

РЕКУРСИВНОСТЬ И ПРОЕКТИВНОСТЬ СОЧИНИТЕЛЬНЫХ СВЯЗЕЙ В РУССКОМ ТЕКСТЕ¹

CO-ORDINATION RECURSIVENESS AND PROJECTIVITY IN RUSSIAN TEXT

*Т.Ю. Кобзарева (stamstam@mtu-net.ru)
РГГУ, Москва*

Анализ сочинительных связей – необходимая составляющая автоматического синтаксического анализа. Рассматриваются свойства структуры предложения при сочинительном сокращении: проективность и рекурсивность сочинительных отношений в структуре предложения. Эти свойства в значительной степени определяют зоны анализа при построении сочинительных и подчинительных связей и систему предпочтений как на этапе сегментации (при поиске простых и придаточных предложений и обособляемых оборотов всех типов) при снятии функциональной омонимии знаков препинания и сочинительных союзов, так и при внутрисегментном анализе.

1. Введение

Свойства сочинения рассматриваются с точки зрения линейной организации русского предложения при сочинении. Постановка вопроса обусловлена проблемами поверхностно-синтаксического анализа (ПСА) предложения, конечной целью которого является представления предложения (S) в виде графа отношений между его компонентами².

Рассмотрены особенности структуры, значимые для стратегии ПСА, принятой в разрабатываемой системе, где свойства проективности и рекурсивности линейной структуры текста существенно определяют организацию динамической составляющей анализа и, соответственно, то, какие именно модули и в каком порядке работают при анализе.

В линейной структуре текста различаются три вида компонент: 1) слова или морфологически автономные группы слов – минимальные морфологически независимые группы (далее обо всех этих единицах будем говорить как о словах, так как чаще всего это слова), эксплицитно заданные пробелами [9]; 2) сегменты: простые и придаточные предложения, деепричастные и другие обороты, требующие обособления, т.е. отрезки с эксплицитно заданными границами и 3) операторы – знаки препинания, сочинительные союзы и их комбинации, служащие границами сегментов или манифестирующие связи слов или сегментов [5]; и два вида отношений слов и сегментов: направленные от хозяина к слуге отношения подчинения (несимметричные, неререфлексивные и нетранзитивные) и ненаправленные отношения сочинения (рефлексивные, симметричные и транзитивные).

Специфика системы состоит в том, что анализ организован как последовательность модулей, и есть отдельный модуль сегментации, строящий сегменты, работа которого предшествует работе модуля внутрисегментного анализа, строящего связи слов в уже построенных сегментах, и модуля межсегментного – строящего граф связей сегментов. Такой порядок работы отличается от принятого в [6,2], где внутрисегментный анализ [6] – это построение связей слов в исходных сегментах (отрезках между двумя ближайшими знаками препинания или союзами), а межсегментный [2] – построение всех остальных связей, и нет отдельного этапа построения сегментов (сегменты в нашей терминологии – это, по терминологии [6,2], – сегменты «окончательные, приведенные»). Кроме того, в отличие от [6,2,1] строятся сразу только правильные связи и истинные на поверхностно-синтаксическом уровне неоднозначности (насколько это позволяет этап разработки лингвистического базиса системы).

Обсуждаемые ниже особенности сочинения весьма существенны для организации процедур ПСА, так как анализ сочинения – одна из важных задач ПСА. В разрабатываемой системе анализ сочинения делится на три области: анализ сочинения любых слов внутри сегмента, кроме предикативных вершин сегментов, анализ сочинения сказуемых – рекурсивная процедура, используемая в ходе построения придаточных и простых-главных предложений; анализ сочинения предикативных вершин сегментов – не сказуемых и сочинения придаточных как их соподчинения, проводимые в ходе сегментации и межсегментного анализа. Суммарно анализ сочинения позволяет разрешать функциональную омонимию операторов [5] при построении сегментов и, устанавливая сочинительные отношения, помогает определить границы сегментов и границы проективных зон поиска подчинительных связей.

¹ Работа подготовлена при частичной поддержке гранта РФФИ № 03-06-80109.

² Проблема рассмотрена в рамках подхода системы ПСА, разрабатываемой в Отделении интеллектуальных систем РГГУ и представленной в [3,4].

2. Функциональные особенности структуры сочиненных групп

2.1. Сочинительное сокращение

При продуцировании отношения сочинения как отношения, возникающего в результате сочинительного сокращения, когда «вынесением за скобки» (в некоторых случаях – с определенными изменениями морфологических характеристик) общего члена сочиненных групп (*старый папа и старая мама* \Rightarrow *старые (папа и мама)*) происходит объединение нескольких исходных синтаксических структур.

В результате «вынесения за скобки» общего хозяина или слуги слова или сегменты оказываются в графе подчинительных связей функционально единой группой в структуре S по отношению к этому общему хозяину или слуге: (*загорелые и веселые*) *дети*; *В Москве (было холодно и топили печи)*; (*На столах и на лавках*) *лежал слой пыли*. При этом и сами «вынесенные за скобки» слова *мальчик и веснушчатая девочка*, (которые уже устали, с которыми никто не разговаривал и которым уже ничего не хотелось) могут быть сочиненными, так как и каждое из сочиненных придаточных здесь можно рассматривать как общий «вынесенный за скобки» член по отношению к первой группе вынесенных за скобки *мальчик и веснушчатая девочка*.

При этом каждое из сочиненных слов (как любое слово в предложении) потенциально является вершиной некоторой ветки графа, то есть может иметь собственных (не общих для всей сочиненной группы) слуг.

Таким образом, сочиненные слова в графе выступают в двух качествах:

- 1) как составляющие сочиненной группы – однородные члены, представляющие в графе функционально единое целое – один узел графа с определенными условиями согласования со своими хозяевами и слугами;
- 2) как несколько отдельных вершин – потенциальные корни веток графа, имеющие собственных слуг.

Именно с этой функциональной двойственностью сочиненных слов связана обсуждаемая ниже специфика моделирования структуры графа предложения.

2.2. Обязательная манифестация сочинения в тексте

При автоматическом анализе свойств текста, не являющегося записью или имитацией разговора, существенно, что 1) каждый знак препинания обязательно имеет хотя бы одно из двух видов значений: манифестирует сочинительную связь или/и является границей сегмента [5] и 2) сочинительная связь в тексте обязательно эксплицитно выражена оператором сочинения (Fk): знаком препинания, сочинительным союзом или их комбинацией. Таким образом, если два слова или сегмента сочинены, то в тексте обязательно существует Fk, манифестирующий их сочинительную связь.

3. Сочинительное сокращение как фактор недревовидности графа

Рассмотрим особенности структуры предложения, характерные для функционирования сочиненной группы в ее первом качестве – как единого целого в графе.

При изображении графа по Теньеру, когда сохраняется порядок следования слов по горизонтали, а слова – узлы графа располагаются на разных уровнях соответственно иерархии подчинительных связей, «для правильных синтаксических структур, изображенных в виде дерева, <...> перпендикуляры, опущенные из узлов дерева, не пересекают его ветвей» [2].

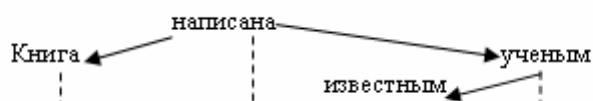


Рис.1.

При сравнении заданного таким образом свойства проективности с изображением связей над цепочкой слов и знаков S «свойство проективности, с одной стороны, и свойства непересечения и необрамления, с другой стороны, точно соответствуют друг другу...» [2].

Естественный граф связей, соответствующий структуре *Папа попрощался, и папа ушел* \Rightarrow *папа (попрощался и ушел)*, получается следующий:

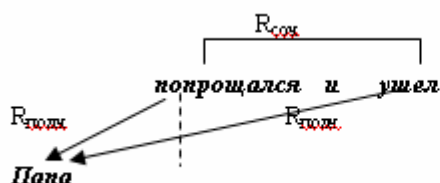


Рис.2.

Или, в ином представлении:

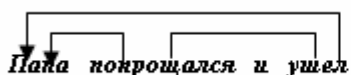


Рис.3.

Для сохранения проективности графа в подобных ситуациях [1,6] обычно вводят три условных приема фиксации связей: 1) для сочиненных групп для каждого двух сочиненных слов первое слева направо по порядку следования в тексте слово считается хозяином ближайшего следующего, 2) для фиксации связей, общих для группы сочиненных, представителем группы выступает первое слева направо сочиненное, 3а) оператор Fk – сочиняющий союз, так как сочиняющая запятая нигде в описаниях не встречается – считается подчиненным правому из сочиненных [6] или 3б) Fk подчинен левому сочиненному, а ему подчинено правое сочиненное [1].

При такой записи мы условно приравниваем симметричное, рефлексивное и транзитивное отношение сочинения нерефлексивному, нетранзитивному и несимметричному отношению подчинения, но этот способ фиксации в конструкции, например, *старые и больные папа и мама*, где сочинены и хозяева и слуги и 4 из 5-ти связей графа заданы условно, обеспечивает его проективность, т.е. дает возможность считать его структуру правильной:



Рис. 4

Существенно, что при этом не удастся фиксировать синтаксическую омонимию для конструкций типа *Старые мужчины и женщины; давно знаю и люблю* и др.

Если же фиксировать все имеющиеся связи, не используя искусственных приемов (подчинительные – со стрелкой, сочинительные – без стрелок), то сочинительное сокращение, даже если всю сочиненную группу рассматривать как единое целое, порождает недревовидность графа и соответствующую непроективность подчинительных связей (пунктиром – связь факультативная. Порождающая неоднозначность интерпретации):

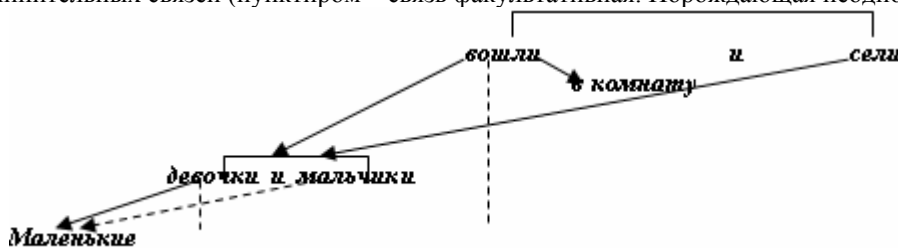


Рис.5.

Т.е. общие слуги или общие хозяева сочиненных слов нарушают древовидность и проективность структуры. В другой форме представления:

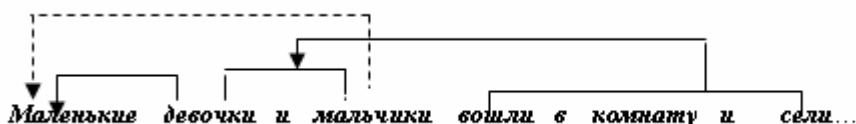


Рис.6.

Для того, чтобы иметь возможность различать синтаксическую омонимию, нужно фиксировать все связи:

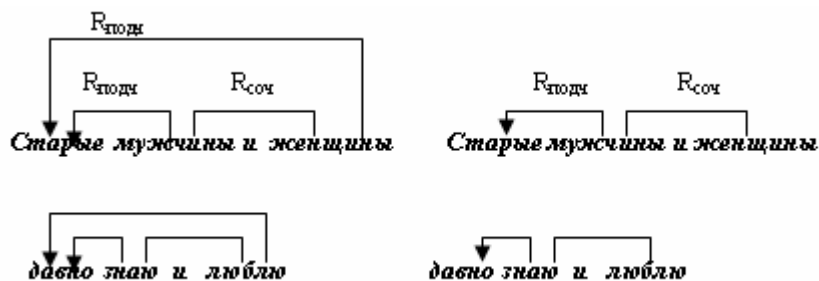


Рис.7.

Как показано в [7], общий хозяин\слуга сочиненных слов в линейной структуре предложения обычно находится вне отрезка, ограниченного сочиненными словами. Например, при вынесении за скобки согласованного определения: «загорелый мальчик и загорелая девочка» ⇒ *загорелые мальчик и девочка*, линейного фрагмента ветки сказуемого с его слугами: «мальчик поет громко, и девочка поет громко» ⇒ *мальчик и девочка поют громко* и др.

Необходимо при анализе иметь в виду, что это правило может нарушаться. Например, подлежащее при сочинении сказуемых:

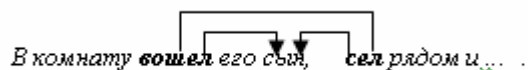


Рис.8.

4. Рекурсивность и проективность сочинительных связей

Рассмотрим особенности структуры линейного текста, связанные со вторым свойством сочиненных слов, то есть с тем, что каждое из сочиненных может иметь своих собственных слуг, т.е. является корнем некоторой ветки в графе подчинительных связей.

О рекурсивности линейной структуры S , используя понятия гнездования, вставления и самовставления, писал еще Хомский [8], рассматривая эту особенность с точки зрения подчинительных отношений.

В примере *Любящий путешествовать и выступать и знающий себе цену, этот артист...* сочиненная группа *путешествовать и выступать* вставлена в сочиненную группу *Любящий ... и знающий*. В линейной структуре S возникает «матрешка» из сочиненных групп.

Рассмотрим несколько свойств таких структур.

1-ое свойство рекурсивной структуры CC : слово, находящееся внутри сочиненной группы (т.е. между двумя сочиненными) может быть сочинено только со словом, находящимся тоже внутри:

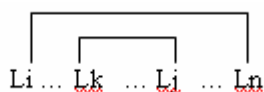


Рис.9: иллюстрирует проективность сочинительных связей: если одно из сочиненных слов группы – L_k – находится между двумя ближайшими сочиненными L_i и L_n в цепочке другой сочиненной группы, то любое слово L_j , входящее в эту сочиненную группу, должно находиться в отрезке между L_i и L_n .

Если бы слово L_k между двумя сочиненными было сочинено со словом L_j вне отрезка, ограниченного сочиненными (Рис.10), это породило бы пересечение сочинительных связей, т.е. их непроективность:

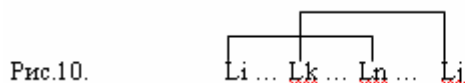


Рис.10.

1-ое свойство рекурсивной структуры CC («матрешки») иллюстрирует Рис.11 - фрагмент картинки результата анализа сочинения в модуле сегментации: CC изображены скобками под текстом (Программная реализация модуля сегментации И.М.Ножова [11])

Он услышал вокруг шум голосов, воюющий звук парадной двери и, осторожно выглянув в окошечко, увидел внизу отца, который, как мальчик, взбежал по лестнице и, не добежав до площадки, опять проворно спустился, двигая врозь коленями. (Набоков)

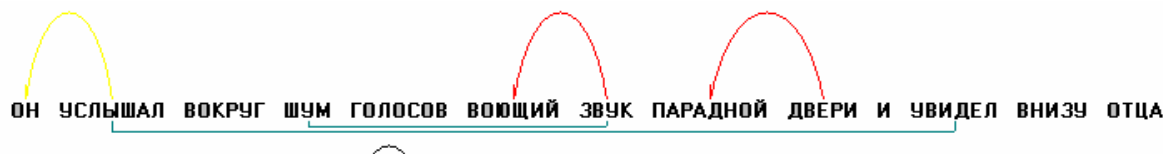


Рис.11

В рекурсивной структуре [навестил [старшего брата и заболевшую гриппом мать] и [долго и охотно рассказывал]] в сочиненную группу [навестил... F_i=и ...рассказывал] вложены две сочиненные группы: [...брата F_k=и ...мать] и [долго F_n=и охотно].

При этом существенно, что все сочиненные вложенной группы *брата* $F_k=и$ *...мать* находятся между одним из сочиненных слов объемлющей конструкции и ее оператором $F_i=и$.

2-ое свойство рекурсивной структуры CC : слово между левым или правым сочиненным объемлющей группы и оператором, ее сочиняющим, не может иметь сочиненного с ним вне отрезка предложения, ограниченного соответственно левым или правым сочиненным и сочиняющим их оператором.

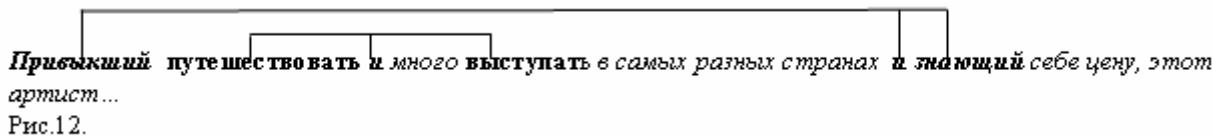
Сочиненные слова потенциально всегда являются вершинами некоторой ветки графа синтагматических отношений.

3-е свойство: если L_s – слуга сочиненного слова L_x находится между этим L_x и F_m – оператором, сочиняющим L_x с L_y , то все слова ветки графа, корнем которого является L_s , находятся между L_x и F_m .

Эти три свойства задают в линейной структуре предложения проективные отрезки рекурсивных линейных структур сочиненных конструкций.

Используя при анализе проективность рекурсивных структур сочиненных групп, легко ограничить зоны линейной структуры предложения, где могут находиться узлы веток, корни которых сочинены: если у сочиненных слов вложенной группы есть слуги, находящиеся между сочиненными словами слева или справа от их оператора сочинения F_m , то и их слуги, и слуги их слуг находятся (при условии сохранения смысла) внутри объемлющей группы соответственно слева или справа от оператора F_m : в конструкции *Привыкший*

путешествовать и много выступать в самых разных странах и знающий себе цену, этот артист... при вложении сочиненной группы в сочиненную группу сочиненные слова обрамляющей группы ограничивают область, где могут находиться слова, непосредственно или опосредованно подчиненные сочиненным вершинам вложенной группы. Кроме того, оператор сочинения – сочинительный союз или запятая – разделяют «зоны влияния» (т.е. участки линейной структуры, где могут находиться слуги этих веток. На условной схеме Рис.12 проиллюстрирована рекурсивная структура (структура каждой сочиненной конструкции изображена над текстом горизонтальными скобками с перпендикуляром внутри от оператора сочинения F_k) из двух сочиненных конструкций и порождаемое этой структурой членение текста на проективные отрезки.



Соответственно членению, проиллюстрированному условной схемой Рис.12:

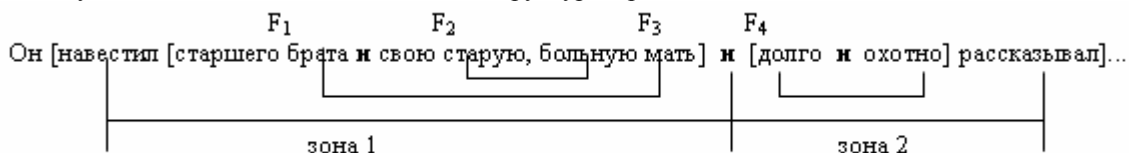
- 1) слова, находящиеся правее **Привыкший** и левее второго **и** не могут вступать в связи со словами вне этого отрезка;
- 2) слова внутри цепочки **путешествовать и ... выступать** не могут вступать в связи со словами вне нее;
- 3) слова между первым **и** и **выступать** не могут вступать в отношения со словами вне этого отрезка;
- 4) слова между **выступать** и вторым **и** не м.б. связаны со словами вне этого отрезка;

В результате зоны, внутри которых м.б. связаны слова, оказываются жестко определены: хозяином **путешествовать и выступать** м.б. только **Привыкший**; **много** м.б. только слугою **выступать**, а хозяином предложной группы *в самых разных странах* может быть только **выступать**.

5. Рекурсивность и иерархия сочинительных связей

Для того, чтобы определить, что именно сочиняет некоторый F_k , нужно выстроить иерархию сочинительных связей – систему предпочтений при поиске сочиненных слов. Насколько это возможно, и как это проявляется в линейной структуре предложения?

В линейной структуре S мы имеем дело с матрешкой сочинительных конструкций – с некоторой последовательностью вложений. При этом каждая очередная вложенная структура м.б. вложена слева или справа от оператора сочинения двух «опоясывающих» сочиненных только целиком со всеми своими слугами – узлами своего подграфа. На Рис.13 в скобках - линейные отрезки сочиненных групп, горизонтальные скобки под текстом - сочинительные связи слов, горизонтальная линия под текстом, разделенная перпендикулярами, опущенными от сказуемых, - условная схема членения линейной структуры предложения на зоны влияния сочиненных вершин)



Приведенная схема иллюстрирует линейную организацию трехъярусной «матрешки» сочиненных групп. Структуру этой сочинительной «матрешки» в первую очередь определяет то, что возникают две проективные зоны внутри проективного фрагмента сочиненной конструкции: между левым сочиненным *навестил* и оператором сочинения $F_3 = \text{и}$) и F_3 и правым сочиненным *рассказывал*.



Проективность отрезка между сочиненным словом и оператором позволяет ограничить зону поиска и сочинительных и подчинительных связей, так как любое слово между ним и оператором не может иметь подчиненных или сочиненных с ним слов за пределами этого отрезка.

При сочинении сказуемых в одном простом-главном или придаточном (т.е. когда при сочинительном сокращении за скобки вынесен субъект), если между сочиненными сказуемыми операторов несколько, их зоны влияния, т.е. отрезки, где могут находиться только слова ветки графа одного из сочиненных и не могут

находиться слова ветки другого сочиненного, определяются тем, какой именно из операторов между каждыми двумя ближайшими сказуемыми сочиняет эти сказуемые. Левее зоны 1 могут находиться слуги *навещал* и общие слуги группы сочиненных, правее зоны 2 – слуги *рассказывал* и общие.

В идеальной структуре:

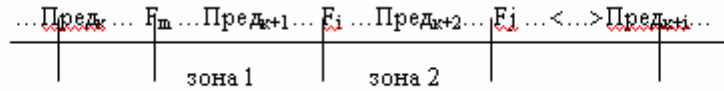


Рис.15.

где Пред_к – первый сочиненный в цепочке сочиненных, а Пред_{к+1} – последний, общие слуги и слуги Пред_к могут быть левее Пред_к, правее же Пред_{к+1} – слуги Пред_{к+1} и общие. В отрезке [Пред_к...F_м] могут находиться только слуги Пред_к, в отрезке [F_м...F_і] – только слуги Пред_{к+1}, а в отрезке [F_і...F_ј] – только слуги Пред_{к+2} и т.п.

Ситуация с сочинением деепричастных и причастных оборотов и их синтаксических аналогов аналогична (Ср.Рис.12, Рис13 и Рис.16).

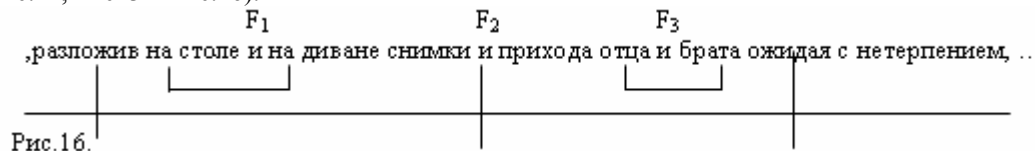


Рис.16.

Между двумя ближайшими сказуемыми не может не быть оператора. Этот оператор разграничивает зоны их влияния. Если два Vf находятся в одном простом предложении (главном или придаточном), то они сочинены, и это отношение в русском тексте обязательно должно быть выражено оператором сочинения Fk – сочинительным союзом или запятой или их комбинацией. Из этого проистекает тривиальное свойство линейной структуры с Fk: если Fk – единственный оператор между двумя сказуемыми, то этот Fk может манифестировать только связь этих глаголов, сочинять их или/и являться границей сегментов, где они – вершины. В этой ситуации в тексте, где справа и слева от Fk есть слова, которые в принципе могли бы быть им сочинены (...*прошел с чемоданами в прихожую и в комнату не заглянул*...), какие бы слова ни оказались слева и справа от этого Fk, предпочтение отдается Vf.

Аналогично, между любыми однородными вершинами сегментов единственный Fk является только границей их зон влияния, или, если у них есть общий хозяин или есть хотя бы один общий слуга, их сочиняет.

Соответственно, можно говорить о первом уровне иерархии сочинения – сочинении предикативных вершин сегментов.

Любой Fk потенциально может иметь одновременно функцию сочинения и быть границей сегмента [5], но, если он сочиняет, то сочиняет единственную пару слов. В силу этого, если между сочиненными, требующими обязательного оператора, мы находим более одного оператора, то встает задача определить, какой из операторов их сочиняет, т.е. **не сочиняет** другие слова.

При анализе каждого Fk могут возникать три рода проблем: 1) факт наличия сочиненной группы для Fk безусловен, но неоднозначен выбор сочиненных (неоднозначности типа *не видел друга брата и его отца*, на уровне предложения, как правило, неразрешимые, и здесь не рассматривается), 2) слева и справа от **каждого** из операторов стоят слова, которые потенциально могут быть сочинены, т.е. все операторы потенциально заняты и система предпочтений не позволяет сделать выбор. Во 2-ом случае возникает неразрешимая на синтаксическом уровне или же истинная, т.е. существующая и в понимании человека неоднозначность. Рис.17а и 17б (под текстом – условная схема деления линейной структуры на зоны влияния предикатов) иллюстрируют случай истинной неоднозначности выбора Fk, сочиняющего сказуемые, порождающей два разных графа: 1. предложная группа *в субботу* (Рис.17а) оказывается в зоне влияния *навещал* и, соответственно, интерпретируется как его слуга или же 2) оказывается слугою *рассказывал* (Рис.17б).

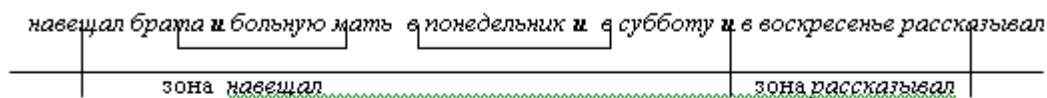


Рис.17а.

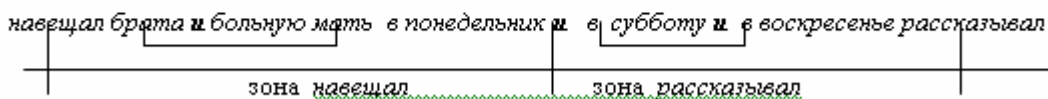


Рис.17б.

Таким образом, с точки зрения иерархии сочинительных связей линейной структуры сегмента, можно выделить два класса связей. Первый – сочинение слов, между которыми не может не быть оператора, их сочиняющего их или разделяющего их сегменты, и второй – сочинение слов, которые могут не быть непосредственно связаны. К первому типу относятся предикативные вершины сегментов. Они не могут быть вложены ни в какие другие сочиненные группы сегмента, тогда как любые другие вкладываются между ними.

Так как линейная сегментная структура предложения рекурсивна [4], и линейная структура синтагматических связей в рамках сегмента тоже рекурсивна [8], определенным в структуре вложений является только порядок этих двух классов сочинительных связей. Во всех остальных случаях иерархию сочинительных связей определяет ситуативная иерархия подчинительных связей.

6. Заключение

Рассмотренные свойства линейной структуры сочиненных групп в значительной степени

1) определяют процедуру анализа знаков препинания и сочинительных союзов в процессе построения сегментов и сочиненных группы и, 2) ограничивая проективные зоны - области поиска потенциально связанных слов, облегчают построение графа подчинительных отношений.

Список литературы

- [1] Апресян Ю.Д., Богуславский И.М., Иомдин Д.Л., Лазурский А.В., Перцов Н.В., Санников В.З., Цинман Л.Л. *Лингвистическое обеспечение системы Этап-2*. //М.: Наука, 1989.
- [2] Иорданская Л.Н. *Автоматический синтаксический анализ. Т.2*. // Новосибирск: Наука, 1967.
- [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. *Принципы сегментационного анализа русского предложения. Московский лингвистический журнал // М., 2004. Т. 8. №1.*
- [5] Кобзарева Т.Ю. *Омонимия и синонимия знаков препинания в русском тексте // Труды Международной конференции Диалог'2005. Москва «Наука» 2005, с. 233-237.*
- [6] Мельчук И.А. *Автоматический синтаксический анализ. Т.1*. // Новосибирск, Ред.-изд. отдел Сибирского отделения АН СССР, 1964.
- [7] Падучева Е.В. *О порядке слов в предложениях с сочинением: сочинительная проективность // НТИ. Сер. 2, 1971, №3.*
- [8] Хомский Н. *Аспекты теории синтаксиса // Изд. Моск. ун-та, 1972.*
- [9] Т.Ю.Кобзарева. *Морфализация in vivo. // Труды Международной конференции Диалог'2004. Москва «Наука» 2004, с 286-291.*
- [11] Ножов И.М. *Процессор синтаксической сегментации русского предложения.//НТИ, сер. 2, 2003, № 11, с. 26-37.*