \rangle$  является обособлением, т. е.  $b$  и  $d$  — это границы обособления,  $(a, c)$  — подчинительная связь. Тогда следует считать ложными и удалить все связи, конфликтующие со связью  $(a, c)$ , а также все связи, охватывающие одну из границ обособления —  $b$  или  $d$ , кроме, разумеется,  $(a, c)$ .

Пусть конструкция  $f = \langle a, b, c, d \rangle$  является сочинительной связью, т. е.  $c$  — это разделитель,  $(a, b)$  и  $(a, d)$  — подчинительные связи. Тогда следует считать ложными и удалить все связи, конфликтующие со связями  $(a, b)$  и  $(a, d)$ , а также все связи  $(x, y)$ , где  $b < x < c < y < d$ . Направление связи  $(x, y)$  не имеет значения.

Отметим, что встраивание разработанных методов анализа пунктуации в синтаксический анализ является самостоятельной задачей и сильно зависит от применяемых подходов к синтаксическому анализу.

В рамках подхода, реализованного в системе синтаксического анализа DictaScore [11, 12], покрытие строится одновременно с деревом разбора. Анализатор позволяет получить в качестве результата не только структуру подчинительных связей, но и информацию о рядах и обособлениях. При этом общая вычислительная сложность алгоритма составляет  $O(n^3)$ , где  $n$  — количество слов в анализируемом предложении. Кроме того, благодаря учету пунктуации удалось существенно поднять качество синтаксического анализа.

## 8. Заключение

Проведенные исследования позволяют сделать следующие выводы:

- Раздельный анализ обособлений и рядов однородных членов невозможен в силу функциональной омонимии запятой;
- Структура обособлений и рядов образует покрытие в любом грамматически правильном предложении русского языка;
- Разработанные модели и методы анализа пунктуации применимы к языкам, составляющим индоевропейскую группу.

Исследование проводилось малым инновационным предприятием ООО «Диктум» при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере.

## Список литературы

1. Бердичевский А. С., Иомдин Б. Л. Роль пунктуации в разрешении неоднозначности. // Труды Международной конференции Диалог'2007.
2. Мальковский М. Г., Старостин А. С. Модель синтаксиса в системе морфосинтаксического анализа «Treeton» // Труды Международной конференции Диалог'2006.
3. Кобзарева Т. Ю., Лахути Д. Г., Ножов И. М. Модель сегментации русского предложения // Труды Международной конференции Диалог'2001. Т. 2. Аксаково 2001.
4. Ножов И. М. Морфологическая и синтаксическая обработка текста (модели и программы) // М., 2003.
5. Гершензон Л. М., Панкратов Д. В. Фрагментационный анализ русского предложения в системе Artefact // Труды Международного семинара Диалог'2002.
6. Кобзарева Т. Ю. Рекурсивность и проективность сочинительных связей в русском тексте // Труды Международной конференции Диалог'2006.
7. Кобзарева Т. Ю. Омонимия и синонимия знаков препинания в русском тексте // Труды Международной конференции Диалог'2005.
8. Кобзарева Т. Ю. Построение графа связей сегментов // Труды Международной конференции Диалог'2008.
9. Розенталь Д. Э., Голуб И. Б., Теленкова М. А. Современный русский язык // Айрис Рольф. — М.: Пресс 2001 г
10. Bernard E. M. Jones. Exploring the role of punctuation in parsing natural text // Proceedings of the 15th conference on Computational linguistics'1994. — V. 1, pp. 421–425.
11. Электронный ресурс [www.dictum.ru](http://www.dictum.ru)
12. Окатьев В. В., Гергель В. П., Алексеев В. Е., Таланов В. А., Баркалов К. А., Скатов Д. С., Ерехинская Т. Н., Котов А. Е., Титова А. С. Отчет о выполнении НИОКР по теме: «Разработка пилотной версии системы синтаксического анализа русского языка» (инвентарный номер ВНИИЦ 02200803750) // М.: ВНИИЦ, 2008