
**IV Международный конгресс
исследователей русского языка**

**РУССКИЙ ЯЗЫК:
исторические судьбы
и современность**

**ТРУДЫ
и МАТЕРИАЛЫ**



**Москва, МГУ
20–23 марта 2010 г.**



Русский язык: исторические судьбы и современность

**IV Международный конгресс
исследователей русского языка**

*Москва, МГУ имени М. В. Ломоносова,
филологический факультет*

20-23 марта 2010 года

Труды и материалы

Lomonosov Moscow State University (MSU)
Faculty of Philology

Russian Language: Its Historical Destiny and Present State

**The Fourth International Congress
of Russian Language Researchers**

Moscow, Lomonosov Moscow State University,
FACULTY OF PHILOLOGY

March 20-23, 2010

Proceedings and materials

Collected by

Marina L. Remneva, Anatoliy A. Polikarpov

Moscow University Press

2010

Московский государственный университет имени М. В. Ломоносова
Филологический факультет

Русский язык: исторические судьбы и современность

**IV Международный конгресс
исследователей русского языка**

Москва, МГУ имени М. В. Ломоносова,
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

20-23 марта 2010 года

Труды и материалы

Составители

М. Л. Ремнёва, А. А. Поликарпов

Издательство Московского университета

2010

в общем случае достаточно сильно перекрывающиеся для разных авторов (или разных жанров). При сильном перекрытии классов задача различения классов (стилей текстов) в многомерном пространстве признаков стилей является достаточно трудной. Дискриминантный анализ позволяет максимально разнести классы друг относительно друга.

4. Пусть $p = \{p_{ij}\}$ — $n \times m$ -матрица относительных частот появления j -го признака в i -м тексте ($i=1, n, j=1, m$, где n — число текстов, m — число признаков). Проранжировав в порядке возрастания величины $\{p_{ij}\}$ по всем текстам (от 1-го до n -го) для каждого j -го признака, получим матрицу рангов $r = \{r_{ij}\}$ признаков (мест признаков среди текстов). При этом рангам совпадающих частот, образующих так называемые связки, припишем средний по связке ранг. Разделив ранги на $n + 1$, приведем их к интервалу $[1 / (n + 1), n / (n + 1)]$. В результате получим матрицу относительных рангов $\{r_{ij} / (n + 1)\}$. Их эмпирическое распределение вероятностей равномерно в единичном интервале. Сопоставим каждому относительному рангу квантиль стандартного нормального распределения уровня этого относительного ранга. В результате такого нелинейного преобразования получим матрицу нормально распределенных величин (нормализованных относительных рангов — НОР) $x = \{x_{ij}\}$ с нулевыми средними и единичными дисперсиями для каждого j -го столбца, причем столбцы будут коррелированы (в общем случае) между собой.

5. Пусть имеется g классов. Вычислив положение центров классов в признаковом пространстве НОР (средние значения координат точек каждого k -го класса), можно подвергнуть оси координат такому линейному преобразованию $y = xV$ (повороту и масштабированию осей), при котором расстояния между центрами классов по отношению к диаметрам классов в новом (дискриминантном) признаковом пространстве станут наибольшими. Дискриминантный анализ предписывает выбирать матрицу V коэффициентов этого преобразования так, чтобы максимизировать отношение

$\lambda_l = (V'BV)_{ll} / (V'WV)_{ll}, l = 1, 2, \dots, q, q = \min(m, g - 1)$. Здесь штрих — знак транспонирования, $B = T - W, T$ — $m \times m$ -матрица ковариаций векторов-столбцов матрицы x, W — $m \times m$ -матрица внутригрупповых ковариаций векторов-столбцов матрицы x . Из этого критерия оптимизации следует [1]; [2], что m -векторы-столбцы $\{V_l\}$ матрицы V — собственные векторы, соответствующие матрицам B и W и удовлетворяющие уравнению $BV_l = \lambda_l WV_l$, а $\{\lambda_l > 0\}$ — q их первых (в порядке убывания) собственных значений, удовлетворяющих характеристическому уравнению $\det(B - \lambda W) = 0$. Столбцы матрицы $y = \{y_{il}\}$ называются дискриминантными функциями и образуют новое признаковое пространство размерности q (пространство новых факторов), в котором обеспечивается наилучшее разделение (дискриминация) классов. Столбцы матрицы коэффициентов $V = \{V_{jl}\}$ называются факторными нагрузками. При соответствующей нормировке они являются коэффициентами корреляции между каждым новым l -м (дискриминантным) признаком и каждым «старым» j -м признаком пространства нормализованных относительных рангов. Содержательная интерпретация дискриминантных функций (новых, дискриминантных признаков) определяется наборами старых признаков, в наибольшей степени коррелирующими с новыми признаками. Статистическая значимость оставляемой в новом признаковом пространстве l -й дискриминантной функции при гауссовом распределении нормализованных относительных рангов рассчитывается по χ^2 -распределению с $v_l = (m - l)(g - l - 1)$ степенями свободы, которому при верной нулевой гипотезе подчиняется величина, пропорциональная логарифму Λ -статистики Уилкса [2].

6. В качестве примера приводятся результаты дискриминантного анализа 81 текста крупных произведений художественной прозы (романов, повестей) 12 русских писателей XIX века с использованием в качестве признаков стилей 55 служебных слов. На рис. 1 видно отличное разделение текстов по писателям в пространстве первых двух дискриминантных функций.

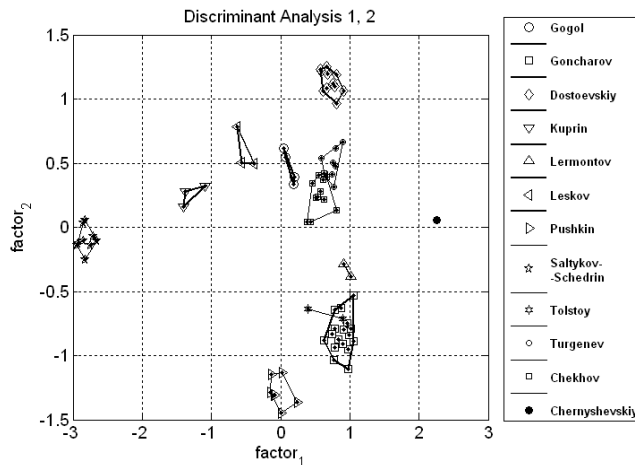


Рис. 1. Дискриминантный анализ текстов русских писателей XIX века по 55 служебным словам.

Литература

1. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды / Пер. с англ. М., 1976.

2. Ким Дж.-О., Мьюллер Ч. У., Клекка У. Р. и др. Факторный, дискриминантный и кластерный анализ / Под ред. И. С. Енюкова; пер. с англ. М., 1989.

Математическое моделирование жизненного цикла языкового знака

В. В. Поддубный

Томский государственный университет
pvv@inet.tsu.ru

А. А. Поликарпов

Московский государственный университет имени М. В. Ломоносова
anatolp@philol.msu.ru

Языковой знак, жизненный цикл, полисемия, математическая модель, диссипативный процесс

Summary. The dissipative nonstationary dynamic mathematical model of the life cycle of the language sign is offered. This cycle is based on the interaction of the processes of the sign polysemy growing and of the losing of the earlier gained sign meanings. It is shown that this model satisfies to the variational principle of the least action. The model is presented in continuous and discrete variants. The stochastic expansion of the discrete variant of the model is built. The numerical modeling of the process of the language sign polysemy is made.

1. Жизненный цикл языкового знака от момента его зарождения до момента выхода из употребления связан с взаимодействием двух процессов его развития: процесса роста полисемии знака, приобретения знаком новых, как правило,

модействием двух процессов его развития: процесса роста полисемии знака, приобретения знаком новых, как правило,

все более абстрактных значений, и процесса постепенного выхода из употребления ранее приобретенных значений, начиная с наименее абстрактных. Второй процесс начинается с некоторым запаздыванием по отношению к первому и протекает более медленно. Разность между количеством приобретенных знаков значений и количеством значений, вышедших из употребления к данному моменту времени, составляет размер активной полисемии знака, т. е. количество живущих в этот момент времени значений знака. Максимальное количество возможных значений знака назовем его ассоциативно-семантическим потенциалом (АСП) [3].

2. Естественно предположить, что скорость роста полисемии знака в каждый текущий момент времени пропорциональна запасу потенциала полисемии этого знака, т. е. разности между АСП и количеством значений, порожденных знаком к этому моменту времени. Это предположение в непрерывной модели развития полисемии приводит к дифференциальному уравнению роста полисемии. К моменту рождения знака запас полисемии максимален и равен АСП. Коэффициент пропорциональности при этом может зависеть от времени, возможно, монотонно убывая, оставляя постоянным или монотонно возрастающим. Но всегда этот коэффициент положителен: скорость роста полисемии тем больше, чем больше запас потенциала полисемии. При этом, очевидно, с ростом полисемии скорость роста падает. Процессы с такими свойствами, описываемые дифференциальными уравнениями первого порядка с правой частью, зависящей от состояния процесса и, возможно, времени, и имеющей отрицательную частную производную по состоянию, называются диссипативными [2]; [4]. В случае независимости коэффициента от времени диссипативную динамическую систему, описываемую таким дифференциальным уравнением, называют стационарной [2], в противном случае — нестационарной [4]. Естественно предположить, что и процесс потери значений знака (процесс выхода значений из употребления) является таким же диссипативным динамическим процессом, как и процесс роста полисемии, но с другим (меньшим) коэффициентом пропорциональности, обеспечивающим более медленное его развитие. Эти предположения определяют математическую диссипативную динамическую модель процессов роста и потерь полисемии, а, следовательно, и математическую модель жизненного цикла языкового знака. Из этой модели математически строго следует, что жизненный цикл языкового знака, начинаясь с единственного исходного значения, со временем проходит стадию роста полисемии, а затем, достигнув пика своего развития, утрачивает все большее число значений до выхода из употребления последнего значения и знака в целом. Таким образом, жизненный цикл языкового знака во времени представляется унимодальной (с одним максимумом) асимметричной кривой с положительной асимметрией (смещением пика в сторону начала процесса), спадающую затем до нуля (выхода знака из употребления).

3. Математически строго показано, что жизненный цикл знака, описываемый математической моделью диссипативного динамического процесса (в общем случае нестационарного), подчиняется вариационному принципу наименьшего действия [1]; [5], лежащему в основе формулировки многих законов природы. Применительно к процессу развития полисемии языкового знака принцип наименьшего действия, обосновывающий его диссипативную математическую модель, можно сформулировать так: из всех возмож-

ных кривых активной полисемии знака в природе реализуется кривая, минимизирующая интеграл, называемый функционалом действия. Функционал действия — это число, выражающее суммарную на траектории жизненного цикла знака разность между «кинетической энергией» полисемии и ее «потенциальной функцией» [1]; [5]. По определению, «кинетическая энергия» пропорциональна квадрату скорости изменения полисемии, «потенциальная функция» — квадрату запаса ассоциативно-семантического потенциала полисемии. Из решения этой вариационной задачи математически строго вытекают наши естественные предположения о диссипативной динамической математической модели жизненного цикла знака.

4. В дискретном варианте диссипативной математической динамической модели жизненного цикла знака условие пропорциональности скорости роста полисемии запасу его ассоциативно-семантического потенциала выражается не дифференциальными, а разностными уравнениями, определяющими моменты рождения и выхода из употребления значений знака. Интервал времени между рождением текущего и следующего за ним значения знака обратно пропорционален текущему запасу потенциала полисемии. То же — для выхода из употребления значений знака, но со своими параметрами. Моменты рождения и выхода из употребления любого значения языкового знака определяют длительность его жизни. Показано, что в рассматриваемой модели длительность жизни значения монотонно возрастает с его номером (уровнем абстрактности), достигая максимума для самого высокого уровня. С выходом из употребления этого значения жизненный цикл знака заканчивается.

5. Проведено расширение дискретной диссипативной динамической модели развития полисемии языкового знака на стохастический случай, когда коэффициенты детерминированной модели становятся случайными неотрицательными функциями времени, обеспечивающими ту же самую последовательность рождения новых и потери ранее приобретенных значений знака, но в случайные моменты времени. Структура такой стохастической модели остается диссипативной и удовлетворяющей принципу наименьшего действия. Эта модель включает в себя мультипликативный случайный фактор, приводящий к случайным колебаниям моментов последовательного рождения и выхода из употребления значений знака. Детерминированная модель определяет медиану стохастического процесса развития полисемии знака, описываемого стохастической моделью.

6. Проведено компьютерное моделирование детерминированных и случайных процессов развития полисемии знака в соответствии с предложенной моделью.

Литература

1. Айзерман М. А. Классическая механика. М., 1980.
2. Лоскутов А. Ю., Михайлов А. С. Основы теории сложных систем. М.; Ижевск, 2007.
3. Поликарпов А. А. Системно-количественный подход в лингвистике // Филологические школы и их роль в систематизации научных исследований. Вестник Смоленского государственного университета. Сер. 1: Филология. Т. 1. Смоленск, 2007. С. 35–59.
4. Шамолин М. В. Динамические системы с переменной диссипацией: подходы, методы, приложения // Фундаментальная и прикладная математика. Т. 14. № 3. М., 2008. С. 3–23.
5. Эльсгольц Л. Э. Дифференциальные уравнения и вариационное исчисление. М., 1965.

Обнаружение события в русских художественных текстах

С. Б. Потёмкин

Московский государственный университет имени М. В. Ломоносова

potemkin@philol.msu.ru

Summary. Techniques for determining the point of event in a fiction text is proposed. Formal detection of events in the discourse is based on text only without its perception by a reader. Our hypothesis is as follows: descriptions of states are differed by numerous pairs of antonyms of adjectives or adverbs. This paper involves examination of this hypothesis and the results obtained.

Проблема выявления событий в тексте. Событие, элементарная составляющая повествовательного текста, было определено Ю. М. Лотманом как «перемещение персонажа через границу семантического поля». Таким образом, событие заключается в некоем отклонении от законного, нормативного в данном мире, в нарушении одного из тех правил,

соблюдение которых сохраняет порядок и устройство этого мира. Определение события как переход между последовательными во времени или причинно-следственно связанными ситуациями покрывает практически все многообразие изменений в любом произведении. Полноценная событийность в нарративном тексте подразумевает выпол-