# **РЕФЕРАТ**

##### **Дистрибутивный анализ**

**Оглавление:**

1. Введение -------------------------------------------------------------------------- 3

2. Основные допущения и термины -------------------------------------------- 4

3. Приемы дистрибутивного анализа ------------------------------------------ 6

4. Основные этапы дистрибутивного анализа ------------------------------- 8

5. Основные процедуры анализа и типы дистрибутивных отношений 10

6. Дистрибутивные методы в фонологии ------------------------------------ 12

а). Процедуры отождествления I ---------------------------------------------- 12

б). Процедуры отождествления II (собственно фонология) ------------- 14

в). Дополнительные критерии для объединения элементов в фонемы. 15

7. Общие итоги дистрибутивного анализа на фонемном уровне ------- 17

8. Выводы. ------------------------------------------------------------------------- 19

Список литературы --------------------------------------------------------------- 20

**Введение.**

Одной из узловых проблем общего языкознания является проблема методов лингвистической науки. Преобладание того или иного метода в определенную историческую эпоху может даже определять общий характер развития языкознания.

Сами конкретные методы лингвистических исследований обладают довольно большим разнообразием. Они сильно варьируют в зависимости от характера объекта, цели исследования и выбранного лингвистом аспекта исследования, например, существуют особые методы расшифровки текстов на неизвестных языках, специальные методы исследования фонетических характеристик звуков языка, методы сбора языкового материала в полевых условиях, особые лингвосоциологические методы, методы изучения истории литературных языков и т.д. Словом, чуть ли не каждая область языка может иметь специфические методы, предназначенные для ее изучения.

Выдвижение в качестве предмета изучения структурных свойств языка требует особых методов, значительно более сложных, чем простое наблюдение и классификация единиц речи.

Дистрибутивный анализ как особая методика исследования языка сформировалась в аме6риканской дескриптивной лингвистике. В 40-60-е годы появились многочисленные работы, посвященные описанию отдельных сторон фонетики, морфологии, синтаксиса языков различного строя.

Одним из важных следствий, вытекающих из работ дескриптивистов, является постановка вопроса о моделировании лингвистического исследования. Модели, имитирующие исследовательскую деятельность лингвиста, предназначаются для того, чтобы обосновать выбор основных понятий, которыми он пользуется при построении грамматики или другого описания языка. В идеале они сводят до минимума роль субъективного фактора в исследовании. По сложности проблематики модели такого типа стоят рядом с моделями творческой деятельности и научного поиска вообще. Эти процессы могут рассматриваться как кибернетические процессы обработки информации.

Дескриптивная лингвистика не выдвинула своей концепции языка как предмета исследования. Ее задачи сводятся к описанию структуры плана выражения, который рассматривается в отвлечении от плана содержания. Структура языка анализируется в терминах дистрибутивных отношений (распределений) элементов различных ярусов. Конечный результат исследования заключается в построении компактного описания языка, которое обеспечивало бы понимание и построение любых высказываний на данном языке. Никаких дополнительных ограничений на форму описания не налагается, поэтому допустимы расхождения в результатах анализа у разных исследователей, работающих с одним и тем же материалом. Описания, выполненные с помощью дистрибутивных методов, предназначаются для использования в других лингвистических разделах (сравнительное изучение языков, диалектология и т.д.).

Такое ограничение задач способствовало разработке практических процедур анализа, которые применялись в процессе построения описания. Была сделана попытка изложить дистрибутивные методы как строгую систему исследовательских приемов, которая не имеет аналогии в языкознание предшествующего периода.

Этот опыт, однако, не дал идеальных результатов. Оказалось, что процесс исследования языка является весьма сложным и не допускает полной формализации процедур на всех этапах работы.

**Основные допущения и термины.**

В качестве общих неопределяемых понятий в дистрибутивном анализе используются следующие:

Универсум речи (анг. - universe of discourse) – объект исследования. Наряду с этим термином в лингвистике принято употреблять и другие: континуум речи, совокупность текстов, корпус; выборка (анг. - sample), представляющая данный язык или его диалект в ограниченном промежутке времени (синхронном срезе).

Конкретный акт речи (к.а.р.) (анг. – act of communication) – любой отрезок речи одного из участников коммуникации, ограниченный с обеих сторон молчанием. Речь состоит из к.а.р., которые являются однократными, происходят в определенный момент времени. Повторяющиеся отрезки речи являются разными к.а.р. (например: снег, снег, снег…). Описать к.а.р. невозможно из-за многообразия случайных моментов, сопровождающих речь индивидуума.

При описании языка выделяют некоторые существенные черты речи, которые влияют на ее содержание. Понятие существенных признаков также неопределяемо. Каждая наука имеет своим представления о том, что существенно.

Лингвист обычно описывает абстрактные акты речи или высказывания (анг. utterance), - освобожденные от конкретных, ситуативных физических признаков, отрезки речи между двумя паузами.

Получение высказываний связано с операцией отождествления к.а.р. по существенным признакам.

Необходимые условия для такого отождествления: сходство содержания к.а.р.; сходство внешних физических признаков звучания к.а.р.. Степень сходства точно не определяется. По-видимому, нельзя отождествить к.а.р. только на основе подобия в смысле (глаза и очи) или только на базе подобия во внешнем звучании (ключ – «источник» и ключ – «инструмент для отпирания замка»).

В результате мы имеем высказывания, которые имеют разные реализации в речи, каждому высказыванию соответствует множество к.а.р..

Термин высказывание соответствует различным текстовым отрезкам: слову, звуку, фразе, предложению.

Принимается фундаментальная гипотеза о том, что непрерывные акты речи (а.р.) делятся на части, называемые сегментами. Высказывание рассматривается как дискретная последовательность отрезков речи, звуковых или морфемных сегментов, цепочек морфем и т.д. Возможны различные способы сегментации текста, повторные операции сегментации на каждом этапе описания. Сегмент – это отрезок речи, вычленяемой в данной последовательности к.а.р..

Понятие сегмента занимает важное место в дистрибутивном анализе. Дескриптивисты строго разграничивают операции сегментирования и последующие операции отождествления сегментов. В этом состоит принципиальное отличие их от традиционной лингвистики, которая обычно не ставит вопроса о сегментации речевого материала. При этом либо безоговорочно принимаются графические и орфографические членения текста, либо процедуры сегментации включаются в само описания и классификацию единиц речи. В методическом отношении оказывается очень важным не смешивать сегментацию с классификацией. При этом подходе допускается гипотетическое расчленение текста, которое в случае неточностей или обнаруженных в последствии ошибочных решений может быть изменено, т.е. может быть проведена пересегментация.

Сегмент противопоставляется элементу. Элемент – это абстрактный сегмент, освобожденный от конкретных ситуативных черт. Элементы определяются путем сопоставления различных отрезков речи, т.е. в результате операции отождествления сегментов по существенным признакам. Элемент – это обозначенный определенным символом сегмент. Так, буквы или транскрипционные знаки используют для обозначения звуковых сегментов. Сегмент принадлежит высказыванию, а элемент – описанию высказываний.

Сегмент и элемент – два важнейших процедурных термина, которые последовательно употребляются на каждом этапе анализа. Остановимся на некоторых дополнительных пояснениях этих понятий:

1. Сегменты – простейшие отрезки, выделяемые при расчленении континуума высказывания на первом этапе анализа текста. Элементы формируются путем объединения эквивалентных сегментов.

Пример: если мы имеем следующие высказывания, расчлененные на звуковые сегменты:

дарога

дорога

дароγа

то можно следующим образом отождествить эквивалентные сегменты:

[дърога]

1. Сегмент есть всегда отрезок текста, принадлежит тексту, элемент же принадлежит описанию, его соотнесенность с текстом опосредована через одну ступень его классификации. Поэтому элемент всегда является сегментом. Обратное же не имеет смысла. Так, в нашем примере [д] = д, [р] = р, [о] = о, [а] = а и другие случаи, когда [ъ] = {а, о}, [г] = {г, γ}.
2. Сегмент можно рассматривать со стороны его состава, его физических признаков (например, взрывность и фрикативность [г]). Элемент не рассматривается по составу, т.к. является цельной единицей.

Аналогично этому сегмент можно рассматривать как часть более сложного сегмента. Элемент обычно рассматривают безотносительно к иерархически более сложным элементам.

В процессе описания многократно осуществляется переход от рассмотрения высказываний как последовательных сегментов к их рассмотрению как цепочек элементов. Шаги перехода можно изобразить на схеме:

к.а.р. высказывание

сегмент элемент

Как видно, движение по циклу (высказывание сегмент элемент высказывание) может проводиться несколько раз. Такая возможность допускается в пределах одного уровня языка, а также при переходе с одного уровня на другой.

Следующее важное понятие – окружение (анг. – environment, position). Окружение элемента – это соседство в акте речи других элементов. Окружения могут быть различными, могут иметь произвольную длину. Так в непосредственное окружение элемента х входят элементы, предшествующие ему (ах, bах и т.д.), или следующие за ним (хс, хсd и т.д.). Полное окружение х включает весь набор элементов в данном высказывании (ахс или bахсd).

Дистрибуция или распределение (анг. distribution) элемента – совокупность всех окружений, в которых он встречается в речи. Дистрибуция представляет собой сумму всех позиций элемента относительно встречаемости других элементов.

О двух (или более) высказываниях или признаках говорят, что они лингвистически, дескриптивно или дистрибутивно эквивалентны, если они идентичны по составу элементов и дистрибутивным отношениям между элементами.

Понятие окружения и дистрибуции коррелятивны между собой так же, как сегмент и элемент. Подобно тому как элементы получаются в результате отождествления эквивалентных сегментов, дистрибуция есть результат типизации всех окружений данного элемента. Процедуры выявления элементов проводятся параллельно с процедурами определения дистрибуций. При этом осуществляется последовательный переход с одного шага анализа на другой и обратно. Учитывая это обстоятельство, можно дополнить схему:

к.а.р. высказывание

сегмент элемент

окружение дистрибуция

Понятия окружения и дистрибуции при нестрогом употреблении терминов часто смешивают друг с другом и отождествляют.

Однако между понятиями дистрибуции и окружения имеется существенное различие: объектами в анализе окружений являются элементы, связанные конъюнкцией (отношением одновременного вхождения – «и»); объектами в анализе дистрибуции являются сами окружения, связанные дизъюнкцией (отношением выбора – «или»). Так, в приведенном примере непосредственное окружение | д | составляет последовательность # | -ъ |. Это окружение, наряду со многими другими входит в дистрибуцию | д |:

# - ъ | дърога |

# - у | дума |

# - р | драка |

Таким образом, дистрибутивный анализ строится на основе синтагматических отношений между элементами. Однако его результаты касаются выделения классов эквивалентных элементов, т.е. охватывают область парадигматики языка.

**Основные экспериментальные приемы, применяемые в дистрибутивном анализе:**

Ю.Д. Апресян выделяет следующие приемы:

1) Добавление элементов к данной форме.

Например, в английском языке из-за омонимии предлогов и наречий типа up, in, on и др. довольно трудно решить вопрос о грамматической природе этих элементов. Возьмем два предложения she was brought up by her grandmother - «она была воспитана бабушкой» и she is going up the hill – «она поднимается в гору». Для того чтобы определить, в каком предложении up является предлогом, а в каком наречием, необходимо поставить между глаголом и спорным элементом качественное наречие и проверить, получается ли в результате правильная фраза. Правильность фразы диагностирует предлог (she is going quickly up the hill – «она быстро поднимается в гору»), а неправильность – наречие (she was brought badly up by her grandmother).

1. Опущение элементов из данной формы.

Например, можно опустить косвенное дополнение, которое менее тесно связано с глаголом, чем прямое; прямое же дополнение опускать нельзя: Martin caused me pain - «Мартин причинил мне боль» → Martin caused pain - «Мартин причинил боль», но не Martin caused me – «Мартин причинил мне».

1. перестановка элементов данной формы.

Возможность перестановки двух простых предложений в составе сложного указывает на наличие связи сочинения между ними, а невозможность такой перестановки свидетельствует о подчинительном характере этой связи.

Им можно, а вам нельзя → Вам нельзя, а им можно.

НО: Она подумала, что собака убежала → Собака убежала, что она подумала.

1. Субституция (замена) элемента другим элементом.

Для построения классов (частей речи, словоформ) используется принцип субституции: два элемента входят в один класс, если они способны замещать друг друга в одних и тех же окружениях без нарушения правильности данной фразы. Два элемента x1 и x2 считаются эквивалентными, если для любой правильной фразы вида A1x1A2 существует правильная фраза вида A1x2A2, и наоборот.

В зависимости от того, что считается элементом, окружением элемента и правильной фразой, субституция дает более или менее дробные классы: чем ближе к текстовой единице элемент и его окружение и чем жестче ограничения на правильность фразы, тем более дробными получаются классы. Например, возьмем фразу я читаю книгу. Здесь x1 является слово читаю, которое мы можем заменить на слова беру, открываю и т.д., т.е. на x2: я беру книгу; я открываю книгу.

Аналогично мы можем заменить x1 на x2 и в словах: т…м ( вместо многоточия мы можем поставить или о, или а и т.д.)

В классической дистрибутивной лингвистике в качестве элементов рассматриваются основы, их окружением считаются грамматические морфемы или классы основ и грамматических морфем, а на правильность фраз налагаются жесткие ограничения.

При замене одного элемента другим нельзя забывать о согласовательной связи: это любая синтаксическая связь между существительным и согласуемым словом, при которой форма согласуемого слова зависит от выбора существительного или от формы, в которой оно стоит. Помимо атрибутивной и предикативной связи, к этому типу синтаксический связи относятся а) связь через предлог из (один из людей, но одна из девушек);

б) связь через глагол (эту затею считают безнадежной, но это дело считают безнадежным);

в) связь со словом «который» в придаточном предложении (дело, в котором, но статья, в которой) и др.

Таким образом, через согласовательную связь определяется ключевое понятие согласовательного класса, т.е. такая совокупность существительных, что любые два ее члена, будучи взяты в любой грамматической форме, требуют при любом типе согласовательной связи одной и той же словоформы любого согласуемого слова.

1. Метод непосредственных составляющих (НС).

Метод НС используется для получения синтаксической информации.

Метод НС основан на следующих содержательных допущениях: 1) существенную роль в синтаксической структуре предложения играет одно единственное отношение – отношение подчинения; 2) предложение не собирается непосредственно из словоформ, но строится последовательно, иерархически: элемент, его части соединяются в простые, «строительные» блоки, из этих блоков строятся более крупные блоки, пока все предложение не будет представлено в виде единичного блока; 3) верхом иерархии является сказуемое, или группа подлежащего и сказуемого.

В ходе анализа устанавливаются как единичные предложения, так и отношения между ними.

Для дальнейших определений, Ю.Д. Апресян вводит понятие ранга. Пара символов XY является синтагмой первого ранга, если в любой содержащей ее правильной конструкции она может быть заменена без нарушения правильности конструкции символом Y, а Y заменен парой XY. Возьмем предложение

### Моя подруга купила очень красивое платье

An Nn V D Aa Na

В данном примере синтагмой первого ранга можно считать пару символов DAn (очень красивое платье), которую можно заменить на An (красивое платье). Синтагмой n-ого ранга называется пара символов XY, если в любой правильной конструкции, где X не входит в состав синтагмы (n – 1)-го ранга, она заменима символом Y, и наоборот. X и Y – непосредственные составляющие синтагмы, причем Y – ядро, X – зависимый элемент. Благодаря этому можно установить синтаксические связи между словоформами и иерархию связей. В приведенном выше примере синтагмами являются AnNn, DAa, NnV и др., причем главным является 2-ой элемент, а зависимым – 1-ый. Каждая синтагма по своим внешним синтаксическим свойствам эквивалентна ядру, взятому в отдельности. Анализ по НС называется свертыванием:

D + Ax→ Ax

Ax + Nx → Nx

Результирующий элемент предыдущей синтагмы является НС следующей синтагмы, т.е. более простая синтагма может вкладываться в более сложную, выступая в качестве НС последней.

Существует несколько правил свертывания по НС: 1) одновременно можно свертывать не более двух символов в один новый; 2) запрещается перестановка символов; 3) порядок применения правил фиксирован.

Весь процесс анализа предложения (свертывания по НС) может быть изображен тремя способами: в виде правил подстановки, в виде скобочной записи и в виде дерева НС.

Однако эта модель имеет много недостатков: 1) в ряде случаев приводит к грубейшим решениям, поскольку не дает возможности фиксировать различия в синтаксической структуре предложений. Возьмем два предложения:

Изучение событий становится интересным

Развитие событий становится интересным.

В первом случае события являются объектом некоторого действия, во втором же – события являются субъектом действия (они сами развиваются). Анализ по НС не вскрывает этих структурных различий. 2) метод НС не показывает связи между активными и пассивными, утвердительными и отрицательными, утвердительными и вопросительными конструкциями, которые связаны отношением производности. 3) с трудом поддаются анализу экзоцентрические конструкции, конструкции с так называемыми разрывными морфемами и конструкции с аппозитивными элементами вида товарищ Иванов, князь Андрей и др. 4) по настоящему эффективной эта модель является только в рамках простого предложения. 5) не дает возможности ответить на вопрос, что является элементарной единицей языка на синтаксическом уровне.

**Основные этапы дистрибутивного анализа.**

I. Долингвистический этап.

Исходные данные здесь – к.а.р. как физические процессы. Их части имеют физические признаки в каждый данный момент времени (характеризуются амплитудой, частотой и т.д.).

В качестве внешнего понятия используются существенные признаки к.а.р. и одинаковость содержания.

Основные операции на этом этапе: а) сегментирование а.р. на фонологические сегменты; б) классификация сегментов по существенным, смыслоразличительным признакам.

Результат – выделение фонемных элементов как классов конкретных фонологических сегментов.

II. Фонологический этап.

Элементы описания: символы фонологических элементов.

Исходные данные: высказывания, представленные как цепочки фонемных элементов.

Внешнее понятие: физическое сходство фонемных элементов.

Операции: классификация фонемных элементов по критерию дополнительной дистрибуции и физическому сходству.

Результат – выделение фонем как классов фонемных элементов (аллофонов).

III. Морфемный этап.

Элементы: символы фонем.

Исходные данные: высказывания как цепочки фонем.

Внешние пон6ятия: связь цепочки фонем с содержанием (наличие значения) и одинаковость значения.

Операции: а) сегментация – выделение морфемных сегментов как цепочек фонем (минимальных, имеющих значение); введение фиктивных морфемных элементов (для устранения неправильностей и исключений);

б) классификация морфемных сегментов по тождеству состава фонем и по дополнительности.

Результат: выделение морфем как классов морфемных элементов. Их очень много, больше, чем фонем, но теперь они связаны с содержанием.

IV. Морфосинтаксический уровень.

Элементы: символы морфем.

Внешне понятие: грамматическая правильность высказывания.

Операции: а) классификация морфем по взаимной подставимости в любых высказываниях без изменения их грамматической правильности.

б) группировка по критерию дополнительности.

Результат: Выделение класса морфем как множества эквивалентных морфем.

Количество элементов уменьшается за счет эквивалентности корневых и других морфем.

V. Синтаксический уровень.

Элементы: символы классов морфем.

Исходные данные: высказывания как цепочки классов морфем.

Внешнее понятие: фраза в данном высказывании или другой фразе.

Операции: а). Сегментация на фразы, т.е. цепочки классов морфем.

б). Классификация цепочек классов морфем по их взаимозаменяемости во всех высказываниях, где они являются фразами, без изменения грамматической правильности высказываний

в). Выписывание синтаксических уровней.

Результат: Синтаксические уравнения и результирующие классы.

В соответствии с этими этапами описания лингвистическая структура плана выражения складывается из элементов 4-х ярусов:

Синтаксический ярус – позиционные классы и их цепочки.

Морфосинтаксический ярус – классы морфем и их цепочки.

Морфемный ярус – морфемы.

Фонемный ярус – фонемы.

Каждый уровень имеет свой перечень элементов и дистрибутивных отношений между ними.

При переходе к каждому следующему уровню описания сохраняется возможность однозначного возврата к элементам предыдущего уровня. Или, имея высказывание, представленное через элементы некоторого уровня, можно восстановить его представление через элементы предыдущего уровня. Для этого применяются дополнительные данные, получаемые на каждом очередном этапе анализа.

Дистрибутивный анализ представляет собой целостную систему методов описания.

На каждом уровне:

1). Определяется форма получаемой части описания или форма записи, представления классификации.

2). Указывается вид процедуры:

а) выявление элементов на основе внешних понятий,

б) способы классификации,

в) способ получения синтаксических уравнений;

3). Даются практические рекомендации и рецепты, как проводить анализ:

а) интуитивные пояснения на конкретных примерах по аналогии,

б) значение критерия экономности и простоты получаемого описания.

**Основные процедуры анализа и типы дистрибутивных отношений.**

Эмпирически данным материалом для дистрибутивного анализа является конкретные акты речи. Практически лингвист имеет дело с абстрактными актами речи, или высказываниями, которые получаются путем элементарного отождествления к.а.р. Процесс анализа сводится к двум основным процедурам, последовательно повторяющимся после завершения каждого этапа обработки данных. Это операции сегментирования и идентификации, или отождествления.

Эти операции противопоставляются друг другу: первая носит неформальный, или полуформальный характер, вторая является формальной. Иногда процедуры сегментирования рассматриваются на долингвистическом этапе обработки текста, к собственно лингвистическому этапу относятся процедуры идентификации.

Исследуемый материал дан нам в виде непрерывного речевого потока, или связного текста. Начальный этап анализа состоит в расчленении его на сегменты определенного типа, например звуковые отрезки. Процедуры сегментирования носят произвольный характер, членение может начинаться от произвольной точки в потоке речи. Механизм и правила сегментации не описаны формально. Здесь нельзя предложить определенного рецепта, можно только высказать некоторые рекомендации «как взяться за дело». Поэтому операции сегментирования занимают особое место в анализе, для них не создана какая-либо теория.

Собственно лингвистический этап начинается с процедур идентификации, которые проводятся последовательно за два шага.

Вначале исходные данные должны быть представлены в виде высказываний, расчлененных на сегменты. Задачи анализа на этом шаге состоят в отождествлении сегментов, сведении их в эквивалентные классы. С этой целью рассматриваются дистрибутивные отношения между разными сегментами. Возможны два типа отношений – эквивалентность и неэквивалентность.

Два сегмента считаются эквивалентными, если они выступают как свободные варианты друг друга, т.е. допускают взаимозамену, не изменяя содержания высказывания. Этот тип дистрибутивных отношений обычно называют отношением свободного варьирования или нефункциональным варьированием.

Примеры свободного варьирован6ия звуков: | г | и | γ | в высказываниях | дарога | и | дароγа |,

| бога | и | боγа | и т.д.

Отношение свободного варьирования представляет собой простейший вид эквивалентности, дистрибутивную эквивалентность «первой степени». На первом шаге анализа устраняется свободное варьирование сегментов, такие сегменты сводятся к одному элементу и в дальнейшем перестают различаться.

Противоположный тип отношений – неэквивалентность – имеет место при контрастной дистрибуции (анг. contrastive distribution) или функциональном варьировании. Два сегмента находятся в отношении контраста, если они встречаются в минимально различающихся парах высказываний. Например, | д | и | т | в дом : том. Варьирование звуков в таких парах высказываний сопровождается изменением их смысла (имеет место ковариативность звучания и значения). Такие звуки не допускают взаимозамены без изменения содержания высказываний. Они безусловно неэквивалентны друг другу и должны быть отнесены к разным элементам.

На втором шаге анализа исходные данные, высказывания представлены в виде последовательностей элементов. Задачи теперь сводятся к рассмотрению отношений между элементами с целью выделения классов эквивалентных элементов. На этом этапе мы обнаруживаем более сложный тип дистрибуции – дополнительную дистрибуцию (анг. complementary distribution). Отношения дополнительной дистрибуции связывают два (или более) таких элемента, один из которых встречается в таком окружении, где не встречается другой, и обратно, второй встречается в таком окружении, где невозможен первый. Оба элемента рассматриваются как позиционные варианты одного и того же класса элементов, для которого устанавливается суммарная дистрибуция. Например: | а | и | ’а’ | в мат – мять.

В итоге анализа исходный перечень элементов преобразуется в классы дистрибутивно эквивалентных элементов, производится количественное сокращение списка элементов. Полученные классы рассматриваются как элементы следующего этапа анализа, как единицы вышестоящего уровня структуры. Цикл процедур повторяется снова: данный набор классов проецируется на высказывания, проводится сегментирование текста в терминах новых элементов; затем приступают к анализу дистрибутивных отношений между сегментами и элементами.

Как было показано, два основных шага идентификации связаны с определением разных типов дистрибуции: нефункциональное и функциональное варьирование на первом шаге и дополнительная дистрибуция на втором. Эти шаги соответствуют двум ступеням анализа, осуществляемого в терминах сегментов – элементов и окружений – дистрибуций. Дополнив соотношения между основными терминами, приведенными выше, мы получим общую схему дистрибутивного анализа:

|  |  |
| --- | --- |
| Объекты языка Объекты метаязыка | Типы дистрибуций |
| К.а.р. высказывание  Сегменты Элементы Элементы Классы элементов + +  окружения Дистрибуция | Свободное варьирование  Контрастная дистрибуция  Дополнительная дистрибуция |

Принципы дистрибутивного анализа, как показывает их рассмотрение, представляет собой стройную систему операций, осуществляемых в терминах определенного метаязыка. Они описывают процесс лингвистического анализа в виде алгоритма. Именно эта особенность дистрибутивного анализа привлекает внимание специалистов, которые занимаются дешифровкой неизвестных систем письма и языковых кодов. Были сделаны попытки построить машинные алгоритмы выделения гласных и согласных в тексте на основе дистрибутивных признаков в сочетании со статическими.

Различают четыре типа дистрибуций, эти отношения можно представить рисунками:

1

**Дополнительная дистрибуция.**

В такой дистрибуции находятся варианты одной фонемы: каждый из них встречается лишь в тех позициях, в которых не встречается другой. Например: в русском языке фонемы [ы] и [и].

2

Контрастная дистрибуция. В такой дистрибуции находятся варианты разных фонем. Например: в русском языке [а] и [о] в словах ток и так, том и там, сон и сан – совпадающие части дистрибуции (на схеме – часть, образованная пересечением кругов); он, вон, кто с одной стороны, и Анна, ванна, стена – с другой стороны, - несовпадающие части дистрибуции.

3

Дистрибуция свободного варьирования. В подобной дистрибуции находятся свободные варианты одной и той же фонемы, например: було[ч]ная и було[ш]ная.

4

Дистрибуция частичной эквивалентности. В подобной дистрибуции находятся, например, русское взрывное [г] (на схеме – большой круг) и русское фрикативное [γ] (меньший круг), т.к. в норме русской речи оно может быть свободным вариантом взрывного [г] лишь в части окружений последнего – в словах Бога, благо и производных от них.

В дистрибутивном анализе формальные процедуры занимают весьма скромное место. Как следует из вышеприведенной схемы анализа, начальные операции сегментирования носят произвольный характер. Обращение к неявным сведениям о языке, которые лингвист привлекает извне (пользуясь интуицией или дополнительными данными о тексте), допускается на каждом этапе описания.

**Дистрибутивные методы в фонологии.**

Корпус материала предстает вначале как непрерывный поток конкретных актов речи. К.а.р. – это последовательности (цепочки) звуковых сегментов. Границы сегментов произвольны, поскольку у нас нет средств для указания, как точно разделить поток звуков на отрезки, которые в дальнейшем соответствовали бы его членению на фонемы. Последующие процедуры будут изменять эту сегментацию до тех пор, пока границы сегментов не совпадут с границами отдельных фонем.

**ПРОЦЕДУРЫ ОТОЖДЕСТВЛЕНИЯ I.**

Эти процедуры применяются, чтобы установить лингвистическую эквивалентность или неэквивалентность звуков-сегментов и перейти от уникальных неповторяющихся сегментов к меньшему числу элементов. Вначале рассматриваются пределы вариативности (различности) отдельных звуков в повторяющихся высказываниях. Имея запись какого-то высказывания, разделенного на сегменты, можно повторить его снова и записать. Говорят, что каждый сегмент одного повторения свободно подставляется вместо соответствующего сегмента другого повторения, или является его свободным вариантом.

Если высказывание АВС есть повторение высказывания А’В’С’, тогда А ~ А’, В ~ В’, С ~ С’.

После этого эксперименты с субституцией звуков проводятся в различающихся высказываниях. Например, проверим, можно ли подставить | б | из бутыль на место | б | в булка. Повторное произношение этих высказываний с заменой звуков будет подобно первому их произношению.

В общем случае берем высказывание с сегментами DEF, конструируем высказывание, составленное из DA’F, где А’ есть повторение А из некоторого высказывание АВС. Если информант воспринимает DA’F, как повторение DEF и если можно также построить E’BC (Е’ есть повторение Е), эквивалентное АВС, тогда А и Е (и А’ и Е’) взаимоподставимы, т.е. являются свободными вариантами друг друга (A ~ E).

Для Дескриптивного анализа важно установить не только эквивалентность некоторых сегментов друг другу, но также определить и неэквивалентность, неподставимость сегментов. Эксплицитно описать различия не менее важно, чем описать сходство, а чтобы иметь средства для различения одних высказываний от других, необходимо определить систему различительных элементов для многих высказываний. Так, чтобы различать дом, дол, дел, мы должны знать несколько различительных признаков между неэквивалентными высказываниями (дом отличается от дол третьим элементом, а от дел - вторым и третьим).

Представление речи как последовательности или аранжировки отдельных элементов связано с выделением звуковых различий между парой неэквивалентных высказываний.

Анализ пар неэквивалентных высказываний оказывается важным при установлении неэквивалентных звуков-сегментов. В особенности надежное определение разных звуков дают так называемые минимально различающиеся пары высказываний типа (дом : том), отличительными признаками которых являются одиночные звуковые сегменты. Эти сегменты | д | и | т | в данном случае имеют одинаковые окружения. Однако подстановка одного из них на место другого приводит к контрасту смысла, т.е. невозможна. Контрастирование смыслов высказываний свидетельствует о том, что различение | д | и | т | существенно для русского языка, т.е. что данные сегменты не эквивалентны, а различны.

В итоге определения эквивалентности и неэквивалентности сегментов получаем классы звуков-сегментов, или фонемные элементы. Случаи свободного варьирования исключены в данном перечне фонемных элементов. Так, если признаны эквивалентными высказываний |дарога | и | дароγа |, то соответственно перестают различаться звуки-сегменты | г | и | γ |. Они получают единое обозначение в виде фонемного элемента | г |. Приведем другие пары сегментов, которые могут считаться свободными вариантами. В русском языке допускается параллельное произношение слов | тэрмос | и | т’ермос |, | критэрий | и | крит’ерий | и т.п. Эти различия носят стилевой характер и не затрагивают смыслового содержания слов, ими можно пренебречь в целях компактного описания текстов. Тогда необходимо признать следующие эквивалентности: | т | ~ | т’ |, | э | ~ | е |. Такое «уравнивание» звуков имеет место только для данных высказываний, а не для любых их употреблений. Так | т | и | т’ | в тапка и тяпка невозможно считать эквивалентными.

Таким образом, получив перечень фонемных элементов, можно перезаписать исходные высказывания, устраняя несущественные для дальнейшего анализа различия звуковых признаков в эквивалентных актах речи. Все высказывания теперь расчленены на фонемные элементы, в каждом из них ровно столько единиц, сколько необходимо для различия высказываний, не являющихся повторениями друг друга, и не более.

Наряду с фонемными элементами, имеющими природу линейных сегментов, в высказываниях выделяются суперсегментные элементы. Такие звуковые черты речи, как тон, последовательность ударений, интонация, рассматриваются как «вторичные фонемы», «просодемы» (анг. contours). Они как бы размазаны по всей длине высказывания. Дистрибутивные методы выявляют эквивалентность высказываний по суперсегментным признакам безотносительно к их сегментному составу. Так парные высказывания с одинаковыми сегментами могут быть различными по смыслу и суперсегментным элементам: Он идет? : Он идет. Для каждого языка определяется весь набор просодем, каждая из них обозначается определенным символом | ? |, | ! |, | , |, | . | и т.д. и записывается после фонемных элементов.

**ПРОЦЕДУРЫ ОТОЖДЕСТВЛЕНИЯ II (СОБСТВЕННО ФОНОЛОГИЯ)**

Имея в качестве данных высказывания, записанные как цепочки элементов, переходим к рассмотрению их сочетаемости. Задача состоит в изучении закономерностей появления фонемных элементов в тексте и нахождении дистрибутивно эквивалентных элементов, которые можно сгруппировать в классы.

Представление актов речи в виде элементов и «просодем» имеет недостатки:

1. Каждый элемент характеризуется конкретными ограниченными окружениями,
2. Количество элементов велико.

Необходимо найти способы для типизации элементов и сокращения тем самым их первоначального числа. Условием для этого является выявление дополнительной дистрибуции.

Подготовка процедуры состоит в установлении окружений элементов и их суммировании.

Рассмотрим эти процедуры на примере. Пусть даны следующие высказывания: | разб’ила |, |бы |, | был |, | быт |, | б’ит’ |, | иба |, | л’уб’и |, | б’иб’и |, | двугорбый |, | падоб’ийе |. Определим окружения элементов | и | и | ы |. Для удобства рассмотрения выпишем все непосредственные окружения в таблицу №1.

табл. №1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Элементы | Окружения | | | | | | | | | |
| б’- л | б-# | б - л | б - т | б’- т’ | # - б | л’- б’ | б’- б’ | б - й | б’- й |
| И | + |  |  |  | + | + | + | + |  | + |
| Ы |  | + | + | + |  |  |  |  | + |  |

Сочетаемость элемента с окружением отмечается знаком + на пересечении строки и столбца. Как видно, число окружений велико. При расширении исходного списка высказываний это количество будет возрастать. Поэтому имеет смысл выявить характерные черты окружений и представить их в более компактном виде.

Просматривая окружения, легко заметить общие признаки их компонентов: наличие мягкости или ее отсутствие. По этим признакам все элементы в окружениях сгруппируем следующим образом: сгруппируем в один ряд мягкие, введя символ класса мягких С’, и все «немягкие» объединим во второй разряд С. С’ = {б’, т’, л’, й}, С = {б, т, л}. Тогда таблицу № 1 можно преобразовать в таблицу № 2.

табл. №2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | C’ - C | C - # | C - C | C’ – C’ | # - C | C – C’ |
| И | + |  |  | + | + |  |
| Ы |  | + | + |  |  | + |

Отождествление некоторых окружений по признакам мягкости\твердости привело к сокращению их числа с 10 до 6.

Продолжим наш анализ окружений. Обращаем внимание на то, что не все члены окружений одинаково существенны для выбора элементов | и | и | ы |. Так, последующие элементы не диагностируют выбор, ср. | и | появляется перед –с и –с’, также и | ы |. Напротив, предшествующие элементы определяют правила выбора гласных. Учитывая это, можно представить окружения только в составе диагностических признаков, исключив постпозитивные члены. Это показано в таблице № 3.

табл. № 3

|  |  |  |  |
| --- | --- | --- | --- |
|  | - C’- | - C - | # - |
| И | + |  | + |
| Ы |  | + |  |

Таким образом, упорядочения списка окружений приводит их к компактному виду. Теперь можно перейти к анализу дистрибуций элементов.

Таблица № 3 показывает, что дистрибуции | и | и | ы | дополнительны друг относительно друга: | и | встречается в таких окружениях, где не появляется | ы |, и наоборот. Отсюда следует, что | и | и | ы | находятся в отношении дополнительной дистрибуции. О таких элементах говорят, что их дистрибуции не пересекаются (в дистрибутивной таблице нет таких столбцов, в которых обе строки были бы заполнены знаком +). В общем виде дистрибутивная таблица имеет вид:

|  |  |  |
| --- | --- | --- |
|  | X - | Y - |
| a | + |  |
| b |  | + |

Элементы a и b не встречаются в одинаковых окружениях (X или Y), их окружения не пересекаются. По критерию дополнительной дистрибуции элементы объединяются в единый класс, который считается элементом следующего шага описания. Для обозначения классов фонемных элементов используем прописные буквы: И = {| и |, | ы |}.

Полученные классы элементов дескриптивисты называют фонемами, а члены класса – аллофонами, или позиционными вариантами фонемы. Для фонемы устанавливается полная дистрибуция как сумма дистрибуций аллофонов. Так, дистрибуция И включает окружения C’-, C -, # -.

**ДОПОЛНИТЕЛЬНЫЕ КРИТЕРИИ ДЛЯ ОБЪЕДИНЕНИЯ ЭЛЕМЕНТОВ В ФОНЕМЫ.**

Критерий дополнительной дистрибуции дает лингвисту строго формальные основания для группировки элементов. Такой тип отношений между элементами в терминах теории множеств интерпретируется как отношение множества к его дополнению. Однако этот формальный метод обеспечивает получение единственной классификации элементов только при условии, если число их не более двух. В практическом анализе приходится иметь дело с большим количеством элементов. В этом случае можно по-разному группировать элементы. Допустим, что дистрибутивная таблица имеет вид:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | X - | Y - | Z - | U - |
| a | + |  | + |  |
| b |  | + |  |  |
| c |  |  |  | + |
| d |  | + |  |  |

Тогда по критерию дополнительной дистрибуции можно построить следующие объединения элементов по два: {a, b}, {a, c}, {a, d}, {b, c}, {c, d}. Можно сгруппировать элементы и одним из следующих способов: {a, b, c}, {d} или {a, c, d} {b}. Таким образом, формальные правила разрешают несколько классификаций элементов.

Для уменьшения неоднозначностей в выделении фонем предлагаются некоторые другие критерии, которые лингвист может использовать по своему усмотрению.

1. Запрещается относить к одной фонеме элементы, которые входят в минимально различающиеся пары высказываний. Например, если будет обнаружено, что помимо учтенных высказываний, | и | и | ы | встречаются еще в такой паре названий букв, как «и», «ы», то это может рассматриваться в пользу необъединения их в один класс. Ср. другой пример из английского языка: при любой дистрибуции не следует объединять | k | и | t |, поскольку они образуют минимальные пары cry - «кричать, плакать» и try - «пытаться, стараться».

Иными словами, отношение дополнительной дистрибуции несовместимо с отношением контрастной дистрибуции у одной и той же пары элементов.

1. Критерий фонетического подобия в ряде случаев подсказывает пути группировки элементов. Покажем это на двух примерах. Пусть имеется дистрибутивная таблица для четырех элементов | и |, | ы |, | е |, | э |, которые встречаются в одинаковых окружениях.

|  |  |  |  |
| --- | --- | --- | --- |
|  | C’ - | C - | # - |
| И | + |  | + |
| Ы |  | + |  |
| Е | + |  |  |
| Э |  | + | + |

Такое распределение допускает двоякое объединение: | ы | может быть включен в класс {и, ы} или {ы, е}. Дополнительный учет степени фонетического сходства фонемных элементов поможет выбрать одну из группировок ( | ы | и | и | имеют общий фонетический признак – степень закрытости). Мы получаем фонемы {и, ы} и {е, э}.

1. Важным критерием при классификации является симметрия окружений. Удобно строить фонемы так, чтобы они имели большую свободу сочетаемости с другими элементами. Для дальнейшего анализа не желательно получать фонемы с ограниченными окружениями, специфичными для одной фонемы, ибо это затрудняет сопоставление фонем друг с другом.

Исходными данными для определения отношения симметрии служит список фонем обследуемого языка {a,b,c,d…} (исходное множество Р). Далее вводится абстрактная система возможностей двух фонемных сочетаний, которая заключается в том, что любая фонема может предшествовать или следовать за любой другой, включая саму себя: aa, ab, ac…ba, bb, bc…

Строится квадрат множества Р x Р:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | a | b | c | d |  |
| a | aa | ab | ac | ad | … |
| b | ba | bb | bc | … | … |
| c | ca | cb | … |  |  |
| d | da | … |  |  |  |

Такой квадрат является эталоном, с которым соотносят реальную дистрибуцию. Последняя же не покрывает абстрактных возможностей сочетания фонем и заключается в том, что только некоторые сочетания действительно встречаются.

Далее устанавливаются отношения каждой фонемы ко всему множеству фонем Р.

Предварительно устанавливаются следующие понятия:

αx – альфа-поле – подмножество фонем, принадлежащих исходному множеству Р и способных выступать в качестве первого члена в последовательностях из двух фонем, если х – второй член последовательности.

βх – бета-поле – подмножество фонем, принадлежащих исходному множеству Р и способных выступать в качестве второго члена в последовательностях из двух фонем, если х – первый член последовательности.

τх – тау-поле, которое является суммой αx и βх (τх = αx U βх)

Так же определяется и пересечение αx и βх (τх = αx ∩ βх)

После этого приступают к выявлению отношений симметрии. Фонема х считается симметричной фонеме y, если встречается как последовательность xy, так и yx.

**Общие итоги дистрибутивного анализа на фонетическом уровне.**

Процедуры отождествления фонемных элементов по критерию дополнительной дистрибуции открывают различные возможности выделения фонем. Полученная классификация может подвергаться дальнейшему анализу и уточнению. Так, можно найти пути для устранения ограничений в дистрибуции фонем за счет учета стыков, т.е. границ между цепочками фонем, отмеченных паузами. Учитывая специфику сочетаемости фонем в началах и концах высказываний, удается сделать классификацию более обобщенной.

Повторная фонемизация также может иметь место в тех случаях, когда оказывается удобным представить какой-либо элемент с резко ограниченной дистрибуцией в виде двух сегментов с большей свободой сочетаемости (например, | č | в виде | tš |).

Конечная цель дистрибутивного анализа на уровне фонем состоит в том, чтобы определить фонемную структуру языка, т.е. дать описание фонемного состава актов речи. Для этого проводится учет комбинаций фонем в речи и их типизация. Так, если | p |, |b |, | t |, | d |, | k |, | g| встречаются перед | a |, | i |, | u |, то вводится символ класса фонем С для каждого из шести согласных и V для каждой из трех гласных, и говорят, что встречаются СV. Это утверждение равносильно утверждению, что встречаются | pa |, | pi |, | pu |, | ba | и т.д. Выделяют классы фонем (типа согласных и гласных), выводятся формулы сочетаемости их в высказываниях. Так, для языка йокут установлена следующая структура высказывания: # [CV(C:)] CV(C)#, где # – знак стыка высказываний и границ контура, C - согласный, V - гласный, : - долгота фонемы; в круглые скобки заключены компоненты, которые иногда встречаются, иногда не встречаются, в квадратные скобки – компоненты, которые встречаются редко. Подставляя вместо символов определенные фонемы, можно построить любое высказывание языка.

Оценивая результаты дистрибутивного анализа в фонологии, следует отметить некоторые важные моменты.

1. Последовательные процедуры идентификации приводят к ограниченному перечню фонем, количество которых значительно меньше, чем исходное множество сегментов речи. В особенности значительную роль при этом играет объединение элементов по критерию дополнительн6ой дистрибуции. Он использует как объективное средство устранения различий между объектами и по существу стоит на службе следующего этапа анализа.
2. Каждая фонема представляет класс аллофонов и имеет дистрибутивную характеристику. Эти дистрибуции обеспечивают однозначное преобразование записи высказываний в виде последовательностей фонем в последовательности аллофонов и обратно. По дистрибутивным таблицам можно составить алгоритмические правила выбора фонемного элемента по данному окружению, или, обратно, выбора окружения по данному аллофону.
3. Дистрибутивный анализ допускает также определение закономерностей комбинаторики фонем в пределах высказываний, укрупнение классов элементов и выявление фонемных структур, которые представляются либо в формулах, либо в диаграммах.
4. В анализе наряду с формальными критериями идентификации используются и внешние данные: оценки эквивалентности высказываний при субституции звуков, а также сведения о распределении звуков в тексте.
5. Дистрибутивные методы, несмотря на громоздкость применяемых операций, могут использоваться в экспериментальных исследованиях по дешифровке системы фонем. По-видимому, для хорошо изученных языков эти методы имеет смысл применять не глобально, а выборочно, при анализе неясных, спорных соотношений между звуками.

**Выводы.**

Принципы дистрибутивного анализа имеют большое научное значение, прежде всего в том плане, что они углубляют наши знания о процессе отождествления единиц речи. Тождества и различия – это то, на чем держится вся языковая система. Нормальная речевая деятельность осуществляется только при условии четкой работы механизмов отождествления и различия языковых элементов. Анализирую факты языка, лингвист также опирается на интуитивное отождествление и различение объектов текста. Основные процедуры дистрибутивного анализа делают шаг вперед в описании этих интуитивных действий исследователя, они расчленяют процесс идентификации на несколько ступеней.

Методы дистрибутивного анализа с успехом применяются в практике лингвистического исследования. Вместе с тем возникает теоретический вопрос о возможности получения универсального алгоритма вывода грамматики из текстовых данных. Для этого необходимо было бы иметь формальные методы описания материала, которые обеспечили бы однозначное построение грамматики без обращения к внешним понятиям. В настоящее время становится ясным, что такая задача в принципе не разрешима. Именно в силу того, что механизмы речевой деятельности протекают в скрытом от наблюдения виде, на современном этапе невозможна полная формализация анализа языка. В зависимости от внешних, неформальных данных о тексте разные исследователи, применяя приемы дистрибутивного анализа, приходят к несколько различающимся описаниям одного и того же языка.

Ценность этих методов состоит не в том, что они обеспечивают получение полного описания языка. Такая цель едва ли оправдана в отношении большинства языков, в особенности хорошо изученных. Дистрибутивный анализ эффективно применяется для задач дешифровки, которые понимаются достаточно широко. Речь идет не только о дешифровке неизвестной письменности, например, майя, карийского письма. Любой фрагмент структуры языка может быть дешифрован дистрибутивными методами. Иначе говоря, структурные закономерности языка нуждаются в реконструкции, они не могут быть полностью обнаружены описательными приемами. Свойства сочетаемости элементов друг с другом являются основными характеристиками в их синтагматической развертке. Систематический анализ дистрибутивных признаков в связи с этим позволяет проникнуть во внутреннюю организацию структуры языка на всех его уровнях, начиная с фонологического и вплоть до синтаксического. Более того, дистрибутивные методы успешно применяются и для исследования структур плана содержания. Семантические оппозиции и семантические множители также манифестируются в дистрибутивных свойствах речевых единиц. В целом дистрибутивный анализ приводит к построению различных классификационных (таксономических) описаний языка.

**Список литературы:**

1. Засорина Л.Н. «Введение в структурную лингвистику», М: «Высшая школа – 1974 г.

2. Апресян Ю.Д. «Идеи и методы современной структурной лингвистики», М: «Просвещение» - 1966 г.

3. Степанов Ю.С. «Методы и принципы современной лингвистики», М: «Наука» – 1975 г.

4. «Общее языкознание: методы лингвистических исследований, М: «Наука» – 1973 г.