

ИНТЕРПРЕТАЦИЯ ЗНАКОВ ПРЕПИНАНИЯ ПРИ АВТОМАТИЧЕСКОМ АНАЛИЗЕ РУССКОГО ТЕКСТА

При автоматическом анализе структуры текста очень важна работа со знаками препинания (ЗП), так как их правильная интерпретация способствует дальнейшей сегментации предложения без порождения множества неправильных вариантов структуры предложения.

Для проведения исследования был создан контрольный корпус русских текстов.

При автоматической обработке структуры текста на этапе сегментационного анализа могут возникать проблемы, связанные с интерпретацией ЗП в предложении.

Во-первых, проблемы будут связаны с **омонимией ЗП**,

При анализе текста важно учитывать, что ЗП могут выполнять две противоположные функции:

- связывать слова в проективные отрезки (например, запятая как оператор сочинения слов) внутри сегмента;
- выступать в качестве операторов членения – границ сегментов, эксплицитно заданных в тексте соответственно правилам русской пунктуации.

Для построения проективных отрезков именных, предложных и сочиненных групп, как и для собственно построения сегментов, необходимо знать исходные функции конкретных ЗП и контекстные условия их дешифровки.

Одной из чрезвычайно значимых особенностей именно письменного текста является проективность подчинительных и сочинительных связей внутри сегмента, из которых следует проективность сегмента, т. е. то, что любое слово сегмента, кроме одного (вершины или "скрепы"), может находиться в отношениях подчинения и/или сочинения только со словами своего сегмента. Именно проективность сегментов позволяет восстанавливать их целостность при разрывах сегментов вложением в них сегментов-слуг или вводных оборотов. Восстановление невозможно без представлений об омонимии и синонимии ЗП, об их способности одновременно исполнять несколько функций в предложении и о возможной неоднозначности ЗП, порождающей неоднозначность сегментации предложения.

Существуют линейные структуры, где рассматриваемый ЗП имеет единственное значение. В других структурах рассматриваемый ЗП имеет в данном случае одновременно два значения (совмещаются две его функции), т.е. мы говорим о бифункциональности ЗП, о поглощении ЗП.

Рассмотрим, например, такой ЗП, как запятая. Функция 1 (ф1) и функция 2 (ф2) запятой – левая и правая границы любого сегмента, вложенного в другой сегмент; функция 3 (ф3) – оператор сочинения внутри сегмента.

Бифункция 1 (бф1) – совмещение функций 1 и 2; бифункция 2 (бф2) – совмещение функций 2 и 3. Например,

Трое крестьян должны, ^{ф1} совершив нападение, ^{ф2} немедленно исчезнуть [2, с. 43].

Время было выбрано с таким учетом, ^{ф1} чтобы на шоссе не было пешеходов и машин, ^{бф1} которые могли бы пострадать во время операции. [2, с. 43].

Кроме того, ^{ф2} за вывоз камней, ^{ф1} добытых в нашей стране, ^{ф3} и отшлифованных, ^{ф3} и необработанных, ^{ф2} необходимо платить огромную пошлину [1, с. 19].

Вьетнамские власти заявили, ^{ф1} что они были похищены бандитами, ^{ф2} и, ^{ф1} выразив свое сожаление посольству, ^{ф2} сказали, ^{ф1} что ничем не могут помочь [2, с. 48].

Дело настолько тревожное, ^{ф1} что его следует отнести к разряду преступлений государственной важности, ^{бф2} и заниматься им помимо прокуратуры должен оперативный отдел жандармерии под личным контролем господина обер-полицейстера [3, с. 7].

Сегментная омонимия конкретной линейной структуры соответствует неразрешимой омонимии ЗП в этом предложении и может быть выявлена только на уровне снятия омонимии ЗП.

В предложении типа *Это было воспоминание, от которого он вздрогнул, когда отец встал, взял чашку и налил в нее чая, открыл форточку и посмотрел на часы* наблюдаются неоднозначности сегментации при определении значения запятой – правой границы сегмента / оператора сочинения. Все интерпретации сегментации синтаксически равноправны и без экстралингвистической информации равно истинны. В данном случае неоднозначность возникла вследствие одинаковых граемемм сказуемых.

Омонимия возникает и при интерпретации ЗП при определении левой границы вложенных сегментов. Например, в предложении типа *Он видел соседа, работающего в огороде отца, брата и деда* падежная омонимия Р.п./В.п. *отца* позволяет проинтерпретировать первую запятую как омоним [левая граница обособленного причастного оборота/оператор сочинения]. Возникает неоднозначность сегментации: 1) все предложение – один сегмент с сочиненной группой *сосед, отец, брат и дед*. 2) первая запятая – левая граница причастного оборота, и все предложение состоит из двух сегментов. При этом омонимия второй запятой [правая граница/оператор сочинения] тоже не снимается и порождает два варианта сегментации: а) *Он видел соседа, работающего в огороде отца, брата и деда*, где вторая запятая – граница сегмента, и б) *Он видел соседа, работающего в огороде отца, брата и деда*, где вторая запятая – оператор сочинения.

Если учесть тот факт, что описанные конструкции, способствующие развитию омонимии ЗП, употребляются преимущественно в книжных стилях, а также принять во внимание, что количество разрывов способствует неоднозначности, то можно прогнозировать встречаемость подобных проблем чаще в текстах художественного и публицистического стиля.

Во-вторых, проблемы возникнут при интерпретации авторских знаков препинания.

Часто авторы прибегают к использованию ЗП с целью смысловой организации предложения, подчиненной творческой задумке, стилю автора. Такую пунктуацию называют "авторской". Элементы подобного использования пунктуации отмечены в своде правил как разрешенные в некоторых позициях при условии, что они направлены на акцентирование внимания последующей после знака информации или для передачи эмоциональности, например: *Он – консультант с правом ничего не решающего голоса. Какая связь между настроением, душой и магазинами – мужчине не понять* [5].

В следующих примерах ЗП также выполняют вышеуказанные функции, но подобные позиции не прописаны в своде правил:

Там сидела лишь та неподвижная дама средних лет – выпрямившись и не обращая никакого внимания на происходившее вокруг [1, с. 20].

Он поднял пухлую ладонь и прижал к ней пальцы – будто щупал пульс невидимого рынка купли-продажи драгоценных камней [1, с. 21].

Джо тоже вздохнул – с облегчением [1, с. 51].

Иногда на его поверхности проскальзывали едва различимые язычки пламени, – неподвижное жгучее солнце заливало белым пыльным светом всё вокруг [4, с. 1].

Мы облазили его весь, а Сашка даже пытался залезть в трубу, но боялся гнева матери... очень уж несло копотью. [4, с. 2].

И – кубарем – с четвёртого этажа вниз... где уже мается Генка, потерявший друзей [4, с. 4].

Отметим, что, проанализировав тексты в экспериментальном корпусе, мы отметили подобные случаи использования ЗП только в текстах художественного стиля. Среди использованных в данных употреблениях ЗП выявлены 2 типа знаков – многоточие и тире. Принимая во внимание позиции знаков, можно сделать такие выводы:

1) некоторые знаки можно опустить и не учитывать при сегментации, в частности используемые для выделения сегментов без предикатной вершины;

2) некоторые знаки могут быть заменены на обычные для данных употреблений запятые, в частности для сегментов с предикатными вершинами.

ЛИТЕРАТУРА

1. Алекс, Джо. Ад во мне / Джо Алекс – Мн., 1994. – 272 с.
2. Чейз, Дж. Х. Венок из лотоса: Остросюжет. детектив / Дж. Х. Чейз; пер. с англ. – Мн., 1990. – 172 с.
3. Акунин, Б. Особые поручения: Декоратор. Приключения Эраста Фандорина – /Б. Акунин [Электронный ресурс]/. – Москва, 1999. – Режим доступа: <http://lib.aldebaran.ru>. Дата доступа: 12.06.2004.
4. Кот Ирвинг Стивенс, эсквайр. Я – после смерти / Кот Ирвинг Стивенс, эсквайр [Электронный ресурс]/. – Москва, 2000. – Режим доступа: http://zhurnal.lib.ru/h/husnullin_a_s/aaa.shtml – Дата доступа: 12.06.2004.

5. Апатова, А. Почему мужчины не любят ходить по магазинам? / А. Апатова [Электронный ресурс] /. – Москва, 2004. – Режим доступа: <http://www.strasti.ru> – Дата доступа: 16.05.2004
6. Гершензон, Л.М, Ножов, И.М., Панкратов, Д.В., Сокирко, А.В. Синтаксический анализ в системе РМЛ [Электронный ресурс]. – Москва, 2006. – Режим доступа: <http://www.aot.ru/docs/synan.html> – Дата доступа: 25.01.2008
7. Кустова, Г.И., Ляшевская, О.Н., Падучева, Е.В., Рахилина, Е.В. Национальный корпус русского языка как инструмент семантико-грамматического исследования лексики / Международная конференция "Корпусная лингвистика – 2004". Тезисы докладов. – СПб., 2004. С. 50-51.
8. Кобзарева Т.Ю. Принципы сегментационного анализа русского предложения. Московский лингвистический журнал., 2004, Т.8, №1, М.: РГГУ, с. 31-80.