О КООРДИНАЦИИ ЯЗЫКА

2 . Начну с примера, в котором у меня есть наиболее полные данные. В /13/, стр.24 приводится сводная таблица /*tab VII*/ частот употребления букв в пяти европейских языках итальянском, французском, испанском, английском, немецком. Число букв 26. Обозначим эти языки буквами: J (Haliano), F (Francese), S (Spagnulo), E (English), T (Tedesco).

Пусть частота *i* буквы /буквы N_i/ есть P_i. Тогда информация, несомая этой буквой[1], есть $\eta_i = -\log p_i$. Однако, рассматривать эту величину неудобно. Напр., частота употребления в итальянском двадцать третьей по счету буквы *W* есть ноль $(p_{23} = 0)$. Тогда $\eta_{23} = \infty$. Оперировать с бесконечностями неудобно. Поэтому мы будем рассматривать вместо p_i величину $\overline{p_i} = 1 - p_i$, т.е. вероятность неупотребления, вероятность запрещения *i* буквы.

Обозначим величину информации об отсутствии *i* буквы ξ_i, т.ч.

$$\xi_i = -\log(1 - p_i) \qquad (2.1)$$

В частности, в итальянском языке информация об отсутствии *W* во фразе есть $\xi_{23} = -\log(1-0) = 0$, что вполне естественно: сообщая, что в таком-то итальянском слове нет буквы *W*, я ничего не сообщаю. Т.к. мы не рассматриваем однобуквенных языков, то всегда $p_i < 1$ и $0 \le \xi_i < +\infty$, но $\xi_i \ne \infty$.

Язык изображаем точкой /вектором/ с координатами $/\xi_v \xi_2, \ldots, \xi_n/$, что мы ради удобства, будем чаще писать в виде столбца.

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_n \end{pmatrix} = \begin{pmatrix} -\log(1-p_1) \\ -\log(1-p_2) \\ -\log(1-p_n) \end{pmatrix}$$
трок – равно

$$\overline{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_n \end{pmatrix} = \begin{pmatrix} -\log(1-p_1) \\ -\log(1-p_2) \\ -\log(1-p_n) \end{pmatrix} \qquad (2.2)$$

где *n* — число строк — равно числу разных букв в этих языках.

---

[1] Утверждение /I/, стр. 104 будто «выражение: ”количество информации, содержащееся в одной букве сообщения”, — бессмысленно», прямо неверно. Напр., если сообщение — фамилия, количество информации, содержащееся в букве П в фамилии Пушкин выше, чем в фамилии Пастернак, ибо при демпфировании первой буквы для « ушкин» возможны варианты: «Пушкин», «Кушкин», «Сушкин», «Тушкин», а для «астернак» — нет вариантов. Информация μ есть функция от букв $\mu = \mu(x)$, а математическое ожидание этой функции $H = \sum_x Px \cdot \mu(x)$ есть негэнтропия.

Пять разбираемых языков изобразятся пятью векторами — столбцами[2]

$$O\overline{J}=\begin{pmatrix}0{,}119\\0{,}007\\0{,}044\\0{,}040\\0{,}119\\0{,}008\\0{,}017\\0{,}009\\0{,}124\\0{,}000\\0{,}000\\0{,}068\\0{,}026\\0{,}075\\0{,}099\\0{,}029\\0{,}004\\0{,}067\\0{,}053\\0{,}072\\0{,}030\\0{,}017\\0{,}000\\0{,}000\\0{,}000\\0{,}010\end{pmatrix};O\overline{K}=\begin{pmatrix}0{,}085\\0{,}009\\0{,}031\\0{,}042\\0{,}197\\0{,}010\\0{,}010\\0{,}007\\0{,}078\\0{,}006\\0{,}000\\0{,}059\\0{,}030\\0{,}076\\0{,}054\\0{,}030\\0{,}014\\0{,}065\\0{,}084\\0{,}076\\0{,}062\\0{,}016\\0{,}000\\0{,}004\\0{,}001\\0{,}001\end{pmatrix};O\overline{S}=\begin{pmatrix}0{,}134\\0{,}014\\0{,}048\\0{,}060\\0{,}146\\0{,}007\\0{,}010\\0{,}007\\0{,}064\\0{,}004\\0{,}000\\0{,}051\\0{,}032\\0{,}069\\0{,}091\\0{,}025\\0{,}009\\0{,}071\\0{,}083\\0{,}047\\0{,}040\\0{,}009\\0{,}000\\0{,}002\\0{,}009\\0{,}005\end{pmatrix};O\overline{E}=\begin{pmatrix}0{,}085\\0{,}014\\0{,}028\\0{,}039\\0{,}140\\0{,}030\\0{,}020\\0{,}054\\0{,}066\\0{,}001\\0{,}004\\0{,}034\\0{,}026\\0{,}074\\0{,}083\\0{,}020\\0{,}001\\0{,}071\\0{,}062\\0{,}111\\0{,}025\\0{,}009\\0{,}016\\0{,}002\\0{,}020\\0{,}001\end{pmatrix};O\overline{T}\begin{pmatrix}0{,}067\\0{,}026\\0{,}027\\0{,}056\\0{,}183\\0{,}021\\0{,}037\\0{,}041\\0{,}082\\0{,}002\\0{,}019\\0{,}029\\0{,}030\\0{,}104\\0{,}023\\0{,}009\\0{,}001\\0{,}068\\0{,}070\\0{,}070\\0{,}038\\0{,}011\\0{,}014\\0{,}000\\0{,}000\\0{,}010\end{pmatrix}; \tag{2.3}$$

Здесь О — условный «язык», в котором ни одна буква не употребляется $/\forall_i p_i = 0\ /$. Видно, что координаты ξ_i близки исходным частотам, которые для облегчения сравнения помещены на обороте стр. 1. Это объясняется тем, что для малых $p_i - \log(1 - p_i) \approx p_i$. Если бы мы взяли не натуральный логарифм, а двоичный, то такое равенство не выполнялось бы.

[2] Все нижеследующие численные данные, будучи в основном иллюстрированными, приблизительными. У меня под рукой нет ни таблиц, ни даже логарифмической линейки, а арифметическими методами вычислять логарифмы и проделывать сотни требуемых умножений, утомительно. Поэтому полезно перепроверить их. Все необходимое для того формулы приводятся полностью.

Обычными методами аналитической геометрии можем определить взаимное расположение точек — языков *J, F, S, E, T.* А именно различие двух языков *X* и *У* /или, если угодно, векторная мера родства этих двух языков/ находится, по формуле:

$$\begin{pmatrix} y_2 \\ \cdots \\ y_n \end{pmatrix} - \begin{pmatrix} x \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} -x \\ y_2 - x_2 \\ \cdots \\ y_n - x_n \end{pmatrix},$$

$$\overline{XY} = \overline{OY} - \overline{OX} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \\ \cdots \\ y_n - x_n \end{pmatrix} \qquad 2.4$$

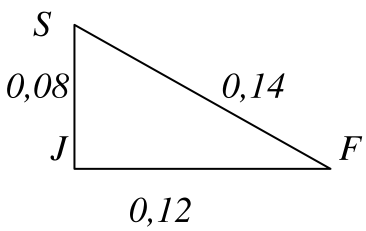
Расстояние двух языков *X* и *У* /т.е. числовая мера их родства/ находится по формуле эвклидовой геометрии, как суммарное квадратическое различие их по всем буквам:

$$XY = \sqrt{\overline{XY}^2} = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \ldots + (y_n - x_n)^2} \qquad 2.5$$

Угол между двумя языковыми различиями $\overline{XY}$ и $\overline{XZ}$ /иначе: угол между напралениями на языки *У* и *Z,* если смотреть с позиции языка *X* /определяется через скалярное произведение/ см. любой курс аналитической геометрии/:

$$\cos \underline{YXZ} = \frac{\overline{XY} \cdot \overline{XZ}}{XY \cdot XZ} = \frac{\sum_{j=1}^{n} (y_i - x_i) \cdot (z_i - x_i)}{\sqrt{\sum_{j=1}^{n} (y_i - x_i)^2} \sqrt{\sum_{j=1}^{n} (z_i - x_i)^2}} \qquad 2.6$$

Так находим, что романовские языки образуют прямоугольный треугольник:

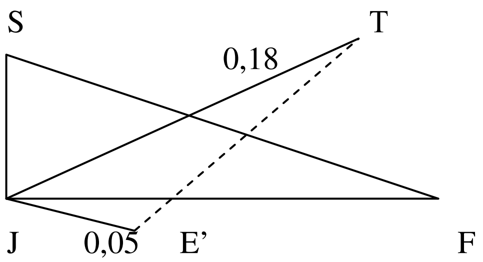


2.7

Посмотрим теперь, как располагаются германские языки относительно этого треугольника романовских языков. Если наш метод не имеет никакого отношения к «подлинной сущности языкового явления»; то может случиться, что точки, изображающие германские языки, разместятся вперемежку с точками — романовскими языками, и даже точки разных языков могли бы слиться друг с другом. Найдем проекции E' и T' точек E и T на плоскость романских языков, проекции находятся по формуле:

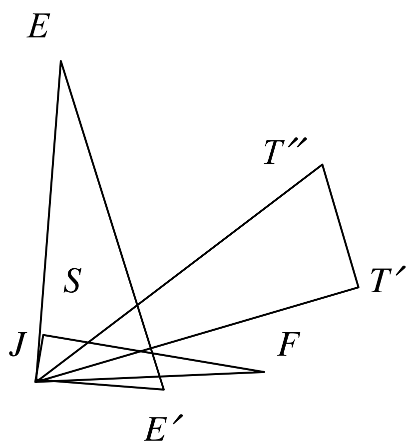
$$\overline{J}E' = JF \cdot \frac{\overline{JF} \cdot \overline{JE}}{JF^{2}} + J\overline{S} \cdot \frac{\overline{JS} \cdot \overline{JE}}{JS^{2}} \qquad 2.8$$

Расположение будет таково /углы откладываются с точностью до нескольких градусов/.



$$\angle FJE' \approx 30°; \qquad \angle EJT' \approx 15°; \qquad 2.9$$

Пунктиром обозначена проекция вектора $E\overline{T}$ (различие немецкого от английского) на плоскость романских языков. Наглядно, п*роекция это те различия двух языков* (здесь между немецким и английским), которые существенны для взаимного различения романских языков (вообще же языков, на плоскость, пространство, которых делается проекция), мы видим, что по отношению к такого рода различиям французский язык как бы лежит между английским и немецким. Но сами точки E и T и лежат на этой плоскости, т.е. английский и немецкий отличаются от романских языков еще различиями, которые несущественны для взаимного расположения романских языков. Изобразить и E и T одновременно на чертеже затруднительно, ибо потребуется четырехмерное пространство. Обозначим T-проекцию немецкого языка на трехмерное пространство, натянутое на треугольник JFS и точку B. Тогда трехмерное расположение J, F, S, E и T” изобразится чертежом (плоскость треугольника горизонтальна, а $E'E$ и $T'T''$ вертикальны):



$\angle E'JE \approx 60°$

Итак, и английский и ненецкий языки лежат существенно вне треугольника романских языков. Мы видим, что романские и германские языки геометрически лежат не вперемежку, но раздельно.

Изучим расположение точек — языков с иной точки зрения. Возьмем точку R — среднее положение романских языков по формуле: $\overline{OR} = \frac{1}{3}\left(\overline{OJ} + \overline{OS} + \overline{OF}\right)$

На чертеже удобнее всего найти как точку пересечения меридиан треугольника JFS. Затем вычислим расстояние до нее и романских языков (по 2.5)

$$\begin{cases} RJ = 0{,}05 \\ RS = 0{,}06 \\ RF = 0{,}08 \end{cases}$$

$$\begin{cases} RG = 0{,}09 \\ RE = 0{,}10 \\ RT = 0{,}11 \end{cases}$$

где

G — среднее германских языков:

$$\overline{OG} = \frac{1}{2}\left(O\overline{E} + O\overline{T}\right)$$

Видим, что германские языки лежат дальше от R, нежели романские. Если с центром в R описать четырехмерную сферу радиуса 0,08, то этот шар охватит все три романских языка, а немецкий, английский и их среднее, окажутся вне этого шара. Опять-таки видим, что романские и германские языки лежат в лингвистическом пространстве отдельными скоплениями (речь идет только о пяти изученных). Для полноты картины отметим еще, что $ET = 0{,}10$. Единица измерения у нас всюду:

$$\log 2 \frac{бит}{букву} \approx 0{,}6 \frac{вит}{букву}$$

3º. Ясно однако., что употребляемость буквы не является ни достаточно характерным признаком языка, ни, тем более, исчерпывающим признаком. Ведь одна и та же буква *j* в пяти разбираемых языках читается четырьмя разными способами, что никак не отразится в модели рубр. 2º.

Поэтому первое, что нужно сделать для уточнения модели — это выбрать подходящий «алфавит» для описания языка. Таким алфавитом, на мой взгляд, является список дифференциальных признаков. Начнем с фонологических д.п., ибо о них у меня больше сведений. /5/, /10/, /11/, /1/.

Согласно /5/ санскритские фонемы определяются девятью парами дифференциальных противопоставлений, а согласно /10/ — польские фонемы девятью частично другими парами противопоставлений. Объединяя эти списки[3], получим 13 пар противопоставлений:
Тогда фонема в санскрите и польском определяется, как 13-местная система[4] /$x_1, x_2, \ldots, x_{13}$/ где x_i принимает одно из трех значений 0, +, — согласно правилу:

Если в фонеме не используется противопоставление признака $N_{(i)}$, то $x_i = 0$;

Если в фонеме маркируется признак $N_{(i)}$, то $x_i = +$.

Если в фонеме используется, но не маркируется противопоставление признак $N_{(i)}$, то x_i = -

Список противопоставлений

1. Гласность — негласность
2. Согласность — несогласность
3. Компактность — диффузность
4. Звонкость — глухость
5. Назальность — ртовость
6. Непрерывность — прерывность

[3] Я отвлекаюсь от возможной унификации, ибо это не моя специальность. Отмечу лишь, что компактность и в /5/ стр. 52, 58 и в /10/ стр. 123 как то выпадает из единообразного метода. Кроме того не лучше ли было бы говорить «дифференцирующие признаки» вместо «дифференциальные»?

[4] Шаумян в /10/, /11/ вместо «система» говорит «пучок».

7. Придыхательность — непридыхательность
8. Переферийность — непереферийность
9. Палатальность — непалатальность
10. Низкая тональность — не низкая
11. Плавность — неплавность
12. Диезная тональность — не диезная
13. Яркость — тусклость

Примеры фонем

$$/a/\,санскрит = \begin{pmatrix} + \\ - \\ + \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};\ /a/\,польск = \begin{pmatrix} + \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};\ /a/\,санскрит = \begin{pmatrix} + \\ + \\ - \\ 0 \\ 0 \\ - \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \qquad 3.1$$

Легко видеть, что каждая фонема изображается столбцом из табл. 3 /5/. Априори возможно, что каждый из 26 признаков будет употреблен 3 раз: если он занимает свое место, то остальные 12 мест могут быть по произволу заполнены любой комбинацией из «+», «-», «О».

Но фактически в табл. 3 /5/ ни один признак не употребляется чаще трех десятков раз, а судя по /I/, стр. 134, это общее явление и ни один признак не употребляется чаще 70 раз вообще.

Рассмотрим список из 26 признаков, получаемый из 3.1

1. Гласность
2. Негласность
3. Согласность

……………………

25. Яркость
26. Тусклость

Покажем, как изобразить язык в этом 26-мерном фонологическом пространстве. Пусть i — с признак встречается в табл. 3 /5/ $N_{(i)}$ раз. Тогда априорная вероятность употребления i

признака есть $\frac{N_i}{3^{12}}$.

Как и в 2° образуем ξ_i — количество информации, содержащееся в высказывании, что i — признак отсутствует в данной фонеме взятого языка.

$$\xi_1 = -\log\left(1 - \frac{N_i}{3^{12}}\right) \qquad (3.3)$$

В силу малости $N_{(i)}$ сравнительно с 3^{12} можно написать

$$\xi_1 \approx \frac{N_i}{3^{12}}$$

а т.к. 3^{-12} входит постоянным множителем во все ξ_i, то можно просто изменить масштаб и оперировать самими $N_{(i)}$. Согласно /5/, табл. 3 санскрит изобразится точкой.

S = / 5, 28, 5, 28, 16, 14, 10, 10, 2, 20, 4, 21, 10, 10, 11, 15, 5, 18, 0, 0, 0, 0, 0, 0, 0, 0, / 3.5

У меня нет аналогичных данных по польскому языку и я не могу определить его положение P различие $\overline{PS}$ расстояние PS и т.п.

Более того, мы не учли, что фонемы сами входят в язык с известными частотами. Подсчитывая $N_{(i)}$ по табл. 3 /5/ мы приняли все фонемы (столбцы) равновероятными. Чтобы правильно подсчитать $N_{(i)}$, надо умножить каждый k-столбец таблицы на P_k, где P_k есть частота (вероятность) употребления фонемы N_k, а затем сложить построчно, полученное положительное число дает N_{2i-1} (частоту употребления маркированного признака, нечетного в списке 3.2), а отрицательное — дает N_{2i} частоту немаркированного, четного в 3.2). У меня нет этих данных.

Чтобы можно было применить эту теорию к нескольким языкам, надо, прежде всего, согласиться насчет единого для всех изучаемых языков перечня фонологических д.п. Если число противопоставлений в этом перечне n, то каждый язык из числа рассматриваемых изобразится точкой в $2n$-мерном фонологическом пространстве. Расположение этих точек изучается далее методами рубл. 2°.

Разумеется, что фонологическое описание не исчерпывает язык. Надо продолжить список признаков, включив в него морфологические д.п._, семантические множители и вообще все характеристикой языка. Тут мы отказывается от господствующей идеи расчленения языка на «уровни» и замкнутого в каждом-«уровне» изучения. Напротив, мы связываем все эти «уровни» в единый комплекс (вектор) вида:

$$\underbrace{x_1,\ldots,x_k}_{\text{фонологические}},\ \underbrace{x_{k+1},\ldots,x_{k+1}}_{\text{морфологические}},\underbrace{x_{k+l+1},\ldots,x_{k+b+m}}_{\text{семантические}}, \qquad 3.6$$

$$\underbrace{x_1,\ldots,x_k}_{\text{фонологические}},\ \underbrace{x_{k+1},\ldots,x_{k+l}}_{\text{морфологические}},\ \underbrace{x_{k+l+1},\ldots,x_{k+l+m}}_{\text{семантические}}$$

Но поскольку у меня нет никаких списков м д.п. и семантических множителей, постольку приходится ограничиться (заведомо неполным) фонологическим изучением. Ближайшей перспективой может явиться изучение фоно-морфологического вектора.

4°. Теперь опишем абстрактно геометрическую модель. С абстрактной точки зрения речь идет об определении лингвистического положения текста (или лучше носителя текста), по отношению к другим текстам, применяется классификационно-количественный метод.

Ищутся чисто формальные признаки, которыми и только которыми различаются между собой все рассматриваемые тексты и только они. Для этого задается n алгоритмов, применимых ко всякому рассматриваемому тексту, и дающих либо ответ «да», либо ответ «нет» (такой-то признак применяется — не применяется). Эти алгоритмы определяют «алфавиты» рассматриваемых текстов:

$\alpha_1, \alpha_2, \ldots, \alpha_n$

В 2° за таком алфавит мы приняли английский письменный алфавит, в 3° — перечень фонологических дифференциальных признаков. Вообще говоря, для полного определения лингвоположения текста требуется алфавит, содержащий фоно-морфологические д.п., семантические множители и возможно иные буквы (ударения, интонация, — если их нельзя отнести к фонологии, и др.). 'Гак происходит классификация.

Количественная сторона заключается в том, что для каждого алгоритма подсчитывается фактическое число ответов «да» на данном тексте, и делится на априорию возможное число таких ответов. Так находится вероятность — P_i употребления буквы α_i.

Положение текста X изображаем вектором $\overline{OX}$:

$$\bar{\xi} = \overline{OX} = \begin{pmatrix} -\log(1-p_1) \\ -\log(1-p_2) \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ -\log(1-p_n) \end{pmatrix} \underset{dfn}{\equiv} \left(-\log(1-p_i)\right)$$

3°. O — начало отсчета, здесь принят «нулевой текст», т.е. текст, в котором ни одна букве, не употребляется. Можно за O выбрать произвольный текст (произвольное начало отсчета). Тогда вместо 4.2 определим вектор лингвоположения X относительно O так. Пусть P_i^X есть вероятность употребления α_i в тексте X. Пусть P_i^O — вероятность употребления α_i в тексте O. Тогда вычислим *ценности информации* о запрещении α_i в X относительно O по формуле Харкевича /9/, формула /2/ и составили $\overline{OX}$ как вектор ценности информации

$$O\overline{X} = \begin{pmatrix} -\log\dfrac{1-P_1^X}{1-P_1^O} \\ \dots\dots\dots\dots\dots \\ -\log\dfrac{1-P_n^X}{1-P_n^O} \end{pmatrix} \qquad 4.2$$

Осмыслим лингвистическое различие двух текстов X и Y определяемое аналитически формулой:

$$\overline{XY} = \overline{OY} - OX \qquad 4.3$$

Согласно 4.1, 4.3 по свойству логарифма

$$\overline{XY} = \left(-\log\frac{1-P_i^Y}{1-P_i^X}\right) \qquad 4.4$$

Отношение $\dfrac{1-P_i^Y}{1-P_i^X} = \dfrac{P_i^Y}{P_i^X}$ выражает отношение числа кодовых запрещений буквы α_i, в

тексте.

Y к числу кодовых ограничений той же буквы в тексте X. Это т.ск., сомнительное запрещение α_I в Y и X.

$A - \log\dfrac{1-P_i^Y}{1-P_i^X}$ есть то, что Харкевич /9/ называет ценностью информации (о

запрещении или отсутствии буквы α_i). Удобнее просто говорить о приращении информации при переходе от X и Y.

Мы видим, что 4.2 вполне согласуется с 4.4. Итак вектор лингворазличия двух текстов (математическое уточнение меры лингвистического родства) выражает (логарифм) отношения кодовых запрещений в этих текстах.

Прибавить к тексту X лингворазличие $\overline{OA}$ значит найти такой текст — Y, который так *относится* (по числу запрещений на каждую букву) к X, как A к O (т.е. лингворазличие $\overline{XY} = \overline{OA}$. Т.о. определению:

$$\frac{1-P_i^Y}{1-P_i^X} = \frac{1-P_i^A}{1-P_i^O}, \qquad 4.5$$

т.е.

$$\log\frac{1-P_i^Y}{1-P_i^O}=\log\frac{1-P_i^Y}{1-P_i^O}+\log\frac{1-P_i^A}{1-P_i^O} \qquad 4.6$$

Умножение лингворазличия на число λ определяется как возведение отношения запрещении в степень λ:

$$\overline{OY}=\lambda\cdot\overline{OX}$$

равносильно

$$\frac{1-P_i^Y}{1-P_i^O}=\left(\frac{1-P_i^X}{1-P_i^O}\right)^{\lambda} \qquad 4.7$$

Другое эквивалентное определение получается так. Умножить лингворазличие на 2 значит сложить его (по 4.6) с самим собой:

$$2\cdot\overline{OX}=\overline{OX}+\overline{OX}$$

Повторяя, приходим к определению умножения на любое натуральное число, а далее приемом известным из IX класса, к определению умножения на любое положительное число. Умножение на - 1 есть просто изменение направления сравнения на противоположное: вместо «различие английского от вьетнамского» рассматриваем «различие вьетнамского от английского».

Расстояние между двумя текстами X и Y в согласии с 2.5 находится как суммарное квадратичное различие текстов по всем буквам:

$$XY=\sqrt{\sum_{i=1}^{n}\left(P_i^Y-P_i^X\right)^2} \qquad 4.9$$

Снова подчеркну, что указанными методами мы не проникаем в суть языковых закономерностей», не даем средств к построению машины-переводчика, не получаем ответа на вопрос: «что есть язык». Мы просто точно определяем положение языка (текста) среди других языков (текстов). Этим решается задача координации языков. Теперь можно точно определить лингвоположение текста среди других текстов. На этой основе можно точно определить, какие фигуры образуют семейства языков, действительно ли они разделяются и в какой мере.

5°. В примерах /2°—3°/ мы оперировали буквами, фонемами. Чтобы трактовать таблицу их употребления, как в /13/, Ш или ы /5/) табл. 5, но дополненную указаниями на частоты употребления этих сочетаний. Возможна двоякая трактовка.

А. Рассматриваем таблицу $\left(P_{ik}\right)=\begin{pmatrix} P_{ii}\ldots P_{im} \\ \ldots\ldots\ldots\ldots \\ P_{m_1}\ldots P_{mm} \end{pmatrix}$

где P_{ik} есть вероятность употребления сочетания α_i a_k. Рассматриваем, как в 3.1, системы, определяющие фонемы (морфемы, самы) $a_k=\begin{pmatrix} x_{k_1} \\ \ldots.. \\ x_{k_n} \end{pmatrix}$. Здесь n — число диф. признаков; m — число фонем, морфем и т.п. первичных сочетаний д.п.

Подсчитываем число употреблений «+» и «-» для каждой x_{kl}, руководствуясь $\left(P_{ik}\right)$. Полученные числа определяют положение текста, трактуемого в виде диаграмм, в $2n$-мерном пространстве диф. признаков.

Я не могу привести примера, ибо в /5/ табл. 5 обозначения таблицы не приведены в соответствие с текстом и не указаны частоты.

Так же поступаем для приграмм, квадриграмм и вообще полиграмм.

В. Пусть можно остановиться на l-грамме. Точнее, пусть можно расчленить текст на сегменты /1/, стр. 50 и пусть длина максимального сегмента есть l. Тогда рассматриваем $2n\cdot l$-мерное пространство, где n — число противопоставлений признаков. Проще, составляем перечень признаков (вместо 3.2).

1 : 1 Гласность на первом месте в сегменте

1.2 Гласность на втором месте

- -

I. Гласность на l— месте 5.1

2.1 Негласность на первом месте

- -

2. Тусклость на l-месте