П 71 24

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

Выпуск 24

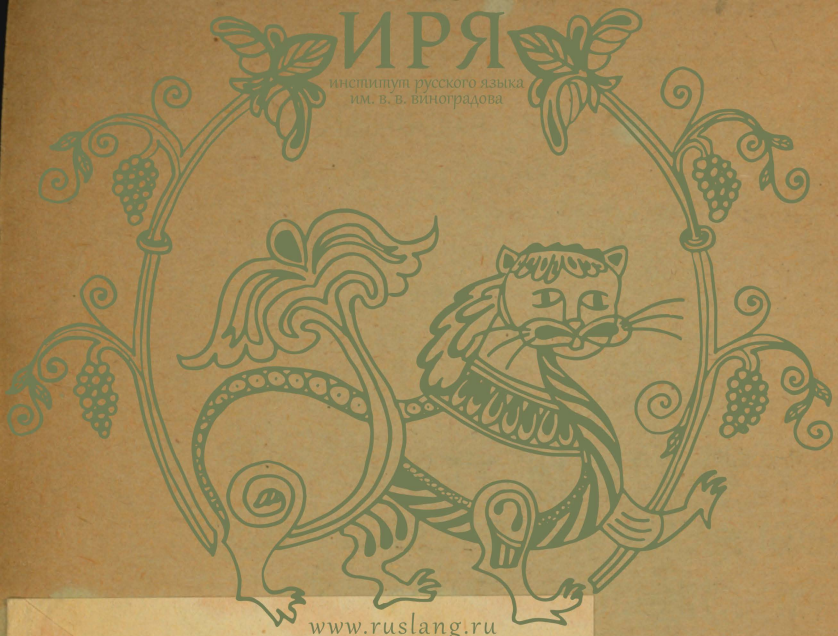
С. И. ГИНДИН

СВЯЗНЫЙ ТЕКСТ: ФОРМАЛЬНОЕ ОПРЕДЕЛЕНИЕ И ЭЛЕМЕНТЫ ТИПОЛОГИИ

Москва — 1971

1. Ол-шарф- шаф. лишнее
2. Ол- синдаклы

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

66157

Институт Русского языка АН СССР
Проблемная группа по экспериментальной
и прикладной лингвистике

Предварительные публикации
Выпуск 24

С.И. Гиндин

СВЯЗНЫЙ ТЕКСТ: ФОРМАЛЬНОЕ ОПРЕДЕЛЕНИЕ И ЭЛЕМЕНТЫ ТИПОЛОГИИ

Москва - 1971

А
П-71
24
БИБЛИОТЕКА
ИНСТИТУТА РУССКОГО ЯЗЫКА
АКАДЕМИИ ЯЗЫКОЗНАНИЯ АН СССР
66157

И, распадаясь на две части,
Уже не связана она.

А.Цыбулевский. Что сторожат ночные сторожа. Тбилиси, 1967, стр. 15.

0. Задачи и структура работы. В предыдущих публикациях[1] мы рассмотрели сущность понятия связности текста в ее противопоставлении жесткому принципу задания текстов с помощью указания допустимых компонентов и их последовательности, а также основные механизмы связности и типы связей между частями текста. В данной работе мы хотим дать формальное определение связного текста в духе и в рамках математической лингвистики, а также наметить контуры некоторой простой типологии связных текстов.

Определение связного текста строится в два этапа: в § 1 описывается используемая модель языка и дается определение того, что считается связью между предложениями или последовательностями предложений. В § 2 выясняется, какое расположение связанных субтекстов относительно друг друга и относительно не связанных с ними субтекстов дает нам связный текст, а какое – нет. В §§ 3-4 дается определение важных подклассов связных текстов – текстов всюду связных и всюду мотивированных, в § 5 рассматривается соотношение текстовой связности со связностью топологической, существенная зависимость связностных характеристик текста от его линейной и уровневой природы. В § 6 определения связности и ее подвидов, определенные для текстов, составленных из предложений, переносится на тексты, сложенные из произвольных совокупностей предложений. Наконец, в §§ 7-8 вводится важное понятие остова текста и рассматриваются возможности типологических классификаций связных текстов по степени устойчивости их связностных характеристик относительно операции изъятия тех или иных частей текста и при переходе от одних разложений текста на части к другим разложениям.

1. Формальное определение связи между субтекстами ("межфразовой связи"). В настоящей работе мы будем полагать возможными

[1] С.И.Гиндин. Онтологическое единство текста и виды внутритекстовой организации. – "Машинный перевод и прикладная лингвистика", вып. 14. М., 1971 (в печати); С.И.Гиндин. Два типа организации текстов и сущность понятия "связность текста". – В кн.: Симпозиум "Семиотические проблемы языков науки, информатики, терминологии". Тезисы докладов. М., МГУ, 1971 (в печ.).

два механизма связи субтекстов: их з а ц е п л е н и е, т.е. введение в состав одного из них специального элемента-связки, значение которого не может быть определено без обращения к другим субтекстам (субституты) или предполагает наличие этих субтекстов (союзы типа и, но, ссылочные слова типа так, следующее и т.п.)[2] и п о в т о р, т.е. уподобление каких-то элементов внутренней структуры субтекстов, выделение в них какого-то одинакового, "рифмующегося" кусочка, будь то звуковой состав, грамматическая категория, совокупность семантических признаков, лексема или целая фраза[3].

Отразить в формальном определении роль связок (зацепляющих элементов) сравнительно нетрудно - достаточно задаться списком этих связок и сказать, что субтексты, в составе которых обнаружен один из членов списка, являются связанными. Сложнее обстоит дело с повторами (роль которых в целом больше роли связок) - ведь мы считаем повторами не только полное совпадение материальной оболочки каких-то элементов текста, которая только и дана нам в тексте, но и совпадение отдельных признаков этих элементов, в том числе таких, которые в тексте, понимаемом как материальный объект, - не зафиксированы. Для того, чтобы сказать, что какие-то элементы являются членами повтора, мы должны уметь как-то отождествить эти элементы. Но подумаем, откуда же берет эти правила человек, воспринимающий текст. Очевидно, они заданы в том ключе к тексту, которым он владеет, - в языке, на котором написан текст. Для построения формального определения связности и связного текста эта мысль является основополагающей - структура тек-

---

[2] Тем самым указание на связь в обоих случаях является существенной компонентой значения слов-связок.

[3] См. классификацию повторов в: С.И.Г и н д и н. Онтологическое единство..., где наряду с зацеплением и повтором выделялся третий возможный механизм - с л е д о в а н и е. По-видимому, следование можно свести к повтору элементов некоторого класса, если ввести в рассмотрение класс элементов, которые можно получить из данного, применяя развернутые правила преобразования. Поэтому в данной работе мы в формальном определении связи отражаем только механизмы зацепления и повтора. Однако при желании нетрудно явно учесть в этом определении и явление следования, для чего в используемое представление языка (см. стр. 8) нужно наряду с классами зацепляющих элементов и разбиениями включить некоторую совокупность правил преобразования.

ста может быть выявлена лишь путем его соотнесения с языком. Впрочем, ничего неожиданного в этом нет – выявление структуры текста есть часть его понимания и так же, как последнее, возможно лишь при условии владения языком и зависит от степени этого владения. И уже выделенный нами класс связок естественно считать одним из важных элементов именно системы языка.

Однако в каком виде включать в язык правила отождествления? Мы думаем, что они с успехом могут быть описаны через классификации, разбиения элементов языка: тождественными (относительно данного разбиения) являются те и только те элементы языка, которые принадлежат к одному классу данного разбиения. То, что таким образом будут исчерпаны все возможные типы отождествления, гарантируется известной теоремой об одно-однозначном соответствии между разбиениями и отношениями эквивалентности[4].

В сущности, уже при описании структуры текста через выявление лексических повторов[5] исследователи пользовались, хотя и неявно, соотнесением текста с языком и некоторым разбиением элементов последнего. Ведь лексема может встречаться в тексте в разных, не вполне тождественных формах и, напротив, текст

День был жаркий. На дне Черного моря много устриц. (1)

нет оснований считать связным. Стало быть, исследование структуры текста по методике И.П.Севбо требует умения отождествлять формы одного слова, т.е. задания некоторого разбиения элементов языка, а именно – простейшего (в известном смысле) разбиения множества словоформ на словоизменительные парадигмы (лексемы). Таким образом, представление языка, лежащее в основе работы И.П.Севбо (Белецкой) может быть формально описано с помощью упорядоченной тройки $\{M, P, F\}$, где M – множество словоформ, P – разбиение M на словоизменительные парадигмы и F – некоторое подмножество свободной полугруппы над M (т.е. предложения, составленные из словоформ)[6].

---

4 См., напр., П.С.Александров. Введение в общую теорию множеств и функций. М.-Л., ОГИЗ ГТТИ, 1948, стр. 23-25; Ю.А. Шиханович. Введение в современную математику. Начальные понятия. М., "Наука", 1965, стр. 263 и след.

5 См. в первую очередь: И.П.Севбо. Структура связного текста и автоматизация реферирования. М., "Наука", 1969.

6 В: С.Марк у с. Теоретико-множественные модели языков. М., "Наука", 1970, стр. 17 – такая тройка называется "языком с парадигматической структурой".

Однако ясно, что разбиение на парадигмы - не единственное разбиение, которое разумно включать в состав адекватной формальной модели естественного языка, и можно надеяться, что введя в эту модель некоторые другие разбиения, мы получим средства для формального определения и других типов повтора.

В математической лингвистике стало уже довольно традиционным рассмотрение разбиений множества словоформ на словоизменительные парадигмы, на пучки грамматических категорий, алфавитного и обратно-алфавитного разбиений[7]. Если первое есть, как мы указали, средство формализации понятия лексического повтора, то второе может служить средством формализации повтора синтаксического, а третье есть пример разбиения, определяющего тождество фонетическое (или графическое, в зависимости от того, в каких символах - фонемных или орфографических - записаны слова из М). Для того, чтобы полностью описать остальные виды фонетического повтора, нужны были бы другие разбиения того же типа - напр., на классы слов, у которых совпадают два последних символа ("рифма глубины 2"), три символа и т.д. Однако, поскольку фонетическая связность в организации текстов, не имеющих "установки на сообщение", занимает периферийное место, желающие могут не включать в понятие языка никаких специальных фонетических разбиений, исключая разве что "поневоле фонетическое" разбиение на парадигмы. Наоборот, интересующиеся, например, поэтическими текстами могут по мере надобности вводить фонетические разбиения нужной глубины.

Сложнее обстоит дело с формализацией повтора семантического. Ведь семантические отношения есть отношения между элементами языка и явлениями внеязыковыми, поэтому возможность их моделирования с помощью каких-то внутриязыковых отношений и операций не представляется очевидной. В свое время И.И.Ревзин в "Моделях языка" отказался от рассмотрения моделей семантики[8], а в наиболее подробном из имеющихся математических определений языка, ста-

---

[7] См., напр., И.И.Р е в з и н. Метод моделирования и структурная типология славянских языков. М., "Наука", 1967, стр. 85.

[8] См.: И.И.Р е в з и н. Некоторые трудности при построении семантических моделей для естественных языков. - В кн.: Симпозиум по структурному изучению знаковых систем. Тезисы докладов. М., Изд-во АН СССР, 1962, стр. 19.

вящим целью как можно более адекватной экспликации интуитивного представления о естественных языках, акад. Л.Кальмар[9] задает семантику с помощью двух внеязыковых множеств A и B и отображений множества слов M α : M $\longrightarrow$ A и множества предложений F β : F $\longrightarrow$ B . Однако, коль скоро такое отображение α задано, на множестве слов M можно ввести следующее отношение Q : $m_1 Q m_2$ тогда и только тогда, когда их значения $\alpha(m_1)$ и $\alpha(m_2)$ близки в некотором строго определенном смысле (напр., если A – метрическое пространство, то можно потребовать, чтобы $\rho[\alpha(m_1), \alpha(m_2)]$ было меньше некоторого фиксированного числа δ).

Отношение Q не будет, вообще говоря, отношением эквивалентности (из-за возможного невыполнения транзитивности: из близости $\alpha(m_1)$ и $\alpha(m_2)$ и близости $\alpha(m_2)$ и $\alpha(m_3)$ не следует близость $\alpha(m_1)$ и $\alpha(m_3)$), но зато оно будет отношением сходства или толерантности (т.е. рефлексивным и симметричным). Следуя Ю.А.Шрейдеру[10], можно показать, что все элементы M можно единственным образом объединить в классы толерантности так, что внутри каждого класса любые два элемента будут находиться в отношении Q (т.е. иметь близкие значения) и для любого элемента m_0 , не входящего в данный класс, в этом классе найдется элемент m_k , не находящийся с ним в отношении Q . Правда, совокупность этих классов не будет разбиением всего множества словоформ, а лишь его покрытием (в смысле Ю.А.Шрейдера), ибо ввиду нетранзитивности отношения Q классы могут пересекаться. Этот факт, однако, не должен нас смущать. Уже членения на парадигмы и на грамматические категории могли считаться разбиениями лишь с некоторыми натяжками – наличие омонимов и омоформ требовало считать инфинитив речь и существительное речь, дне (предложный от день) и дне (предложный от дно) разными словами, иначе классы получились бы пересекающимися. Но там это были отдельные неувязки при общей тенденции к непересечению классов, а в семантике сама сущность предмета требует именно пересекающихся классов, свидетельством чему – существующие семантические словари, являющиеся именно покрытиями,

---

9 L.C a l m a r. Le Langage comme Structure Algébrique. – "Cahiers de Linguistique Théorique et Appliquée", IV. Bucarest, 1967, pp. 79–80.

10 Ю.А.Ш р е й д е р. Пространства толерантности. – "Кибернетика", 1970, № , стр. 125.

а не разбиениями (отношение вхождения пары слов в общую семантическую рубрику симметрично и рефлексивно, но не транзитивно).

Итак, если допускать в формальной модели языка не только разбиения в собственном смысле, но и покрытия, то можно надеяться на формализацию определения всех видов повтора – и фонетического, и синтаксического, и лексического, и семантического. В силу этого точной экспликацией понятия "язык" нам будет служить упорядоченная шестерка $L = \langle M, \{R^i\}, Ц_Л, Ц_П, F, \{G^h\} \rangle$, где

M – словарь (множество словоформ);

$\{R^i\}$ – конечный набор разбиений или покрытий множества $M \setminus (Ц_Л \cup Ц_П)$, причем некоторые из этих разбиений и покрытий могут быть связаны друг с другом отношениями включения или измельчения.

F – некоторое подмножество свободной полугруппы над M (множество предложений);

$\{G^h\}$ – конечный набор разбиений или покрытий множества F[11];

$Ц_Л \subset M$ – класс связок (зацепляющих элементов) с левой областью действия, т.е. сцепляющих отрезок, в который они входят, с предшествующим текстом;

$Ц_П \subset M$ – класс связок с правой областью действия, сцепляющих отрезок, в который они входят, с последующим текстом.

---

[11] Разбиения множества предложений введены нами для того, чтобы отразить трактовку синтаксических повторов не как повтора грамматического значения (категории), выраженной в отдельной словоформе, занимающей определенную позицию, но как повтора элементов, принадлежащих к одному синтаксическому классу – парадигме предложений. См.: Б.М.Гаспаров. Опыт классификации синтаксических структур современного русского языка на основе строения их парадигм. – В кн.: Труды по русской и славянской филологии. XIV. Серия лингвистическая. Тарту, 1970 (Уч. зап. Тартуского Гос. Ун-та, вып. 245). Б.М.Гаспаров. Некоторые лингвистические аспекты изучения структуры текста. – В кн.: Третья летняя школа по вторичным моделирующим системам. Тезисы. Тарту, 1968. По-видимому, обе трактовки синтаксического повтора эквивалентны, и желающие могут, включив в число R^i разбиение П на пучки грамматических категорий, считать набор $\{G^h\}$ пустым. В свою очередь, F вместе с этим набором может быть заменено каким-либо другим заданием синтаксической системы – напр., – списком конфигураций.

Нетрудно заметить, что такая модель языка включает в качестве частных случаев все представления языка, неявно ранее использовавшиеся исследователями структуры текста. Для того, чтобы получить уже разбиравшееся выше представление И.П.Севбо, достаточно взять в качестве единственного R^i разбиение P на парадигмы, а $Ц_л$, $Ц_п$ и $\{G^h\}$ положить пустыми. Представление, используемое И.А.Фигуровским[12], получим, включив в $Ц_л$ и $Ц_п$ союзы и соотносительные слова и оставив пустыми $\{R^i\}$ и $\{G^h\}$. Представление языка, достаточное для подхода, развиваемого Б.М.Гаспаровым, оставляет пустыми $Ц_л$ и $Ц_п$, в качестве единственного G^h берет разбиение на синтаксические парадигмы $G_{парад.}$, а из возможных R^i использует разве что разбиение на пучки грамматических категорий П (и то не для прямого установления связности, а в качестве вспомогательного средства для построения $G_{парад.}$). В дальнейшем мы не будем использовать никакого конкретного уточнения нашей общей схемы понятия "язык", но во всех определениях и рассуждениях предполагается фиксированным некоторый язык L, заданный в соответствии с описанной схемой.

Имея такую общую модель языка L, мы можем теперь построить общее же определение "межфразовой связи", т.е. связи между предложениями и последовательностями предложений. Чтобы не вызывать излишних ассоциаций, мы, как это сейчас принято в работах по математической лингвистике, будем далее называть элементы P не предложениями языка L, а цепочками; не исключено, что при исследовании, напр., стихотворного текста потребуется рассматривать связи между строками, т.е. цепочками, меньшими, чем "предложение" в общепринятом смысле этого термина. Может оказаться полезным при исследовании конкретных текстов и рассмотрение в качестве цепочек - максинтем, т.е. максимальных

---

12 И.А.Фигуровский. Структура текста художественного произведения (тезисы доклада). - В кн.: Вопросы русского языка, диалектологии и методики преподавания. Воронеж, 1965 (Уч. зап. Липецкого пед. ин-та, вып. IV). Сравнение разных подходов см.: С.И.Гиндин. [Рец. на:] И.П.Севбо. Структура связного текста и автоматизация реферирования. - "Вопросы языкознания", 1971, № 5.

отрезков, построение которых предопределяется синтаксической системой языка: максинтемами являются простые предложения, сложноподчиненные предложения и каждая из сочиненных частей сложносочиненных[13].

Итак, пусть нам дан язык L и две цепочки τ_1 и τ_2 (не требуется, чтобы обязательно $\tau_1 \in F$ или $\tau_2 \in F$, т.е. чтобы они были цепочками языка L - содержательно это отражает тот факт, что связная речь может включать и неправильно построенные предложения). Определим на множестве словоформ каждого из предложений семейство отображений $\{r_i\}$ в язык L следующим образом:

$$\begin{cases} r_i(m) = m \equiv m \in M \& i = 0 \\ r_i(m) = R^i_j \equiv m \in R^i_j \& i > 0 \end{cases}$$

где m - некоторая словоформа, а R^i_j означает j-й класс разбиения R^i. Для сцепляющих элементов из классов $Ц_Л$ и $Ц_П$, не входящих в разбиения R, определено только отображение r_0.

Соответственно, определим семейство отображений $\{g_h\}$ и для каждой цепочки в целом[14]:

$$\begin{cases} g_h(\tau) = \tau \equiv \tau \in F \& h = 0 \\ g_h(\tau) = G^h_k \equiv \tau \in G^h_k \& h > 0 \end{cases}$$

где τ - одна из цепочек, а G^h_k - k-й класс разбиения G^h.

O_1 : Множество всех образов всех элементов цепочки τ при некотором фиксированном отображении r_{i_0} называется R^{i_0} - характеристикой данной цепочки $R^{i_0}(\tau)$. Множество образов при отображении r_0 называется М-характеристикой τ. Аналогично R^i - характеристике цепочки определяется ее G^h-характеристика.

O_2 : Множество всех образов всех элементов цепочки τ (в том числе и образов самой τ) при всех отображениях r и g

---

[13] Понятие максимальной синтаксической единицы введено в: И.Б.Долинина. Проблема представления синтаксической структуры в грамматике членов предложения. - В кн.: Проблемы моделирования языка. Вып. 3. 2. Keele modeleerimise probleeme. Tartu, 1969, стр. 29 (Уч. записки Тартуского университета, вып. 228). В отличие от И.Б.Долининой, мы не рассматриваем в качестве максинтем придаточные предложения, которые заведомо включены в более крупные максинтемы.

[14] При другом, напр., конфигурационном способе задания синтаксиса отображения надо было бы определять и на множестве синтаксических составляющих цепочки.

будем называть структурной характеристикой цепочки τ : $S(\tau)$[15]. Структурная характеристика цепочки есть теоретико-множественное объединение всех ее R^i- и G^h-характеристик.

O_2 : Цепочки τ_1 и τ_2 называются $Ц_Л$ - связанными ($Ц_П$ - связанными), если и только если M-характеристика τ_2 содержит хотя бы один элемент из $Ц_Л$: $M(\tau_2) \cap Ц_Л \neq \emptyset$ (соответственно $M(\tau_1) \cap Ц_П \neq \emptyset$)[16].

Например, цепочки

Подошли они к залу.
И Петя открыл дверь. (2)

являются $Ц_Л$ - связанными в силу наличия во втором из них левосторонней связки и. Цепочки же

Петя увидел вот что.
На столе лежали остатки завтрака. (3)

являются $Ц_Л$ - связанными благодаря имеющемуся в первом из них правостороннему связочному элементу вот что. Заметим, что $Ц_Л$ - и $Ц_П$ - связанности не коммутативны - переставив цепочки в примерах (2) или (3), мы получим пары, не являющиеся ни $Ц_Л$, ни $Ц_П$-связанными.

Цепочки τ_1 и τ_2 называются R^i-связанными, если и только если R^i-характеристики τ_1 и τ_2 имеют непустое пересечение: $R^i(\tau_1) \cap R^i(\tau_2) \neq \emptyset$, т.е. существует такой класс разбиения (покрытия) R^i , ему принадлежат одновременно

---

15 Поскольку каждому элементу цепочки может быть сопоставлено много образов, понятие структурной характеристики (даже если ввести в ней отношение порядка) не совпадает с принятым в математической лингвистике понятием B-образа (или B-структуры) цепочки (см., например, И.И.Р е в з и н. Метод моделирования, стр. 85-86). R^i-характеристика предложения с тем же отношением порядка между образами, что между их прообразами в цепочке, будет R^i-образом (R^i-структурой) этой цепочки тогда и только тогда, когда R^i - разбиение, а не покрытие.

16 Тем самым мы ограничиваем круг связок теми, которые подобно и или нижеследующее могут присоединять любые отрезки. В отличие от них связки типа суоституотов предполагают выполнение в этих присоединяемых отрезках некоторых дополнительных условий (напр., наличие антецедентов). Мы не выделяем таких связок, поскольку они могут рассматриваться как разновидность повтора (ср. восстановление антецедентов в работе И.П.Се бо), желающие ввести их в рассмотрение должны заменить в определении языка единственные классы $Ц_Л$ и $Ц_П$ семействами классов $\{Ц_Л^{(x)}\}$ и $\{Ц_П^{(x)}\}$ и соответственно дополнить определение $Ц_Л$- и $Ц_П$- связанности. Другим возможным подразделением связок является деление на контактные и дистантные.

по крайней мере хотя бы один элемент из τ_1 и хотя бы один из τ_2.

Аналогично определяется понятие G^h – связанных цепочек. Цепочки называются **связанными через повтор**, если они R^1– связаны или G^h – связаны.

Приведем примеры, поясняющие смысл введенных определений. Пусть сначала $R^1 = R^0 = M$. Тогда M-связанными будут предложения

Папа ест яичницу.
Папа пойдет на работу., (4)

содержащие общую словоформу папа: $R^0(\tau_1) \cap R^0(\tau_2) = \{папа\}$, – но не будут предложения

Папа ест яичницу.
Папе яичница нравится., (5)

в которых нет ни одной общей словоформы.

Однако предложения из (5) являются P – связанными, где P – разбиение на парадигмы (лексемы), т.к. в каждой из них представлены словоформы лексемы папа и лексемы яичница. Предложения

Продукты из яиц полезны для здоровья.
Папа ест яичницу, (6)

P-связанными уже не будут, так как общих лексем в них нет, но зато они будут W-связанными, где W – разбиение M на словообразовательные гнезда; второе из них содержит яичницу, а первое – яиц. Syn -связанными, где Syn – покрытие M классами синонимичных слов, будут предложения

Отец пришел с работы.
Папа ест яичницу., (7)

потому что папа и отец являются синонимами. Если D^3 – разбиение на классы слов, у которых совпадают три последних фонемы, то D^3-связанными придется считать предложения

Летит большая птица.
Закрылась последняя страница. (8)

Если включить в один класс вместе со словом m_0 все, являющиеся семантическими параметрами (в смысле А.К.Жолковско-

го-И.А.Мельчука) от него (считая, что слово является нейтральным параметром от себя самого), то получим покрытие К параметрическими классами. К-связанными будут предложения

Страна большая.
Народ занялся охотой., (9)

поскольку страна является значением параметра S_{loc} от слова народ.

Пусть теперь G^{constr} есть разбиение множества предложений F на классы предложений, имеющих идентичную синтаксическую конструкцию. G^{constr}-связанными будут предложения

Солнце светит с неба.
Машина едет с горы. (10)

После приведенных определений и разбора примеров мы в состоянии дать теперь общее определение:

O_4: Цепочки τ_1 и τ_2 называются с в я з а н н ы м и (между τ_1 и τ_2 существует м е ж ф р а з о в а я с в я з ь) тогда и только тогда, когда τ_1 и τ_2 или $Ц_л$ - связаны, или $Ц_п$ - связаны, или существует такое $R^i \in L$, что τ_1 и τ_2 R^i - связаны, или существует такое G^h, что τ_1 и τ_2 G^h - связаны. Пересечение структурных характеристик τ_1 и τ_2 или содержащаяся в них связка называются в этом случае а г е н т о м м е ж ф р а з о в о й с в я з и.

Если мы договоримся, что R^i- характеристик (G^h - характеристика, структурная характеристика) конечной последовательности цепочек $t = \tau_1, \ldots, \tau_m$ есть теоретико-множественная сумма R^i-характеристик (G^h- характеристик, структурных характеристик) входящих в нее цепочек, то все выше приведенные определения разных видов связи можно обобщить на случай двух конечных последовательностей цепочек, заменив всюду в определениях слова "цепочки τ_1 и τ_2 " на "последовательность цепочек t_1 и t_2 ".

Наши определения связанности цепочек или последовательностей цепочек относятся к цепочкам и последовательностям, взятым самостоятельно, вне какого-либо определенного текста. Поэтому в любом тексте мы относительно любой пары цепочек можем сказать, связаны они или нет, но ясно, что эта связанность никак не за-

висит от контекста, в который они погружены в данном конкретном случае, и потому, являясь п о т е н ц и а л ь н о й, характеризует не столько текст, сколько систему языка, на котором написаны цепочки. Для характеристики же конкретного текста нужно ввести понятие н е п о с р е д с т в е н н о й связи цепочек в некотором тексте: как в логическом выводе доказываемая теорема "связана" с начальной формулой вывода, сколь бы он ни был длинен, но непосредственно выводится лишь из нескольких формул, используемых на последнем шагу вывода, так и в произвольном тексте естественно предполагать "бинарность" и "последовательность" передачи каждой связи и считать цепочку непосредственно связанной лишь с ближайшими к ней слева и справа из связанных с ней данным агентом цепочек.

O_5 : Цепочки τ_i и τ_j $(1 \leq i < j \leq m)$ называются непосредственно связанными в последовательности $t = \tau_1, \ldots, \tau_i, \ldots, \tau_j, \ldots, \tau_m$, если и только если они связаны и ни одна из промежуточных цепочек τ_{i+k} $(1 \leq k < j - i)$ не связана с ними тем же агентом связи.

Например, в тексте

Блюда из яиц полезны для здоровья.
Мама сделала яичницу.
Папа уплетает яичницу. (11)

первое и третье предложения, вообще-то связанные между собою через словообразовательное гнездо "яйцо" (см. выше пример (6)), не являются непосредственно связанными, поскольку каждое из них тем же агентом связи связано со средним предложением, которое и служит передаточным звеном связности.

Напротив, в тексте

Блюда из яиц полезны для здоровья.
Папа пришел с работы.
Папа вымыл руки и сел уплетать яичницу. (12)

среднее предложение связано с третьим уже не через словообразовательное гнездо "яйцо", а через лексему папа и грамматическое значение прошедшего времени. и потому первое и третье предложения, являясь ближайшими друг к другу предложениями, связанными повтором данного гнезда, в этом тексте непосредственно связаны.

Не менее естественно различение потенциальной связи и непосредственной связи в тексте в случае рифмы и других типов связи[17].

2. Определение связного (по цепочкам) текста. Теперь, когда у нас есть определение межфразовой связи, мы можем приступить к построению определения связного текста. Определение связи между цепочками (последовательностями цепочек) дает нам возможность в некотором упорядоченном наборе цепочек относительно любой конкретной пары цепочек (последовательностей цепочек) решить, являются ли они связанными (как потенциально, так и непосредственно). Но как правило, в наборах будут и связанные пары, и несвязанные, поэтому сам факт наличия в наборе скольких-то связанных пар еще ничего не говорит о связности набора, и для того, чтобы установить, будет ли весь набор в целом связным текстом, нужны какие-то дополнительные критерии. (Напомним, что и в теории графов существуют два понятия – связанность вершин графа и связность графа в целом.) В поисках таких критериев, на наш взгляд, естественнее всего обратиться к взаимному расположению связанных и несвязанных пар. Так, набор

Маша варила кашу.
"Иванов придет".
Каша вышла вкусная. (13)

мы склонны рассматривать как связный текст, а полученный из него перестановкой второго и третьего членов

Маша варила кашу.
Каша вышла вкусная.
"Иванов придет". (14)

[17] Именно отсутствие подобного различения представляется наиболее уязвимым местом подсчетов числа смысловых связей между фразами в текстах различных стилей, выполненных М.И.Бэлзой и Т.Н.Яловой, включавшими в число связей данного текста любую потенциальную связь независимо от расстояния между ее членами. См.: М.И.Б э л з а. Некоторые параметры семантической связности текста. – "Автоматизация информационных работ и вопросы математической лингвистики". Киев, ИК АН УССР, 1968, 2; Т.Н.Р я л о в а. Некоторые формальные критерии выявления семантических связей между предложениями текста. – "Автоматизация информационных работ и вопросы прикладной лингвистики", Киев, ИК АН УССР, 1969, 1.

на наш взгляд, таковым не является (что, однако, не мешает связности

Маша варила кашу. Каша вышла вкусная. "Иванов придет". (15)
Иванов был охоч до вкусненького.

так как связный текст может содержать несвязные куски).

Возможно и другое интуитивное понимание термина "связность текста", не допускающее никаких "перерывов" в связности, т.е. предложений, не связанных с предшествующим отрезком текста (согласно этому пониманию набор (13) пришлось бы считать несвязным). Мы предпочитаем придерживаться первого понимания, второе же будет формализовано в § 3 в определении "всюду связного текста" и, как мы там увидим, всюду связные тексты образуют собственный подкласс связных.

Третье, уже не интуитивное, а достаточно формализованное понимание, согласно которому связный текст – это такой текст, которому соответствует связный граф связей между предложениями[18], кажется нам слишком узким, т.к. не позволяет описать как связные тексты со вставными[19] предложениями или с разного рода "чужими голосами". Поэтому в даваемом ниже определении связного текста мы, обобщая и формализуя определение Н.Н.Леонтьевой[20], данное ею для текстов на предложенном ею семантическом языке, будем исходить из представления о связном тексте как таком наборе предложений, в котором, независимо от того, между какими двумя предложениями мы проведем границу, получившиеся две последовательности будут связанными в определенном выше смысле. Подчеркнем, что мы не случайно определяем связный текст, используя только понятия цепочки и набора цепочек, а не текста, – как мы стремились показать в рабо-

[18] См., например: Э.Ф.Скороходько. Автоматическая компрессия текста на основе анализа его семантической структуры. – В кн.: Проблемы прикладной лингвистики. Тезисы межвузовской конференции 16–19 декабря 1969 г., ч. II, М., 1969 (МГПИИЯ им. М.Тореза).

[19] О вставных предложениях и их функциях см. В.А.Ицкович. К вопросу об отношениях между самостоятельными предложениями в тексте. – В кн.: Вопросы языкознания, кн. I. Львов, 1955, стр. 56–60 (Львов, Гос. Ун-т).

[20] Н.Н.Леонтьева. О смысловой неполноте текста (в связи с семантическим анализом). – В кн.: Машинный перевод и прикладная лингвистика, вып. 12, М., 1969, стр. 106 (МГПИИЯ, Труды института).

те "Онтологическое единство текста...", понятие "текста вообще", в отличие от связного текста, не определимо в собственно лингвистических терминах.

O_6 : Итак, пусть нам дана последовательность цепочек $t = \tau_1, \tau_2, \ldots, \tau_m$, причем не имеет место $\forall_i \forall_j (\tau_i = \tau_j)$, т.е. t содержит хотя бы две не вполне совпадающих цепочки.

Мы будем говорить, что последовательность t **связна в точке** i (i – целое и $1 \leqslant i \leqslant m-1$), если между подпоследовательностями $t^{(i)}_{\wedge} = \tau_1, \ldots, \tau_i$ и $t^{(i)}_{\Pi} = \tau_{i+1}, \ldots, \tau_m$ существует межфразовая связь.

O_7 : Последовательность цепочек t **называется связным текстом**, если каково бы ни было натуральное i , такое, что $1 \leqslant i \leqslant m-1$, t связана в точке i.

Если приложить эти определения к последовательностям в примерах (13) – (15), то получим, что (13), связная и в точке 1, и в точке 2, является связным текстом, (14) несвязная в точке 2, не будет связным текстом, а (15) – опять-таки связный текст, поскольку связна и в точке 1, и в точке 2, и в точке 3. Вообще, можно показать, что имеет место следующая "теорема о вставках":

T_1 : Если даны связанные повтором последовательности предложений $t_1 = \tau^{(1)}_1, \ldots, \tau^{(1)}_k$ и $t_2 = \tau^{(2)}_1, \ldots, \tau^{(2)}_\ell$ $(k \geqslant 1,\ 1 \geqslant 1)$ такие, что последовательность $t' = t_1 t_2 = \tau^{(1)}_1, \ldots, \tau^{(1)}_k, \tau^{(2)}_1, \ldots, \tau^{(2)}_\ell$ является связным текстом, то, какова бы ни была последовательность $t_{вст}$, последовательность $t_{рам} = t_1 t_{вст} t_2$ также будет связным текстом.

Ограничение "связанность повтором" введено в условия T_1 потому, что, если t_1 и t_2 являются $Ц_д$ – или $Ц_п$ – связанными, то со вставлением $t_{вст}$ область действия связочных элементов изменится и либо $t_1 t_{вст}$ и t_2, либо t_1 и $t_{вст} t_2$ могут оказаться несвязанными. Напр., текст

В Лиссабоне землетрясение.
А Петя пойдет в кино. (16)

связный только благодаря зацеплению а, со вставлением в середину предложения

Театр Моссовета покажет завтра премьеру (17)

станет несвязным (связность нарушена в точке 1)

66151
БИБЛИОТЕКА
ИН-ТА РУССКОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР

В Лиссабоне землетрясение.
Театр Моссовета покажет завтра премьеру.
А Петя пойдет в кино. (18)

Теорема о вставках показывает, что данному нами определению связного текста будет удовлетворять, в частности, любая рамочная конструкция ("кольцо" в терминологии О.М.Брика - В.М.Жирмунского, т.е. такая последовательность, в которой первое и последнее предложение связаны между собой каким-нибудь повтором) независимо от того, какова будет срединная часть. Это отражает такую отмечавшуюся нами выше особенность связного текста, как возможность неотмечаемого особыми знаками включения "чужого голоса", цитаты, которая может быть посвящена совсем другому предмету, а то и вовсе быть бессвязной (напр., герой находит пострадавший от времени документ, от которого сохранились лишь обрывочные фразы, включаемые в общий текст повествования). Хорошим примером такого кольца является стихотворение В.Я.Брюсова "О последнем рязанском князе Иване Ивановиче"[21], первый и последний строфоиды которого (почти совпадающие) - обращение к балалаечке-многознаечке с просьбой спеть, а четыре срединных посвящены последнему рязанскому князю и как бы представляют собой песнь балалаечки, но не выделены ни кавычками, ни тире, ни какими-либо другими знаками "чужой речи". Количество этих срединных строф могло бы быть увеличено без ущерба для связности текста (но, конечно, не для его художественной силы).

Отсутствие ограничений на длину вставки может показаться странным. Действительно, у каждого вида повтора есть по-видимому, предельное расстояние между членами, при превышении которого их тождество уже не замечается. Для повтора, например, ритмической структуры строки А.Белый считал таким предельным интервалом десятистишие[22]. Но, во-первых, у каждого вида повтора этот предельный интервал восприятия может быть другим, во-вторых, заметность повтора зависит от окружения его членов в данном кон-

[21] В.Я.Брюсов. Стихотворения и поэмы. Л., "Сов. писатель", 1961, стр. 128-129 ("Б-ка поэта. Большая серия, 2-е изд.").

[22] См.: А.Белый. Ритм как диалектика и "Медный всадник". М., "Федерация", 1929.

крепком тексте[23], в трехлых, величина этого интервала может сильно варьировать от индивидуума к индивидууму – что ощущается как повтор Ивановым, может уже не ощущаться как повтор Сидоровым[24]. Поэтому вводить в определение связного текста какое-нибудь максимальное число предложений между предложениями, связанными повтором, и при превышении этого максимума считать текст несвязным было бы чистейшей воды произволом. Теоретически последовательным будет считать связными все тексты, удовлетворяющие определению O_7, независимо от расстояния между связанными предложениями. Для отражения же влияния длины интервала между членами повтора на связность текста мы должны уже ввести функцию (от натурального аргумента) с т е п е н и с в я з а н н о с т и членов повтора ("заметности" повтора), убывающую с возрастанием этой длины, но никогда, однако, не обращающуюся в нуль, что, конечно, не мешает, в зависимости от тех или иных практических целей, считать ее равной нулю, начиная с некоторого номера[25].

3. Всюду связный текст. Тексты типа рамочной конструкции (кольца) составляют, однако, достаточно специфический подкласс связных текстов. Это своего рода связность a posteriori – нуж-

---

23 Ср. мнение Ю.Н.Тынянова о том, что рифма 10-й строки тютчевского "Кончив пир, мы поздно встали..." с 3-й и 4-й строками ощущается слабо не только из-за 5-строчного интервала, но и из-за того, что это третий рифмующий член, тогда как в начале строфы доминирует двучленная смежная рифма (Ю.Н.Тынянов. Проблема стихотворного языка. Статьи. М., "Сов. писатель", 1965, стр. 59-60). Ср. у Б.Ахмадулиной в "Он утверждал..." ("Лит. Грузия", 1962, № 5) ритму Куры – дарй, почти не ощутимую на фоне установившейся в промежутке между ее членами инерции рифмования через одну (графическую) строку.

24 В этом отношении очень показательным могло бы быть сопоставление максимальных интервалов между рифмующими членами в астрофическом стихе разных авторов.

25 Так фактически поступает в своем "счислении ритма" А.Белый (Цит. соч., измеряющий ритмическое сходство по формуле $\frac{1}{n}$ (а ритмический контраст – по формуле $1 - 1/n$) и при $n > 10$ приравнивающий $1/n$ нулю. См. в этой связи: С.И.Гиндин. Внутренняя семантика ритма и ее математическое моделирование. – В кн.: Проблемы прикладной лингвистики. М., МГПИЯ, 1969; С.И.Гиндин. Пути моделирования ритмической организации текста. – В кн.: Структурно-математические методы моделирования языка. Киев, КГУ, 1970.

но прочесть текст до конца, чтобы ощутить его связность, в процессе же чтения прочитанный отрезок текста может в некоторый момент, напр., в первом предложении вставки, представляться несвязным[26]. Следующее определение вводит класс текстов, в развертывании которых нет этих временных перерывов в связности.

O_8 : Последовательность цепочек $t = \tau_1, \ldots, \tau_m$ называется всюду связным текстом, если и только если каково бы ни было $1 < i \leqslant m$ субпоследовательность $t^{(1)}_{\wedge} = \tau_1, \ldots, \tau_i$ является связным текстом.

Из определения следует, что всюду связный текст всегда является связным. Обратное же автоматически выполняется только для связных текстов длины 2. Более длинные связные тексты уже не обязаны быть всюду связными. Напр., связные тексты (I3) и (I5) не являются всюду связными, ибо в первом несвязен субтекст $t^{(2)}_{\wedge}$, а во втором субтекст $t^{(3)}_{\wedge}$.

Во всюду связном тексте каждое предложение связано хотя бы с одним из предшествующих. Поэтому имеет место:

T_2 : Текст t всюду связен только тогда, когда соответствующий ему граф зависимостей между предложениями N_t связен[27].

Прежде, чем доказывать эту теорему, необходимо уточнить, что значат слова "граф зависимостей, соответствующий тексту t". Сопоставим каждой цепочке τ_i текста t вершину $N(\tau_i)$ и соединим эту вершину ребрами с вершинами, сопоставленными тем и только тем цепочкам, которые непосредственно связаны с τ_i в тексте t. Построенный таким образом граф (обозначим его N^o_t) является минимальным из всех графов, которые естественно считать соответствующими тексту t и отражающими его структуру.

---

26 Ср. различение "последовательного развертывания произведения" и "анализа композиции, завершившей свое развитие" в: Н.В. К о т р е л е в. К вопросу о взаимоотношении метра и ритма. – В кн.: Материалы XXII научной студенческой конференции. Поэтика. История литературы. Лингвистика. Тарту, ТГУ, 1967, стр. 4–5, а также критику представления об одновременном сосуществовании элементов ритма в завершенном целом и утверждение процессуальной природы ритма в: С.И.Г и н д и н. Пути моделирования..., стр. 33.

27 Определение связного графа, а также понятий "ребро", "вершина", "маршрут", "цепь" см. в: О.О р е. Теория графов. М., "Наука", 1968.

все потенциальные межфразовые связи, не являющиеся непосредственными, выразятся в таком минимальном графе уже не непосредственно ребром, а цепью ребер (ср. различение[28]) отношений управления и руководства в грамматиках зависимостей для предложений). Нам представляется, что такой минимальный граф зависимостей лучше отражает структуру текста, чем графы, более насыщенные ребрами. Но так или иначе, включая в качестве самостоятельных ребер еще и те или иные из потенциальных связей, мы, во всяком случае, получим графы с тем же множеством вершин, что в N_t^o и с множествами ребер, включающими множество ребер N_t^o как подмножества. Поэтому наш выбор минимального графа N_t^o оправдан тем, что, доказав связность N_t^o, мы тем самым докажем связность любых других графов, которые имеет смысл считать "соответствующими t".

Доказательство T_2. Возьмем теперь граф N_t^o и допустим, что он несвязен. Тогда в нем имеются по крайней мере две вершины $N(\tau_x)$ и $N(\tau_y)$ такие, что не существует ни одной простой цепи с концами $N(\tau_x)$ и $N(\tau_y)$. Возьмем из вершин $N(\tau_x)$, таких, что существует $N(\tau_y)$, где $y > x$ и не существует цепи от $N(\tau_x)$ до $N(\tau_y)$, вершину с наименьшим x: $N(\tau_{x_{min}})$. Из всех вершин $N(\tau_y)$, с которыми $N(\tau_{x_{min}})$ не соединяет ни одна простая цепь, опять-таки возьмем вершину с наименьшим номером: $N(\tau_{y_{min}})$. По предположению, не существует ни одной соединяющей их простой цепи. Пусть $x_{min} > 1$. В силу связности субтекста $t^{(x_{min})}$ $\tau_{x_{min}}$ связано хоть с одним своим предшественником и среди таких предшественников есть один ближайший к нему $\tau_{предш}$ и, следовательно, $N(\tau_{x_{min}})$ и $N(\tau_{предш})$ можно соединить цепью. В свою очередь, по условию выбора $\tau_{x_{min}}$ $N(\tau_{предш})$ можно соединить цепью с любой вершиной, в том числе и с $N(\tau_{y_{min}})$. Соединив цепи, связывающие $N(\tau_{x_{min}})$ с $N(\tau_{предш})$ и $N(\tau_{предш})$ с $N(\tau_{y_{min}})$, получим цепь от $N(\tau_{x_{min}})$ до $N(\tau_{y_{min}})$, что противоречит принятому допущению.

К такому же противоречию приходим и при $x_{min} = 1$, ибо в силу связности субтекста $t^{(y_{min})}$ $N(\tau_{y_{min}})$ непосредственно

---

28 См., напр.: Ю.А.Шрейдер. Свойства проективности языка. – "Научно-техническая информация", 1964, № 8.

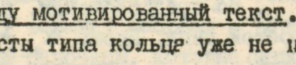
связана ребром с некоторым $N(\tau_{z_0})$, где $z_0 < Y_{min}$, а по условию выбора $\tau_{Y_{min}}$ $N(\tau_{X_{min}})$ можно соединить простой цепью со всеми $N(\tau_u)$, такими, что $X_{min} < u < Y_{min}$, в том числе и при $u = z_0$. Соединив эту цепь и это ребро, получаем цепь от $N(\tau_{X_{min}})$ до $N(\tau_{Y_{min}})$.

Итак, наше допущение неверно и в графе N_t^o не существует пары вершин, которые нельзя было бы соединить цепью. Следовательно, граф N_t^o связен, что и требовалось доказать.

Теорема, обратная T_2, неверна. Именно, текст t, минимальный соответствующий граф которого N_t^o связен, может не быть всюду связным (хотя, конечно, будет связным). Напр., текст:

> От Ростова до Любима в небе звезды и луна.
> "У меня в моей светелке на окошке снег лежит,
> а за шторкою сквозь щелку свечка горит".
> Ни звезды, ни луна, от Ростова до Любима
> в небе свечечка одна[29], (19)

если рассматривать его как состоящий из трех предложений и ограничиваться только лексической связностью, не всюду связен (субтекст $t_\wedge^{(2)} = \tau_1 \tau_2$ не является связным), но сопоставленный ему минимальный граф (лексических) зависимостей будет связен:



4. Всюду мотивированный текст. Под понятие всюду связного текста тексты типа кольца уже не подходят, но вместе с такими текстами, содержащими произвольные инородные вставки, под эту категорию не подходят и тексты, содержащие несвязные субтексты $t_\wedge$, в которых предложения, нарушающие всюду-связность, оказываются не связанными лишь с предыдущим текстом, но одновременно как бы "подготавливают" последующие предложения. Когда длина текста равна 3, тексты такого типа естественно было бы назвать "углами" – сюда относятся, напр., все тексты выводов в исчислении предикатов по правилу введения конъюнкции

$$P(x),\ R(y),\ P(x)\ \&\ R(y) \qquad (20)$$

---

[29] Адаптированный и сокращенный текст "Старинной мелодии" Ушакова, см.: Н.Н.У ш а к о в. Я рифмы не боюсь глагольной. Стихи. М., 1970, стр. 16.

когда предикаты P и R и термы x и y не являются членами какого-нибудь повтора. Примером такого текста на естественном языке может послужить (19), если, повторяем, учитывать в нем только лексические повторы.

Класс подобных текстов, занимающий промежуточное положение между связными и всюду связными текстами, может быть, на наш взгляд, адекватно описан с помощью следующего определения:

O_9 : Текст $t = \tau_1, \tau_2, \ldots, \tau_m$ называется всюду мотивированным, если для любого i $(2 \leqslant i \leqslant m)$, либо субпоследовательность $t_{\wedge}^{(i)}$ является связным текстом, либо существует такое k, $1 \leqslant k \leqslant m - i$, что цепочки τ_1 и τ_{i+k} связаны.

Из определения следует, что связный текст длины 2 всегда всюду мотивирован.

Класс всюду мотивированных текстов включает в себя класс текстов, которым соответствуют связные графы. Именно, выполняется

T_3 : Текст t, минимальный соответствующий граф которого N_t^o связен, является всюду мотивированным.

Доказательство. Пусть граф N_t^o связен. Возьмем вершину $N(\tau_i)$. В силу связности графа существует простая цепь соединяющая ее с $N(\tau_1)$. Возможны два случая: 1) эта простая цепь проходит хотя бы через одну вершину $N(\tau_{i+k})$, где $k > 1$ и 2) эта цепь проходит только через вершины $N(\tau_{i-k})$. В первом случае существует такое минимальное k, что τ_1 и τ_{i+k} связаны, во втором – $t_{\wedge}^{(i)}$ является связным субтекстом. Так как i было выбрано произвольно, то доказанное верно для любого i, $2 \leqslant i \leqslant m$ и, следовательно, t – всюду мотивированный текст.

Обратная теорема неверна, ибо не всякому всюду мотивированному тексту соответствует связный минимальный граф. Противоречащим примером может служить текст, получаемый перестановкой предложений текста (15)

Маша варила кашу.
Иванов придет.
Каша вышла вкусная.
Иванов охоч до вкусненького. (21)

которому соответствует граф



По этому примеру можно предположить, что несвязный граф будут иметь те из всюду мотивированных текстов, которые представляют собой нечто вроде вложенных одна в другую зубьями гребенок, т.е. переплетение двух или более всюду связных текстов. К текстам такого типа мы еще вернемся ниже в связи с вопросами типологии связных текстов по поведению их связностных характеристик при переходе от одного разложения к другому.

5. Связность текста по повторам и топологическая связность. Зависимость связностных характеристик текста от его линейной и и уровневой природы. Средства связи, учтенные в нашем определении связного текста, имеют двоякую природу. С одной стороны, это зацепляющие элементы, связки, аналогичные связкам внутри предложений (характерно, что в языках математической логики все связки, даже если они вынесены вперед, как в нотации Лукасевича, действуют внутри формулы, а не между формулами, и даже при кванторах дескрипции, специально созданных для ссылки на другие формулы, то, к чему они отсылают, повторяется внутри данной формулы). С другой стороны, это повторы - само определение повторов как пересечений структурных характеристик наталкивает на топологические аналогии, а определение связного текста и вытекающее из него понимание несвязной последовательности как такой, которую можно разбить на две подпоследовательности с непересекающимися структурными характеристиками, прямо напоминают топологические определения связного и несвязного пространства.

Пусть нам дан текст $t = \tau_1, \dots, \tau_m$. Введем в структурную характеристику $S(t)$ следующую топологию: объявим открытыми множествами структурные характеристики всех левых и всех правых членов любого бинарного разложения, т.е. $S(t^{(i)}_{\wedge})$ и $S(t^{(i)}_{\sqcap})$ для любого i, а также любые их пересечения и суммы и пустое множество. Из самого определения вытекает, что суммы и пересечения открытых множеств будут открытыми. Всё $S(t)$ будет открытым как $S(t^{(m)}_{\wedge})$. Тем самым аксиомы топологического простран-

ства[30] выполнены.

Ясно, однако, что из связности текста t связность S(t) как топологического пространства еще не следует. Для того, чтобы топологическое пространство было связным, требуется[31], чтобы его нельзя было представить в виде суммы двух его непересекающихся непустых открытых подмножеств. Текст же, как мы помним, может быть связным благодаря концевому расположению связанных частей, даже если "сердцевина" никак не связана с этими концевыми частями. В этом случае S(t) можно представить как сумму структурной характеристики "сердцевины" с объединением структурных характеристик концевых частей. Оба слагаемых – открытые множества, они не пересекаются и потому S(t) – несвязное топологическое пространство.

Даже структурная характеристика всюду связного текста может оказаться несвязным топологическим пространством. Рассмотрим, например, текст

Миша сел.
Саша сел.
Миша встал.
Саша встал., (22)

всюду связный относительно языка L, включающего только единичное разбиение на словоформы. Тогда структурной характеристикой предложения будет множество его словоформ. Открытыми подмножествами структурной характеристики этого текста будут

Λ, {Миша}, {Саша}, {сел}, {встал},
{Миша, сел}, {Саша, сел}, {Миша, встал}, {Саша, встал},
{Миша, Саша, встал}, {Миша, Саша, сел},
{Миша, встал, сел}, {Саша, встал, сел},
{Миша, Саша, встал, сел}.

Очевидно, что S(t) здесь легко представимо в качестве суммы двух непустых непересекающихся открытых множеств, напр.,

{Миша} ∪ {Саша, встал, сел}

и потому несвязно. Причина этого в том, что в тексте (22) все

---

30 П.С.Александров. Введение в общую теорию множеств и функций. М.-Л., ОГИЗ ГТТИ, 1948, стр. 288.

31 П.С.Александров. Цит. соч., стр. 294, 256.

слова совершают "полный оборот" (нет только предложения "Встал сел"), т.е. для любого слова можно подобрать два предложения, которые только им и отличаются. Поэтому оказывается, что если какая-то часть структурной характеристики предложения – открытое множество, то и оставшаяся часть структурной характеристики этого предложения также будет открытым множеством. И потому все $S(t)$ оказывается возможным представить в виде суммы открытых подмножеств, ни одно из которых не является ни структурной характеристикой предложения или последовательности предложений, ни суммой структурных характеристик. Между тем, определение связного (и всюду связного) текста учитывает особый характер описываемого объекта – линейное расположение (линейный порядок) его элементов (каждый из которых есть некоторая выборка из $S(t)$), и требование пересечения открытых подмножеств применяется в нем не ко всем парам непустых открытых подмножеств, сумма которых равна $S(t)$, а лишь к таким, которые, кроме того, соответствуют элементам, следующим друг за другом в тексте, т.е. к реально сделанным последовательным выборкам из $S(t)$. В этом смысле определение даже всюду связного текста слабее определения связного топологического пространства, именно, верна следующая

T_4: Структурная характеристика $S(t)$ всюду связного текста не может быть представлена в виде $S(t) = S' \cup S''$, где S' и S'' – непустые непересекающиеся открытые множества, каждое из которых является либо структурной характеристикой некоторой последовательности цепочек $\tau_i, \tau_{i+1}, \ldots, \tau_{i+k-1}, \tau_{i+k}$ из текста t (где $1 \leqslant i \leqslant m$, $0 \leqslant k \leqslant m - i$), либо суммой структурных характеристик таких последовательностей.

Доказательство будем вести от противного. Допустим, что наше $S(t)$ с описанной типологией, где t – всюду связный текст, несвязно как топологическое пространство. Это значит, что существуют такие открытые S' и S'', что $S(t) = S' \cup S''$, $S' \cap S'' = \emptyset$ и при этом S' и S'' не любые открытые множества, а являются каждое либо структурной характеристикой некоторой последовательности предложений, либо объединением таких структурных характеристик. Тогда возможны два случая: либо а) S' и S'' являются структурными характеристиками соответственно левого

и правого членов некоторого бинарного разложения текста t, т.е. $S(t) = S(t_{\wedge}^{(i)}) \cup S(t_{\sqcap}^{(i)})$ для некоторого i, либо

β) S' является суммой структурных характеристик нескольких отделенных друг от друга (и, возможно, одноэлементных) последовательностей предложений, а S'' – сумма структурных характеристик всех остальных последовательностей: $S' = S(\tau_{i_1}, \tau_{i_1+1}, \ldots, \tau_{i_1+k_1}) \cup S(\tau_{i_2}, \tau_{i_2+1}, \ldots, \tau_{i_2+k_2}) \cup \ldots \cup S(\tau_{i_r}, \tau_{i_r+1}, \ldots, \tau_{i_r+k_r})$, а $S'' = S(\tau_1, \tau_2, \ldots, \tau_{i_1-1}) \cup S(\tau_{i_1+k_1+1}, \ldots, \tau_{i_2-1}) \cup \ldots \cup S(\tau_{i_{r-1}+k_{r-1}+1}, \ldots, \tau_{i_r-1}) \cup S(\tau_{i_r+k_r+1}, \tau_{i_r+k_r+2}, \ldots, \tau_m)$, где $1 < i_1 < i_2 < \ldots < i_{r-1} < i_r \leqslant m$; $0 \leqslant k_p < i_{p+1} - i_p$ $(p = 1, 2, \ldots, r)$; $1 \leqslant r \leqslant m - 2$.

Рассмотрим эти случаи по отдельности и покажем, что они оба невозможны.

α) Всюду связный текст заведомо является связным, т.е. при любом бинарном разложении существует такое R^j или G^h, что правый член $t_{\sqcap}^{(i)}$ R^j – связан или G^h – связан с левым $t_{\wedge}^{(i)}$. Это означает, что существует класс R_x^j, принадлежащий разбиению R^j (или класс G_y^h, принадлежащий G^h) и являющийся R^j – образом (G^h – образом) хотя бы одного элемента из $t_{\wedge}^{(i)}$ и хотя бы одного из $t_{\sqcap}^{(i)}$. Но тогда этот R_x^j (G_y^h) должен входить и в $S(t_{\wedge}^{(i)})$ и в $S(t_{\sqcap}^{(i)})$ и, следовательно, S' и S'' не могут быть непересекающимися. Случай α), таким образом, невозможен.

β) Существующие по нашему допущению S' и S'' являются суммами структурных характеристик нескольких подпоследовательностей цепочек. Возьмем самую левую подпоследовательность из тех, структурная характеристика которых входит в S'', именно – подпоследовательность $\tau_1, \ldots, \tau_{i_1-1}$. Возьмем также τ_{i_1} – первую цепочку самой левой подпоследовательности из тех, структурная характеристика которых входит в S'. По определению всюду связного текста, левый член $t_{\wedge}^{(j)}$ любого бинарного разложения связен. Следовательно, связна и подпоследовательность $\tau_1, \ldots, \tau_{i_1-1}, \tau_{i_1}$, являющаяся левым членом i_1-го бинарного разложения текста t. В частности, эта подпоследовательность связна в точке $i_1 - 1$ и тем самым структурные характеристики $S(\tau_1, \ldots, \tau_{i_1-1})$ и $S(\tau_{i_1})$ имеют непу-

тое пересечение. Соответственно, непустое пересечение будут иметь и S'' с S' и наше допущение неверно, случай β) также невозможен.

Как видим, наше предположение, что $S(t)$ можно представить в виде суммы S' и S'', где $S' \cap S'' = \emptyset$ и S', S'' – открытые множества, приводит к противоречию и потому неверно. Топологическое пространство $S(t)$ нельзя представить в указанном виде, Ч.Т.Д.

Заметим, что в доказательстве теоремы 4 мы существенно использовали всюду-связность текста t, ибо, если случай α) невозможен для любого связного текста, то случай β), в котором не требуется линейного предшествования элементов S' элементам S'', может иметь место в любом тексте, который связен, или даже всюду мотивирован, но не всюду связен, может иметь место. Именно в таком тексте $W = \omega_1, \ldots, \omega_\ell$ существует некоторый несвязный субтекст $\omega_1, \ldots, \omega_i$. Если субтекст ω_i, ω_ℓ несвязен, то возьмем все такие ω_{i+k} $(0 \leqslant k < \ell - i)$, что субтексты ω_i и ω_{i+k} связаны, а субтекст $\omega_1, \ldots, \omega_{i-1}, \omega_{i+k}$ несвязен (множество таких ω_{i+k} заведомо непусто, ибо в него войдет по крайней мере само ω_i). Структурную характеристику последовательности таких ω_{i+k} обозначим через S_1, а структурную характеристику последовательности остальных предложений через S_2. Тогда $S(W) = S_1 \cup S_2$, где S_1, S_2 непустые открытые множества вида, описанного в условиях T_4, и $S_1 \cap S_2 = \emptyset$, т.е. структурная характеристика связного текста оказывается несвязным топологическим пространством.

Доказанная теорема показывает, в силу каких факторов структурная характеристика всюду связного текста может оказаться несвязным топологическим пространством. Именно, для этого необходимо и достаточно, чтобы нашлось такое открытое подмножество, меньшее (вопреки уровневой членимости текста на предложения, а не на их составные части), чем структурная характеристика одного предложения, дополнение к которому, включающее "остаток" структурной характеристики данного предложения, также было бы открытым. Таков был рассмотренный нами выше текст (22), включавший почти все возможные перестановки словоформ. В то же время из теоремы вытекает, что, если каждое предложение всюду связ-

ного текста содержит в своей структурной характеристике хотя бы один элемент (напр., словоформу), не входящий в структурные характеристики других предложений, то структурная характеристика такого текста будет связным топологическим пространством.

Заканчивая рассмотрение структурных характеристик текстов как топологических пространств, заметим, что они не обязательно являются даже T_0 – пространствами[32], – ведь, чтобы не выполнялась нулевая аксиома отделимости, достаточно, чтобы некоторые два элемента, напр., два семантических поля, встретились в тексте лишь однажды и при этом – в одном предложении, которое является пересечением всех окрестностей, содержащих каждый из этих элементов. Однако тезис, что "топологические пространства, не являющиеся T_0-пространствами, едва ли представляют какой-нибудь интерес"[33], обусловлен нуждами теорий точечных множеств и метрических пространств, и возможно, что в топологическом изучении текстов его придется пересмотреть. А то, что такое изучение имеет смысл – видно уже из того, что даже условия, при которых не выполняется нулевая аксиома, непосредственно связаны с распределением в тексте лексем, семантических полей и других языковых сущностей.

6. Связность и иерархическая организация текста. Текст, связный относительно произвольно фиксированного разложения. Связность текста по цепочкам (предложениям) не исчерпывает его внутренней организации по гибкому принципу, так как не отражает возможности возникновения в тексте всякого рода частей, "сгустков", объединений предложений, связанных друг с другом теснее, чем с остальными предложениями и потому противостоящих этим остальным, как некое единое цельное образование. Особенно явным такое выделение частей становится тогда, когда оно зафиксировано самим автором в иерархическом членении на главы, тома и т.п. Между такими сгустками так или иначе объединенных (напр., определенным видом повтора) предложений, в свою очередь, могут возникать отношения

32 См.: П.С.Александров. Цит. соч., стр. 300.
33 Там же.

связностного типа. Во-первых, связочные элементы могут сцеплять не только два предложения, но и две последовательности предложений, заканчиваемые первым и начинаемые вторым из непосредственно связанных предложений (не говоря уже о связках типа *вышеизложенное*, изначально призванных сращивать именно целые последовательности). Напр., в стихотворении Некрасова "Смолкли честные, доблестно павшие..." Но в четвертой строке, непосредственно присоединяющее лишь эту строку к трем предшествующим, отзывается затем средством сцепления первого трехстишия, описывающего завершившееся действие, со всем последующим десятистишием, посвященным настоящему моменту[34]. Во-вторых, ввиду того, что большинство предложений по крайней мере двучленно, то даже при ограничении одним типом связей существует возможность иерархизации структуры: связав ближайшие предложения лексическим повтором, напр., сказуемого или дополнения, мы можем получившиеся группы связать затем повтором первого подлежащего. Например:

Пикассо стал заниматься скульптурой.
Многие художники сразу же перешли на скульптуру.
Пикассо принялся за графику.
И опять множество народу увлеклось графикой. (23)

В-третьих, существование разных типов связей, в том числе – повтора, делает возможным несовпадение структур текста, выявленных с учетом каждый раз только одного из этих типов[35]. В частности –

---

34 См.: С.И.Гиндин. Опыт анализа структуры текста с помощью семантического словаря (стихотворение Некрасова "Смолкли честные, доблестно павшие...") (в печати).

35 Ср. о взаимоотношениях и возможном несовпадении "точек зрения" разного типа в: Б.А.Успенский. Поэтика композиции. Структура художественного текста и типология композиционных форм. М., "Искусство", 1970, гл. 5. Нам представляются, однако, крайне неудачными употребляемые в этой главе выражения "точки зрения, выделяемые на разных уровнях анализа" и тем более "точки зрения на разных уровнях в произведении" – ведь поскольку между "оценкой", "фразеологией", "пространственно-временной характеристикой" и "психологией" трудно установить какие-нибудь иерархически соподчиняющие отношения типа тех, что конституируют уровни языка, термин "уровень" лишается здесь всякого конкретного содержания и лишь дезориентирует читателя. Термин "план", фигурировавший в названиях других глав, больше отвечает существу дела.

возможно применение связей типа X к группам предложений, уже скрепленным связями типа Y. В тютчевском "Восток белел. Ладья катилась..." объединение внутри строф достигается синтаксическим повтором глагольного времени (в первой и второй строфах – еще и глагольного вида), а также семантическими повторами чело – уста – взоры, склонилась – поникла и выя – ланиты. Вторая и третья строфы объединяются семантическими повторами алель – вспылать, чело – уста – взоры – выя – ланиты и лексическим она – она. Наконец, все три строфы объединены лексическим повтором первого подлежащего Восток, сквозной рифмой первых строк и семантическим повтором элементов поля "цвета" белел – алел – вспылал.

Таким образом, организация текста в каждой данной точке предстает как диалектическое единство связанности по одним параметрам (агентам связи) и несвязанности по другим. В определении же связности по цепочкам момент несвязанности учитывался разве что в требовании, чтобы не все цепочки были тождественны. Момент же иерархичности организации текста вовсе не был отражен в нем.

Чтобы отразить все это и формально охарактеризовать подобные формальные связностные отношения между субтекстами, введем следующие определения.

O_{10} : Разложением $t \langle j_1, j_2, \ldots, j_l \rangle$ последовательности цепочек $t = \tau_1, \tau_2, \ldots, \tau_m$ будем называть такую последовательность ее подпоследовательностей $\theta_1, \theta_2, \ldots, \theta_l$, что $\theta_1 = \tau_1, \ldots, \tau_{j_1}$; $\theta_2 = \tau_{j_1+1}, \tau_{j_1+2}, \ldots, \tau_{j_2}$, ... ; $\theta_l = \tau_{j_{l-1}+1}, \tau_{j_{l-1}+2}, \ldots, \tau_m$, где $1 \leqslant j_1 < j_2 < \ldots < j_{l-1} < j_l = m$.

O_{11} : Последовательность $t = \tau_1, \tau_2, \ldots, \tau_m$ называется текстом связным относительно некоторого разложения $t \langle j_1, \ldots, j_l \rangle$ если и только если она связна в каждой из точек $j_1, j_2, \ldots, j_{l-1}$

Поскольку точки $j_1, j_2, \ldots, j_{l-1}$ составляют собственное подмножество точек 1, 2, ..., m, текст, связный по предложениям, будет связен относительно любого своего разложения на субтексты. Обратное неверно. Возможность построения противоре-

чащих примеров дает нам

T_5 : Если $t = \tau_1, \ldots, \tau_m$ – связный по цепочкам текст, то какова бы ни была последовательность цепочек $t' = \tau'_1, \tau'_2, \ldots, \tau'_k$, последовательность $t_{рез} = tt' = \tau_1 \tau_2, \ldots, \tau_m, \tau'_1, \tau'_2, \ldots, \tau'_k$ будет связным текстом относительно бинарного разложения $t_{рез} \langle m-1, m+k \rangle$.

Согласно этой теореме мы можем рассматривать как связный текст типа

Отец дал мне тетрадь.
"Прочти", – сказал он.
Я раскрыл тетрадь.
"Эне. Бене. Рес. Квинтер. Минтер Жес" (24)

– даже, если само "Эне Бене Рес..." кажется нам бессвязным, оно, несомненно, связано с предыдущей фразой, что и отражается в факте его объединения с этой фразой в один член бинарного (или тернарного) разложения, относительно которого текст (24) в целом является связным.

Как и в случае связности по цепочкам, для класса текстов, связных относительно некоторого l-арного разложения, естественно ввести его сужения – класс текстов, всюду связных относительно данного разложения, и класс текстов, всюду мотивированных относительно данного разложения.

O_{12} : Последовательность цепочек $t = \tau_1, \ldots, \tau_m$ называется текстом, всюду связным относительно l-арного разложения $t \langle j_1, j_2, \ldots, j_l \rangle$ $(1 \leqslant m)$, если и только если

каково бы ни было k, где $1 < k \leqslant l$, субпоследовательность $t^{(jk)} = \tau_1, \ldots, \tau_{j_{k-1}}, \tau_{j_{k-1}+1}, \ldots, \tau_{j_k}$ связна в точке j_{k-1}.

Смысл этого определения в том, что каждый следующий субтекст данного разложения должен быть сам связан хоть с одним из предыдущих субтекстов. Текст, всюду связный по цепочкам, всюду связен относительно любого своего разложения, а обратное неверно.

T_6 : Связный по цепочкам, но не всюду связный (по цепочкам же) текст t всюду связен относительно любого бинарного разложения (по определению связности по предложениям), но для не-

го обязательно существует такое l-арное разложение, где $l > 2$, относительно которого t не всюду связен.

Доказательство. В силу не всюду-связности t существует такое i_1, что $t^{(i_1)}_{\wedge}$ - минимальный из несвязных левых субтекстов. Но в силу связности существует такое $i_2 > i_1$, что $t^{(i_2)}_{\wedge}$ связен. Из всех таких i_2 опять-таки возьмем минимальное. Уже относительно тернарного разложения $t\langle i_1 - 1, i_2 - 1, m\rangle$ текст не всюду связен. Более того, соответствующим образом разбивая подпоследовательности $t^{(i_1-1)}_{\wedge}$ и $t^{(i_2-1)}_{\sqcap}$ и оставляя в неприкосновенности последовательность $\tau_{i_1}, \ldots, \tau_{i_2-1}$, мы для любого l, где $3 \leqslant l \leqslant m - (i_2 - 1) + (i_1 - 1) + 1$ можем получить l-арное разложение t, относительно которого t не будет всюду связным.

O_{13}: Последовательность цепочек $t = \tau_1, \ldots, \tau_m$ называется текстом всюду мотивированным относительно разложения $t\langle j_1, j_2, \ldots, j_l\rangle$ если и только если, каково бы ни было $1 < k \leqslant l$, либо $t^{(j_k)}_{\wedge} = \tau_1, \tau_2, \ldots, \tau_{j_{k-1}}, \tau_{j_{k-1}+1}, \ldots, \tau_{j_k}$ связен в точке j_{k-1}, либо существует такое d, $k < d \leqslant l$, что подпоследовательность $\tau_{j_{k-1}+1}, \ldots, \tau_{j_k}, \tau_{j_k+1}, \ldots, \tau_{j_d}$, связана в точке j_k.

Короче говоря, текст всюду мотивирован относительно данного разложения на субтексты, если каждый субтекст непосредственно связан хотя бы с одним из предшествующих или последующих субтекстов.

Текст, всюду мотивированный по цепочкам, всюду мотивирован относительно любого своего разложения. Текст, всюду связный по цепочкам, всюду связен, а значит, и всюду мотивирован относительно произвольного разложения. Для текста, связного по цепочкам, но не всюду мотивированного по цепочкам, существует хотя бы одно небинарное разложение, относительно которого он не всюду мотивирован. (Алгоритм построения такого разложения аналогичен примененному выше в доказательстве теоремы 6).

Между свойствами связности, связности минимального соответствующего графа, всюду-связности и всюду-мотивированности отно-

сительно некоторого фиксированного разложения сохраняется то же соотношение, что и между связностью, всюду-связностью и всюду-мотивированностью по цепочкам:

Связность ⟵ Всюду-мотивированность ⟵ Связность минимального соответствующего графа ⟵ Всюду-связность

– из более правого следует более левое, но не наоборот.

7. Типология текстов, связных относительно фиксированного разложения, на основе комбинаторики их субтекстов. Пусть нам дано некоторое фиксированное разложение текста t. Будем обозначать субтексты – члены этого разложения через $\theta_1, \ldots, \theta_1$. Поставим вопрос: как влияет на связностные характеристики элиминация некоторых членов этого разложения или их перестановка? Выше, в разделе 2 мы различали потенциальную связанность цепочек вне контекста и их непосредственную связанность в конкретном тексте. Возьмем некоторый субтекст, являющийся членом данного разложения. Допустим, что мы вынули из текста тот субтекст, с которым он был непосредственно связан. Тогда возможны два случая: либо среди более удаленных субтекстов найдется потенциально связанный с нашим вне контекста и тогда их связанность актуализируется в новом, "сокращенном" тексте и станет в нем непосредственной связью, либо такого элемента не найдется. И в том и в другом случае связностные характеристики сокращенного текста могут отличаться от характеристик исходного "полного" текста – в частности, связный текст может, вообще говоря, превратиться в несвязную последовательность цепочек. Реакция связностных характеристик текста (в фиксированном разложении) на изъятие тех или иных субтекстов показывает степень "прочности" организации этого текста, его насыщенность потенциальными запасными связями, способными актуализироваться и заменять изъятые связи. Поэтому соотношение связностной характеристики "полного" текста с характеристиками его различных сокращений может быть положено в основу некоторой достаточно содержательной классификации связных текстов, начатки которой мы здесь рассмотрим.

Предварительно введем важное понятие о с т о в а текста. Так как количество даже непосредственных связей между субтекстами некоторого текста может существенно превышать само число суб-

текстов (т.е. длину текста), представляет интерес вопрос о "минимизации", сокращении числа связей, не оставляющей, однако, без связей ни один из субтекстов.

O_{14}: Остовом данного разложения текста $t = \theta_1, \ldots, \theta_l$ называется такая система субтекстов – членов этого разложения, что каждый субтекст данного разложения потенциально связан хотя бы с одним субтекстом данного разложения. Вообще "остовом текста" естественно называть его остов относительно $(m-1)$-арного разложения на предложения, но поскольку ниже речь всюду будет идти о фиксированном разложении текста, мы, говоря об остове, будем опускать слова "данного разложения" и говорить просто "остов текста" – ведь если разложение фиксировано и других разложений нет, то рассматриваемые субтексты данного разложения есть единственные составляющие текста, своего рода "укрупненные предложения".

Для ряда задач может быть полезно также понятие внешнего остова текста – такой совокупности не входящих в текст цепочек, что каждый субтекст текста (потенциально) связан хотя бы с одной из цепочек этой совокупности. Если внутренний остов текста может рассматриваться как его реферат, полученный путем выбора отрезков из самого текста, то внешний – как "осмысляющий" реферат, при построении которого все "пересказывалось своими словами".

Примером остова (в общем случае – внешнего) текста, рассматриваемого с точки зрения фонетической организации, может служить так называемый "звукообраз" или "звуковая тема" произведения[36].

---

[36] См.: В.И.Иванов. О "Цыганах" Пушкина. В кн.: В.И.Иванов. По звездам. Спб., "Оры", 1909, стр. 146–150; В.И.Иванов. К проблеме звукообраза у Пушкина. – В кн.: Московский Пушкинист, II., М., "Федерация", 1930; Д.Выгодский. Из эвфонических наблюдений ("Бахчисарайский фонтан"). – В кн.: Пушкинский сборник памяти С.А.Венгерова. Пушкинист, IV. М.–Пг., Гиз, 1922; F. de Saussure. Les anagrammes. – "Mercure de France", 1964, Fevrier; Л.М.Аринштейн. Фонологические средства экспрессии в современном русском стихе. – В кн.: Сборник докладов и сообщений Лингвистического общества. I, вып. 2. Калинин, КГПИ, 1969, стр. 11–13 (в последней работе удивляет утверждение о фонологической, а не фонетической природе данного явления и его отнесение исключительно к современному стиху).

Ясно, что у каждого текста есть по крайней мере один внутренний остов – он сам. Интерес представляет отыскание других, нетривиальных остовов.

Понятие остова полезно для типологии связных текстов по изменению их связностных характеристик при изъятии субтекстов – ведь сам остов также есть результат изъятия каких-то субтекстов. Вместе с тем, поскольку в определении остова не предполагается существование связей между его элементами, это понятие может быть применено и к несвязным последовательностям.

Перейдем теперь к обещанной типологии. Мерой прочности организации текста может служить наименьшее число субтекстов, которое нужно изъять, чтобы текст стал несвязным.

Минимальной прочностью обладает связностная организация того текста, в которой достаточно убрать один субтекст, чтобы обратить его в несвязный. Дальнейшая классификация таких текстов может опираться на положение этого "ключевого" субтекста. Если текст обращается в несвязный лишь при изъятии одного из концевых субтекстов, то перед нами "кольцо" с произвольной "сердцевиной" (причем, если текст становится несвязным и при изъятии обоих концевых субтекстов сразу, то "сердцевина" кольца является сама по себе несвязным текстом).

O_{15} : Особенно интересен такой связный текст, который перестает быть связным при изъятии из него любого, в том числе и неконечного субтекста (а тем более любой пары, любой тройки и т.д. неконечных субтекстов). Назовем такой текст ц е п ь ю б е з в о з в р а щ е н и й. Про него могут быть доказаны следующие теоремы:

T_7 : Цепь без возвращений не содержит связей по зацеплению.

Следует из определений $Ц_л$- и $Ц_п$-связанности предложений. Если цепочка τ' содержит левостороннюю связку, то $\tau''\tau'$ связны, каково бы ни было τ''. Поэтому, если бы хоть один субтекст цепи без возвращений (кроме начального) содержал левостороннюю связку, то изъятие предыдущего субтекста не превращало бы ее в несвязный текст. Аналогично доказывается отсутствие правосторонних связок во всех субтекстах, кроме последнего.

T_8 : Цепь без возвращений является всюду мотивированным текстом.

Доказательство. Возьмем произвольный субтекст θ_i. По определению цепи без возвращений ее связность с изъятием θ_i нарушается по крайней мере в одной точке k. Возможны два случая: 1) $k < i$, 2) $i < k$. Рассмотрим оба случая, причем возьмем точки нарушения связности, ближайшие к точке i.

1) Нарушение связности показывает, что один из субтекстов $\theta_1, \ldots, \theta_k$ непосредственно связан с субтекстом θ_i, а значит субтекст $t^{(i)}_{\wedge} = \theta_1, \ldots, \theta_k, \ldots, \theta_{i-1}, \theta_i$ связан в точке $i-1$.

2) Нарушение связности показывает, что один из субтекстов $\theta_{k+1}, \ldots, \theta_m$ непосредственно связан с θ_i.

Мы получили, что для любого i либо субтекст $t^{(i)}_{\wedge}$ связен в точке $i-1$, либо субтекст θ_i непосредственно связан с одним из последующих субтекстов, а это и означает, что текст t всюду мотивирован.

Из доказанной теоремы T_8 видно, что хотя прочность связностной организации цепи без возвращений невелика, сама степень организованности довольно высока. Более того, цепь без возвращений может быть даже всюду связным текстом – таков, напр., текст, в котором каждый субтекст одним повтором связан с предыдущим, а другим повтором с последующим, причем каждый конкретный повтор используется для связи не более, чем двух субтекстов. При изъятии субтекста связность такой цепи нарушается именно в том месте, где произошло изъятие. Относительно разложения на предложения такова организация следующего отрывка из известной пародии А.Г.Архангельского на В.Б.Шкловского[37]).

Не каждый, умеющий садиться, умеет писать.
Садятся и на извозчика.
От Страстной до Арбата извозчик берет рубль.
Седок сердится.
Я тоже ворчу. (25)

37 А.Г.Архангельский. Избранное. Пародии. Эпиграммы. Сатира. М., ОГИЗ ГИХЛ, 1946, стр. 97.

Теоретически такая цепь может длиться сколь угодно долго в обе стороны и связность ее (и даже всюду-связность в нашем примере) будет при этом сохраняться. Языковое и стилистическое чутье пародиста проявилось, однако, в том, что так построена не вся пародия, а лишь маленький кусочек. Поскольку каждый связующий повтор может встретиться не более двух раз, то длинный текст должен очень быстро удалиться от тематики начальных предложений. Поэтому в реальных текстах рано или поздно член какого-нибудь повтора встречается в третий раз и происходит "возвращение" к каким-то из предыдущих предложений. Поэтому стиль реальных текстов (напр., пародируемого автора) будет скорее отличать не подное их построение по схеме всюду связной цепи без возвращений, а само наличие в них (и, возможно, повышенная частота) участков, устроенных по этой схеме.

Любой остов всюду связной цепи без возвращений длины m описанного вида должен, как легко убедиться, содержать не меньше $\left[\frac{m-1}{2}\right]$ субтекстов.

Другим полюсом прочности связностной организации являются

O_{16} : Максимально устойчивые тексты, т.е. такие, из которых можно вынуть любое число субтекстов от 1 до $(m-2)$, но они при этом сохраняют связность (при изъятии $(m-1)$ субтекста остается всего один субтекст и говорить о связности - несвязности бессмысленно).

Важной чертой максимально устойчивых текстов является возможность сопоставления им одноэлементных остовов. Имеет место

T_9 : Если текст $t = \theta_1, \ldots, \theta_m$ сохраняет связность при изъятии из него k произвольно выбранных субтекстов, где K - любое число от 1 до $m-2$, то t имеет по крайней мере один одноэлементный остов.

Доказательство будем вести от противного. Допустим, что t не имеет ни одного одноэлементного остова. По определению остова это означает, что для любого субтекста θ_1 $(1 \leqslant i \leqslant m)$ найдется такой субтекст θ_{j_1} $(j_1 \neq i, 1 \leqslant j_1 \leqslant m)$, что θ_i, θ_{j_1} - несвязна. Фиксируем некоторое i_0 и возьмем субтексты

θ_{i_o} и $\theta_{j_{i_o}}$, тогда все прочие субтексты образуют группу из $(m-2)$ членов, при изъятии которой текст t теряет связность. Но существование такой группы противоречит условию теоремы. Значит, наше допущение неверно, что доказывает теорему.

Более того, искомый одноэлементный остов может быть указан и вполне конструктивно: поскольку в максимально устойчивых текстах все субтексты, промежуточные между θ_1 и любым θ_k, и все, следующие за θ_k, могут быть изъяты без нарушения связности, каждый субтекст θ_k потенциально связан с θ_1 и потому в качестве одноэлементного остова всегда можно выбрать первый субтекст θ_1, который как бы "задает тон" всему тексту. Эта ключевая роль первого субтекста максимально устойчивых текстов станет яснее после разбора возможных типов таких текстов.

Сначала коснемся двух "чистых" разновидностей таких текстов. Первая - это "тривиально устойчивые" тексты, в каждом субтексте которых есть связка. Другую, несравненно более интересную, составляют тексты, которые можно назвать г и м н и ч е с к и м и, - такие, в которых можно выделить элемент повтора, входящий в структурную характеристику любого субтекста: в разных местах текста могут возникать дополнительные, локальные повторы и связки, но и в этих по-иному связанных субтекстах должен присутствовать сквозной повтор. Среди стихотворных текстов, если считать субтекстом каждую строку, а круг возможных повторов ограничить концевыми созвучиями, гимническими будут так называемые моноримы - стихотворения, стянутые одной рифмой.

В плане лексической и семантической связности к этой категории относятся всевозможные стихотворные и прозаические гимны (откуда и название всего класса текстов) - т.е. тексты, посвященные воспеванию какого-то одного объекта, перечислению его свойств и характеристик, допускающие отступления от этого перечисления лишь для того, чтобы обосновать очередную характеристику (такое отступление естественно объединить в один субтекст с обоснованной в нем характеристикой). Такова известная Гетевская статья "Природа"[38], каждое предложение которой содержит либо слово природа, либо заменяющее его местоимение она, тако-

[38] И.В.Г е т е. Избранное. М., Детгиз, 1963, стр. 488-490.

во в переводе А.А.Ахматовой древнеегипетское "Прославление писцов"[39] относительно разложения на субтексты, заключенные между двумя интервалами: в каждом из таких субтекстов встречается либо слово писцы, либо его субститут они, либо слова – писания, написанный, написать. Гимническим будет и проспект какого-нибудь лекарства:

X помогает при головной боли.
Принимающий X избавляется от желудочных колик.
X снимает все виды судорог.
Регулярное употребление X способствует хорошему сну. (26)

В гимнических текстах особенно прозрачна ключевая функция первого субтекста, на которую указывалось выше: первый субтекст задает в них "тему" или "героя" всего текста.

Было бы заманчиво предположить, что любой максимально устойчивый текст может быть разбит на отрезки, каждый из которых, в свою очередь, является либо гимническим, либо тривиально устойчивым текстом. Следующий простой текст служит, однако, опровергающим примером:

Петя вошел в комнату.
Ваня вошел в комнату.
Петя и Ваня сели обедать. (27)

Однако возможности построения длинных максимально устойчивых текстов, не поддающихся указанному разбиению, все же весьма ограничены – при отсутствии единого сквозного повтора каждый субтекст должен содержать количество элементов, достаточное для вступления в повторы со всеми другими субтекстами. Так, если ограничиться разложением текста на предложения и лексическим типом повторов, то следующий короткий текст:

Петя вошел в комнату.
Ваня вошел в комнату.
Петя и Ваня друзья.
Сережа тоже друг Пети и Вани.
Сережа не стал обедать с Петей и Ваней. (28)

представляется едва ли не предельным, ибо для того, чтобы текст не перестал быть максимально устойчивым, шестое предложение долж-

[39] Лирика Древнего Египта. М., "Худ. лит.", 1965, стр. 89-93.

содержать уже не менее трех различных повторяемых лексем, а последующие предложения и того больше, если только текст не станет с некоторого места гимническим (например, текст (28) может получить как угодно длинное продолжение, в котором каждое предложение будет содержать имена Петя, Ваня и Сережа). Поэтому можно высказать гипотезу, что отрезки максимально устойчивых текстов, не являющиеся гимническими, имеют в реальных текстах сравнительно небольшую длину.

Некоторые другие типы поведения текста при изъятии субтекстов будут рассмотрены нами ниже в § 8, так как их удобнее описывать в терминах перехода от одного фиксированного разложения текста к другому.

8. Элементы типологии связных текстов на основе соотношения их связностных характеристик относительно различных разложений.

Введение в рассмотрение произвольных разложений текстов позволяет классифицировать связные тексты по тому, как изменяются или, наоборот, не изменяются их связностные характеристики (связность, всюду-связность, всюду-мотивированность) при переходе от одного разложения к другому.

Напр., текст

Мяч упал в воду.
Таня заплакала.
Мяч поплыл по реке.
Таня заревела еще громче.
Папа достал мяч.
Таня успокоилась. (29)

если ограничиться учетом только лексических средств связи, относительно разложения на предложения только всюду мотивирован и и не всюду связен, но характерно, что можно так объединить предложения (напр., попарно), что относительно полученного разложения t ⟨2, 4, 6⟩ он будет всюду связен. Текст

Мяч упал в воду.
Таня заплакала. ЭВМ выдала неверный ответ.
Таня заревела уже по-настоящему.
По лицу Тани в два ручья лились слезы.
А мяч все плыл и плыл по реке. (30)

по предложениям даже не всюду-мотивированный, относительно разложения $t\langle 2, 4, 6\rangle$ становится не только всюду мотивированным, но и всюду связным. За исключением бинарных разложений, относительно которых, как отмечалось выше, всюду связен и тем более всюду мотивирован любой связный текст, все другие разложения могут быть диагностическими для классификации текстов по поведению их связностных характеристик.

Текст, всюду связный (всюду мотивированный) по цепочкам, будет, как мы помним, всюду связным (мотивированным) относительно любого своего разложения. С другой стороны, существуют связные тексты любой длины, не являющиеся всюду связными (всюду мотивированными) ни для одного своего небинарного разложения. Примером такого текста длины 3 может служить текст (I3), из которого нетрудно получить пример любой заданной длины.

O_{I7} : Будем называть у к р у п н е н и е м разложения $t\langle j_1, \ldots, j_{k-1}, j_k, j_{k+1}, \ldots, j_l\rangle$ любое разложение $t\langle j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_l\rangle$ при $k = 1, \ldots, l-1$.

Исчерпать тексты длины m, всюду связные (всюду мотивированные) относительно хотя бы одного l-арного разложения, где $2 < l < m-1$, помогает следующая очевидная

T_{I0} : Если текст $t = \tau_1, \ldots, \tau_m$ связен (всюду связен, всюду мотивирован) относительно $(l+1)$-арного разложения $t\langle j_1, \ldots, j_{l+1}\rangle$, то он будет связен (всюду связен, всюду мотивирован) относительно любого l-арного укрупнения этого разложения.

Вследствие T_{I0}, чтобы проверить, будет ли t связным (всюду связным, всюду мотивированным) относительно хотя бы одного небинарного разложения, достаточно проверить его связностные характеристики относительно всевозможных тернарных разложений. Если t не является связным (всюду связным, всюду мотивированным) относительно хотя бы одного тернарного разложения, он не будет связным (всюду связным, всюду мотивированным) и при любом l-арном разложении $(3 < l \leqslant m)$.

Тексты, всюду связные по цепочкам, и тексты, связные по цепочкам, но не всюду мотивированные ни при одном своем разложении, составляют как бы два полюса связных текстов со связностны-

ми характеристиками, инвариантными относительно всевозможных разложений. Все остальные связные тексты занимают промежуточное положение и могут менять свою связностную характеристику. Среди них могут быть выделены самые разнообразные классы. Нам представляется особенно интересным класс текстов, выделяемый следующим определением.

O_{18}: Число 1 называется **минимальным размахом** связного текста $t = \tau_1, \ldots, \tau_n$, если и только если текст t является всюду связным относительно любого разложения, в котором $j_1 = i$, и 1 – наименьшее из чисел, обладающих данным свойством. Текст t в этом случае называется **текстом с размахом 1**.

Связный текст $t = \tau_1, \ldots, \tau_m$ называется текстом **с ограниченным** размахом, если и только если его минимальный размах меньше $m - 1$.

В связных текстах с ограниченным размахом существует возможность "упрятать" все субтексты, нарушающие всюду-связность текста, в самый левый член некоторого небинарного разложения. Наиболее простым подклассом текстов с ограниченным размахом являются тексты, у которых субтекст $t_\Lambda^{(i)}$ может быть взят в качестве одноэлементного остова текста. При $i = 1$ этот подкласс совпадает с выделенным нами выше классом максимально-устойчивых текстов. Последний факт наводит на мысль, что целесообразно изучать влияние перехода от одного фиксированного разложения к другому не только на сами связностные характеристики но и на описанные в предыдущем разделе реакции связностных характеристик на удаление тех или иных субтекстов фиксированного разложения. Например, текст, являющийся гимническим относительно разложения на цепочки, будет гимническим и относительно любого другого разложения. Понятия максимально-устойчивого, тривиально-устойчивого и гимнического текста можно подвергнуть такому же обобщению, какому было при введении понятия текста с ограниченным размахом подвергнуто понятие всюду связного текста. Определение соответствующих понятий – текст, максимально-устойчивый (тривиально устойчивый, гимнический) вправо от точки 1, – строится, однако, очевидным образом, поэтому мы обра-

тимся сейчас к другому классу текстов – одному из наиболее интересных среди промежуточных между цепями без возвращений и гимническими текстами. Тексты этого типа, напр.:

Ерема ушел.
Фома убежал.
Ерема был крив.
Фома с бельмом. (31)

естественно характеризовать именно в терминах перехода от одного фиксированного разложения к другому. Отличительной чертой этих текстов является существование в них интервала некоторой фиксированной длины ε такого, что:

1) субтексты θ_j, $\theta_{j+\varepsilon}$, $\theta_{j+2\varepsilon}$, ... образуют гимнический текст, каково бы ни было $1 \leqslant j \leqslant m - \varepsilon$, где m – число субтекстов в данном разложении.

2) укрупнение исходного разложения, при котором в один укрупненный субтекст объединяются тексты с θ_1 по θ_ε, с $\theta_{\varepsilon+1}$ по $\theta_{2\varepsilon}$ и т.д., также является гимническим. Подобные тексты (их можно было бы назвать "кратно-гимничными") фактически представляют собой несколько текстов, вложенных друг в друга наподобие шестерен или гребенок. По этому типу устроены, например, русские переводы стихотворения Бодлера "Каин и Авель"[40], где $\varepsilon = 2$ (за субтекст принимаем двустишие).

Если требование гимничности заменить в условии 1 связностью, то получим "кратно-связные тексты". При $\varepsilon = 2$ и дополнительном требовании наличия при любом j связи между θ_j и θ_{j+1} получим аналог "амебейной композиции" В.М.Жирмунского[41].

БИБЛИОТЕКА
ИН-ТА РУССКОГО
ИН-ТА ЯЗЫКОЗНАНИЯ

---

40 Ш.Бодлер. Цветы зла. М., "Наука", 1970, стр. 201, 436–437.

41 В.М.Жирмунский. Композиция лирических стихотворений. Пг., ОПОЯЗ, 1921, стр. 35.

Т-18166 от 15.XI.71. 2,75 п.л. Тир. 150 Зак. 1075
Т п. МГПИИЯ

ИРЯ
институт русского языка
им. в. в. виноградова
www.ruslang.ru