

# Тайные знаки пунктуации

## Secret punctuation marks

**Окатьев В. В.** (oka@dictum.ru), **Ерехинская Т. Н.** (te@dictum.ru),  
**Ратанова Т. Е.** (rt@dictum.ru)

ООО «Диктум», г. Нижний Новгород

Вот два петуха,  
Которые будят того пастуха,  
Который бранится с коровницей строгою,  
Которая доит корову безрогую,  
Лягнувшую старого пса без хвоста,  
Который за шиворот треплет кота,  
Который пугает и ловит синицу,  
Которая часто ворует пшеницу,  
Которая в темном чулане хранится  
В доме,  
Который построил Джек,,,,,,,,,,,,,

С. Маршак

В работе предложена математическая модель пунктуации, справедливая для различных языков и применимая в различных подходах к синтаксическому анализу. Предложен метод использования разработанной модели в синтаксическом анализе на примере алгоритма Эйснера.

### Введение

Существует обширный класс текстов, содержащих грамматически правильные предложения. Это и художественная литература, и публикации в СМИ, и нормативно-правовые документы, и многое другое. Для таких текстов применимы методы формального синтаксического анализа. В настоящее время точность этих методов оставляет желать лучшего. Например, для чешского языка процент правильно построенных деревьев составляет 30,6–35,9 % [11]. Для других языков оценки отличаются ненамного.

Известно, что количество правильно построенных деревьев с увеличением длины предложения уменьшается, как указано в [19], до 9 %. Известно также, чем длиннее предложение, тем больше в нем знаков препинания [7]. Очевидна важность учета пунктуации в синтаксическом анализе [1, 9]. Однако, знаки препинания в компьютерном анализе похожи на занозы — мелкие, но очень неудобные.

Обзор подходов к анализу пунктуации приведен нами в [8]. Дополнительно отметим, что в статистических методах анализа [11, 13, 16, 19] знаки

препинания могут лишь влиять на веса связей при машинном обучении. Подходы на основе грамматик [17, 18] позволяют описывать пунктуацию достаточно подробно, однако сам процесс описания — чрезвычайно кропотливый. Кроме того, эти правила, в основном, интегрированы с правилами, описывающими связи, что затрудняет их перенос на другие естественные языки.

В обзоре [15] отмечается отсутствие математической модели пунктуации, как общий недостаток известных подходов к анализу пунктуации. Указанное замечание по нашему мнению остается справедливым до настоящего времени. Качественный учет пунктуации представляется сомнительным без адекватной математической модели.

Математическая модель пунктуации, чтобы быть применимой для формального анализа текстов, должна компактно описывать такие явления, как омонимия знаков препинания, их многофункциональность, транспозиция, утрата функциональной самостоятельности и др. Обзор этих явлений дан Кобзаревой в работе [4], которая, по мнению автора, «необходима как лингвистический базис любых

актуальных задач анализа русского текста». Однако, в указанной работе не было предложено математической модели. Более ранний обзор для английского языка составил Nunberg в [10].

Построение модели с указанными свойствами было начато нами в [8], где приводится математическая постановка задачи анализа пунктуации, основанная на формальной модели синтаксической конструкции. Там же дано формальное описание вложенных конструкций, в частности вложенных рядов однородных членов и обособлений.

В данной работе, являющейся продолжением [8], предложена модель пунктуации и схема ее применения в формальном синтаксическом анализе на примере алгоритма Эйснера.

## Пунктуация vs Математика

Границы обособлений в естественном языке принято обозначать парными знаками препинания. «Пунктуально» выполняя данное правило, в конце стихотворения «Дом, который построил Джек» Маршака следовало бы перед точкой поставить 9 (девять) запятых — по числу вложенных придаточных предложений.

Желание явно обозначить правую границу каждого обособления, по всей видимости, берет начало из аналогии с алгебраическими выражениями, где каждой левой скобке соответствует своя правая. Например,  $a*(b-(c-(d+e)))$ . На такую аналогию указывают также Nunberg [11] и Кобзарева [4]. Однозначность в алгебраических выражениях сильно облегчает их синтаксический разбор.

Представим себе пунктуацию, удобную для компьютерного анализа, т. е. построенную по аналогии с алгебраическими выражениями. Для этого достаточно трех символов: “(”, “)” и “|”. Каждое обособление будем выделять своей индивидуальной парой круглых скобок, а между однородными членами будем ставить разделитель “|”. Таким образом, «знаки препинания» будут прямо соответствовать той роли, которую они выполняют в предложении: выделяющей и разделяющей. Тогда, например, предложение:

(71) *К пешеходам приравняются лица, передвигающиеся в инвалидных колясках без двигателя, ведущие велосипед, мопед, мотоцикл, везущие санки, тележку, детскую или инвалидную коляску.*

запишется следующим образом:

(72) *(К пешеходам приравняются лица (передвигающиеся в инвалидных колясках без двигателя) | (ведущие велосипед | мопед | мотоцикл) | (везущие санки | тележку | детскую | инвалидную коляску))*

Замена знаков препинания их ролями вносит определенность в синтаксическую структуру предложения. Конечно, такая «пунктуация» не снимает омонимию полностью, в частности для слова «детскую» в последнем примере, но, без сомнения, она могла бы ощутимо облегчить синтаксический анализ и проблема фрагментации предложений [2,3,5] не была бы такой серьезной. Заметим, что вложенные конструкции — однородные члены и обособления — подробно рассмотрены нами в [8], поэтому здесь мы на них не останавливаемся.

Если же применить обратную аналогию — от пунктуации к математике, то получим выражение, которое покоробит любого цивилизованного человека:  $a*(b-(c-(d+e)$ . Куда девались скобки? Может быть, их поглотила точка?

К сожалению, в пунктуации нет однозначности, присущей алгебраическому выражению. Однако же, аналогия с математикой может оказаться весьма продуктивной для целей математического моделирования пунктуации, что и будет показано ниже.

## Пунктуация + Математика

Посмотрим, что будет, если границы каждого обособления обозначить своим парным знаком, добавив при необходимости виртуальные знаки. Перепишем пример 1:

(73) *^К пешеходам приравняются лица, передвигающиеся в инвалидных колясках без двигателя,, ведущие велосипед, мопед, мотоцикл,, везущие санки, тележку, детскую или инвалидную коляску,.*

Символом ^ явно обозначена левая граница (начало) предложения, для которой точка является правым закрывающим символом (сравните с испанским: ¿Donde debo firmar?). Три стоящие рядом запятые выполняют три различные функции, каждая — свою: закрытие причастного оборота, разделение однородных причастий и открытие следующего причастного оборота.

В [8] были введены обозначения ролей знаков препинания и сочинительных союзов. Их всего три: “(” и “)” — границы обособлений, “|” — разделитель однородных членов. Здесь мы также будем использовать эти обозначения, как интуитивно понятные и максимально соответствующие их ролям. Конечно, из-за идентичности скобок, как парного выделяющего знака, и скобок, как обозначения ролей возможна путаница. Но мы старались строить изложение таким образом, чтобы смысл этих символов был понятен из контекста.

Итак, если в примере 3 заменить знаки препинания и союзы на обозначения их ролей, то получим

ту же запись, что и в примере 2. То есть, мы получим ту самую, удобную для компьютерного анализа, пунктуацию, где каждый знак выполняет строго одну роль.

Пример 3 (и эпиграф) интересны еще и тем, что их можно привести в соответствие с общепринятыми нормами грамматики с помощью последовательного применения двух формальных операций «сложения»:

$$, + , = , \quad (1)$$

$$, + . = . \quad (2)$$

Данные операции применяются к паре соседних знаков препинания, в результате чего виртуальный знак исчезает — поглощается соседом. Такой процесс поглощения виртуальных знаков Nunberg называл «absorption», Кобзарева — «стяжение». При этом, важно иметь в виду, что функцию исчезнувшего знака принимает на себя оставшийся знак препинания.

Однако, не будем торопиться с ликвидацией виртуальных знаков и рассмотрим более подробно свойства виртуальной пунктуации. Прежде всего, в такой пунктуации отсутствует явление многофункциональности: знаки препинания не могут выполнять несколько функций одновременно. Действует принцип: один знак — одна роль в предложении.

### ^ Разделяй и властвуй!

У знаков препинания в виртуальной грамматике появляется строгая специализация. Например, знак “;” уже не может выполнять роль правой границы обособления.

(74) ^ *Нельзя сказать, чтобы это нежное расположение к подлости было почувствовано дамами; однако же во многих гостиницах стали говорить, что, конечно, Чичиков не первый красавец, но зато такой, как следует быть мужчине,...* (Гоголь)

Здесь роль правой границы придаточного предложения выполняет виртуальная запятая, а знак “;” выполняет исключительно свою основную роль разделителя однородных членов. В “обычной” пунктуации, из-за отсутствия виртуальной запятой, закрывающая роль для знака “;” является дополнительной.

Разделим знаки препинания на три группы — по выполняемым ими функциям:  $M^{\wedge}$  — левая граница обособления,  $M^{\prime}$  — правая граница обособления и  $M^{\downarrow}$  — разделитель однородных членов. При этом, в каждую группу будем включать только такие знаки, которые выполняют соответствующую роль как основную. Например, знак “;” попадет только в множество  $M^{\downarrow}$ . Выпишем эти множества:

$$\begin{aligned} M^{\wedge} &= \{ , ( - : ^ \} \\ M^{\prime} &= \{ , ) - . ! ? \dots \} \\ M^{\downarrow} &= \{ , ; \} \end{aligned}$$

Во избежание путаницы, опустим запятые при перечислении элементов множеств.

Как видно, такое разделение по ролям омонимии запятой не устранило: она включена во все три множества. Осталась также омонимия тире. В данной статье мы не рассматриваем случай, когда тире обозначает эллипсис. Восстановление таких эллипсисов рассмотрено в [7]. Кроме того, мы не рассматриваем точки, обозначающие сокращения, например, инициалы, точки — разделители полей в составных лексемах (даты, электронные адреса и т. д.) и другие знаки с аналогичными функциями.

В [8] было показано, что сочинительные союзы в контексте компьютерного анализа обладают теми же свойствами, что и знаки препинания. Например, союз «и» является разделителем однородных членов. На этом основании включим союз «и» в множество  $M^{\downarrow}$ .

Известны случаи, когда функциональная самостоятельность запятой утрачивается. Например, при сочинении простых предложений «и» функционирует, как единый разделяющий символ — запятая и союз “вдвоем” исполняют одну роль. В английском языке встречается так называемая оксфордская запятая перед союзом при перечислении: *The flag is red, white, and blue* [12]. То же относится и к русским противительным союзам «,а» и «,но». Добавим «,а» в качестве представителя этой группы в  $M^{\downarrow}$ .

Кроме того, в русском языке существуют конструкции, в которых союз «и» выполняет открывающую роль, что является основанием для его включения в  $M^{\wedge}$ . Например, в случае итерации: ^ ( *Случилось и ( то, ) и ( другое, ) и ( третье. )* Здесь роли союзов и знаков препинания обозначены верхним индексом. Другие союзы также могут, в случае итерации, исполнять открывающую роль, в частности союзы *или/ да/ либо/ ни/ то/ не то/ то ли*. Их также необходимо включать в  $M^{\wedge}$ , однако, чтобы не загромождать изложение, для целей построения модели ограничимся включением в  $M^{\wedge}$  только одного представителя этой группы — союза «и». Разумеется, остальные союзы можно подвергать анализу совершенно аналогично. На этих же основаниях уберем из  $M^{\prime}$  знаки *!/?/...* и оставим только точку в качестве представителя группы знаков, заканчивающих предложение.

В итоге получаем:

$$\begin{aligned} M^{\wedge} &= \{ , ( - : ^ \text{ и} \} \\ M^{\prime} &= \{ , ) - . \} \\ M^{\downarrow} &= \{ , ; \text{ и } ',a' \} \end{aligned}$$

Элементы этих множеств для краткости будем называть знаками, не делая отличия между знаками препинания и сочинительными союзами.

### Формула пунктуации

Зададимся вопросом — какие последовательности контактно расположенных знаков возможны между соседними словами, а какие — невозможны, помня при этом, что мы рассматриваем свойства виртуальной пунктуации (один знак — одна роль в предложении!). Прежде, чем ответить на него, абстрагируемся от внешнего вида знаков и зафиксируем их роли. Оказывается, всевозможные сочетания ролей в последовательности знаков, расположенных между двумя соседними словами, а также в начале и в конце предложения, описываются следующим регулярным выражением:

$$)*|?(* \tag{3}$$

Символ “\*” в шаблоне означает повторение элементов, либо, в частном случае, их отсутствие. Символ “?” означает необязательное присутствие разделительной роли.

Поскольку при “физическом” поглощении виртуального знака препинания его роль в обязательном порядке остается (она переходит к поглотившему знаку), то указанная формула будет справедливой при переходе от виртуальной пунктуации к общепринятой, в которой знаки препинания, как известно, обладают свойством многофункциональности. Таким образом, формула справедлива как для последовательности контактно расположенных знаков, так и для одиночного знака, выполняющего либо одну, либо одновременно несколько ролей.

Из формулы видно, что следующие сочетания ролей: |), (|, ||, () — невозможны. Перечислим допустимые сочетания соседних ролей, их будет пять: ))), ((, |), |(, |(, (.

К случаю “))” относятся формулы поглощения (1) и (2). Для описания взаимодействия друг с другом всех элементов множества  $M^1$ , удобнее будет составить таблицу  $M^1 \times M^1$ .

Табл. 1

| $M^1 \times M^1$ | ,  | )  | -  | .  |
|------------------|----|----|----|----|
| ,                | ,  | )  | ,- | .  |
| )                | ), | )) | )- | ). |
| -                | ,  | )  | -  | .  |
| .                | #  | #  | #  | #  |

В левом столбце таблицы перечисляются элементы множества  $M^1$ , выполняющие роль первой закрывающей скобки, а в верхней строке — элементы множества  $M^1$ , выполняющие роль второй закрывающей скобки. На пересечении строки и столбца записан результат взаимодействия. Например, точка, оставшаяся после поглощения запятой по формуле (2), находится на пересечении второй строки и пятого столбца. В случае, если ре-

зультат операции не определен, в клетке ставится символ “#”.

Таблицы для остальных случаев составляются аналогично.

Табл. 2

| $M_c \times M_c$ | ,  | (  | -  | : | и  | ^ |
|------------------|----|----|----|---|----|---|
| ,                | ,  | (  | ,  | # | #  | # |
| (                | (  | (( | (  | # | #  | # |
| -                | -  | -( | -  | # | #  | # |
| :                | :  | :( | :  | # | #  | # |
| и                | и, | и( | и- | # | #  | # |
| ^                | ^  | ^( | ^  | # | ^и | # |

Табл. 3

| $M_j \times M_j$ | ,  | ;  | и  | ,a  |
|------------------|----|----|----|-----|
| ,                | ,  | ;  | ,и | ,a  |
| )                | ), | ); | )и | ),a |
| -                | ,  | ;  | -и | ,a  |
| .                | #  | #  | #  | #   |

Табл. 4

| $M_i \times M_i$ | ,   | (   | -   | : | и  | ^ |
|------------------|-----|-----|-----|---|----|---|
| ,                | ,   | (   | ,   | # | ,и | # |
| ;                | ;   | ;(  | ;   | # | ;и | # |
| и                | и,  | и(  | и-  | # | #  | # |
| ,a               | ,a, | ,a( | ,a- | # | #  | # |

Табл. 5

| $M_j \times M_c$ | ,  | (  | -  | :  | и  | ^ |
|------------------|----|----|----|----|----|---|
| ,                | ,  | (  | ,- | :  | ,и | # |
| )                | ), | )( | )- | ): | )и | # |
| -                | ,- | #  | #  | :  | -и | # |
| .                | #  | #  | #  | #  | #  | # |

Последняя таблица, в частности, описывает транспозицию тире и запятой в примере: (5) Мужчины пили, спорили и хохотали, — словом, ужин был чрезвычайно весел (Пушкин).

В дальнейшем понадобится еще одна таблица — для описания парных знаков (табл. 6).

Табл. 6. Парность

| $M_c \setminus M_j$ | , | ) | - | . |
|---------------------|---|---|---|---|
| ,                   | + |   |   |   |
| (                   |   | + |   |   |
| -                   |   |   | + |   |
| :                   |   |   | + |   |
| и                   | + |   |   |   |
| ^                   |   |   |   | + |

Она отличается от предыдущих таблиц: в клетке указан “+”, если соответствующие символы могут выделять обособление.

## Применение модели пунктуации в синтаксическом анализе

Итак, переходим к ключевому вопросу — применение разработанной модели пунктуации в формальном синтаксическом анализе. Применение модели будет показано на примере модификации алгоритма Эйснера [14], который в исходном виде наличие знаков препинания позволяет учитывать лишь с помощью весов связей.

В указанном алгоритме объединение противоположно направленных путей  $p^+ = a, g, \dots$  и  $p^- = b, \dots, d$  под связью  $(a, b)$  иллюстрируется двумя треугольниками, скошенными навстречу друг другу, как показано на рисунке 1. Каждому пути  $p$  соответствует поддерево, включающее  $p$  и дуги, покрытые дугами из  $p$ .

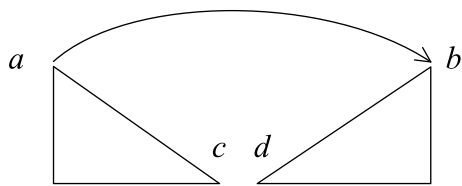


Рис. 1. Иллюстрация к алгоритму Эйснера

Стык треугольников соответствует паре соседних слов  $c$  и  $d$ , являющихся конечными вершинами встречных путей. В частном случае пути вырождены и представлены одной вершиной, т. е.  $a=c$  и/или  $b=d$ . Между словами  $c$  и  $d$  могут находиться знаки препинания и/или сочинительный союз. Их роли необходимо определить именно при объединении путей  $p^+$  и  $p^-$  под связью  $(a, b)$ .

Рассмотрим путь  $p^+$ . Отдельные дуги, входящие в путь, могут покрывать знаки препинания и союзы. К тому моменту, когда  $p^+$  сформирован, их роли уже известны. Нас будут интересовать только те знаки, которые покрываются дугой из  $p^+$  непосредственно, т. е. *другие* покрывающие (нижележащие) дуги отсутствуют. По построению среди них будут знаки с разделяющей ролью и открывающей ролью. Закрывающих знаков не будет.

Открывающие знаки, непосредственно покрытые дугами из  $p^+$ , могут иметь свой парный закрывающий знак исключительно на стыке двух встречных путей, т. е. между словами  $c$  и  $d$  (Рис. 2).

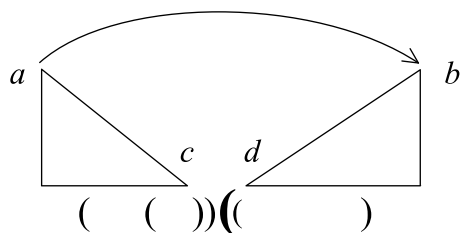


Рис. 2. Обособления

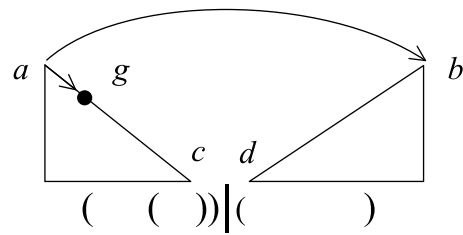


Рис. 3. Сочинение

Аналогично (или симметрично) рассуждая относительно встречного пути  $p^-$ , приходим к выводу, что закрывающие знаки, непосредственно покрытые дугами из  $p^+$ , могут иметь свой парный открывающий знак также исключительно между  $c$  и  $d$  (Рис. 2).

Далее, в зависимости от дуги  $(a, b)$ , возможны три варианта:

1. дуга  $(a, b)$  непосредственно покрывает левую границу обособления (например, связь от существительного к причастию) и предполагает открывающий знак (Рис. 2);
2. вершины  $b$  и  $g$ , где  $(a, g) \in p^+$ , соответствуют однородным членам, которые предполагают разделительный знак (Рис. 3);
3. дуга  $(a, b)$  не предполагает знака препинания или союза.

Отметим, что условия 1) и 2) могут выполняться одновременно.

Таким образом, в конце пути  $p^+$  при его построении необходимо добавлять парные закрывающие знаки, в соответствии с таблицей 6, отдельно для каждого открывающего, непосредственно покрытого дугой из  $p^+$ . В результате получится последовательность виртуальных закрывающих знаков  $)^*$ . Аналогично в конце  $p^-$  сформируется последовательность виртуальных открывающих знаков  $(^*$ . Между этими последовательностями в зависимости от дуги  $(a, b)$ , возможно, понадобится вставка разделителя и/или открывающего знака  $|?(? ( с учетом фактических знаков между  $c$  и  $d$ . В итоге получаем некую ожидаемую последовательность знаков, роли которых соответствуют формуле (3).$

Далее, с помощью таблиц 1–5 приводим ожидаемую последовательность знаков к фактически имеющимся между  $c$  и  $d$ . Если приведение завершается успешно, то знакам препинания между  $c$  и  $d$  назначаются роли, а дуга  $(a, b)$  получает вес в соответствии с алгоритмом Эйснера. В противном случае к весу дуги  $(a, b)$  добавляется штраф, который сводит к минимуму возможность включения рассматриваемой конструкции в дерево синтаксического разбора.

Для случая левонаправленной дуги  $(b, a)$  рассуждения строятся аналогично, а результат будет симметричным.

Отдельно следует остановиться на одном важном случае — пунктуации в конце предложения. В этом случае имеется только путь  $p^+$ , в конце ко-

того сформирована последовательность закрывающих знаков. В алгоритме Эйснера анализируемое предложение содержит формальное слово *root* с номером ноль. Поскольку дуга  $(root, x)$  всегда непосредственно покрывает начало предложения “^”, то парным для него следует назначить фактический знак в конце предложения из множества { . ? ! ... }. После того, как построен путь от *root* к последнему слову предложения, производим операцию поглощения по таблице 1. Для примера из эпиграфа получим:

»»»»»»»»»» → .

### Результаты экспериментов

Тестирование синтаксических анализаторов русского языка затруднено из-за неудобного доступа к корпусу размеченных текстов, который на сайте [ruscorpora.ru](http://ruscorpora.ru) доступен лишь в режиме online. По-видимому, этим объясняется отсутствие количественных результатов экспериментов в публикациях о синтаксическом анализе русского языка.

В работах [16, 19] описан близкий подход, также основанный на поиске оптимального дерева. В указанных работах приведены результаты экспериментов для различных европейских языков, в том числе чешского — одного из славянских языков. Количество правильно установленных связей в среднем составляет 82,26 %. Процент правильно

разобранных предложений не превышает 40 %. Оба показателя существенно снижаются с ростом длины предложения.

Применение разработанного подхода в системе DictaScore позволило добиться показателей 77,3 % правильно построенных деревьев и 92,1 % правильно установленных связей для предложений, имеющих длину более 10 слов и содержащих хотя бы один знак препинания. Для экспериментов были случайным образом выбраны 400 предложений из правовых и новостных текстов.

### Заключение

Предложенная модификация алгоритма Эйснера позволяет построить дерево разбора с учетом пунктуации, выявить функции знаков препинания и сочинительных союзов и определить тем самым фрагментацию предложения. Фрагментация, как промежуточный этап, не требуется.

В данной работе усилия авторов были сосредоточены, прежде всего, на разработке модели, пригодной для анализа пунктуации в различных языках и применимой в различных подходах к синтаксическому анализу. Содержимое отдельных клеток в приведенных таблицах, возможно, требует уточнения.

Предложенные модели и методы используются при разработке синтаксического анализатора DictaScore.

## Литература

1. Бердичевский А. С., Иомдин Б. Л. Роль пунктуации в разрешении неоднозначности. //Труды Международной конференции Диалог'2007.
2. Гершензон Л. М., Панкратов Д. В. Фрагментационный анализ русского предложения в системе Artefact //Труды Международного семинара Диалог'2002.
3. Кобзарева Т. Ю., Лахути Д. Г., Ножов И. М. Модель сегментации русского предложения // Труды Международной конференции Диалог'2001. Т. 2. Аксаково 2001.
4. Кобзарева Т. Ю. Омонимия и синонимия знаков препинания в русском тексте // Труды Международной конференции Диалог'2005.
5. Кобзарева Т. Ю. Построение графа связей сегментов // Труды Международной конференции Диалог'2008.
6. Электронный ресурс [www.dictum.ru](http://www.dictum.ru)
7. Окательев 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
8. Окательев В. В., Ерехинская Т. Н., Скатов Д. С. Модели и методы учета пунктуации при синтаксическом анализе предложения русского языка // Труды Международной конференции Диалог'2009.
9. 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.
10. Nunberg, G. The Linguistics of Punctuation. CSLI Lecture Notes, No. 18. Stanford: Center for the Study of Language and Information. 1990.
11. McDonald, R., Pereira, F. (2006). Online Learning of Approximate Dependency Parsing Algorithms. Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Trento, Italy, pp. 81–88.
12. Truss, L. 2003. Eats, Shoots and Leaves. Profile Books. London.
13. Kübler, S., McDonald, R., Nivre, J. 2009. Dependency Parsing. Synthesis Lectures on Human Language Technologies, G. Hirst (ed.) Morgan & Claypool Publishers.
14. Eisner, J. (1996) Three new probabilistic models for dependency parsing: An exploration. In: Proceedings of the 16th International Conference on Computational Linguistics (COLING), pp 340–345
15. Say, B. and Akman, V. (1997) Current approaches to punctuation in Computational Linguistics. Computers and the Humanities, 30, pp. 457–469.
16. Nivre, J. and McDonald, R. (2008) Integrating graph-based and transition-based dependency parsers. In Proceedings of ACL-HLT.
17. Sleator, Daniel D. and Temperley, D. (1993) Parsing English with a link grammar. In Proc. Third International Workshop on Parsing Technologies, pp. 277–292.
18. Dekang Lin. 1998a. Dependency-based evaluation of MINIPAR. In Workshop on the Evaluation of Parsing Systems, Proceedings of the First International Conference on Language Resources and Evaluation, p. 234–241, Granada, Spain, 28–30 May.
19. McDonald, R., Pereira, F. Discriminative learning and spanning tree algorithms for dependency parsing, University of Pennsylvania, Philadelphia, PA, 2006