



Русский язык: исторические судьбы и современность

**V Международный конгресс
исследователей русского языка**

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

18–21 марта 2014 года

Труды и материалы

Lomonosov Moscow State University (MSU)
Faculty of Philology

Russian Language: Its Historical Destiny and Present State

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

Moscow, Lomonosov Moscow State University,
FACULTY OF PHILOLOGY

March 18–21, 2014

Proceedings and materials

Collected by

Marina L. Remneva, Anatoliy A. Polikarpov, Olga V. Kukushkina

Moscow University Press

2014

Русский язык: исторические судьбы и современность

**V Международный конгресс
исследователей русского языка**

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

18–21 марта 2014 года

Труды и материалы

Составители

М. Л. Ремнёва, А. А. Поликарпов, О. В. Кукушкина

Издательство Московского университета

2014

УДК 82
ББК 80/82
Р 89

*Печатается по постановлению Редакционно-издательского совета
филологического факультета МГУ имени М. В. Ломоносова*

МАТЕРИАЛЫ ПУБЛИКУЮТСЯ В АВТОРСКОЙ РЕДАКЦИИ

Р 89 Русский язык: исторические судьбы и современность: V Международный конгресс исследователей русского языка (Москва, МГУ имени М. В. Ломоносова, филологический факультет, 18–21 марта 2014 г.): Труды и материалы / Составители *М. Л. Ремнёва, А. А. Поликарпов, О. В. Кукушкина*. — М.: Изд-во Моск. ун-та, 2014. — 848 с.

ISBN 978-5-19-010920-7

Научное издание

Русский язык: исторические судьбы и современность

V Международный конгресс исследователей русского языка
(Москва, МГУ имени М. В. Ломоносова, филологический факультет, 18–21 марта 2014 г.)

Труды и материалы

Составители

М. Л. Ремнёва, А. А. Поликарпов, О. В. Кукушкина

Зав. редакционно-издательским отделом филологического факультета

Е. Г. Домогацкая

edit@philol.msu.ru

Оригинал-макет

Л. М. Захаров

Подписано в печать 06.03.2014. Формат 60х90/8. Гарнитура Таймс.

Издательство Московского университета.

125009, Москва, ул. Б. Никитская, 5.

Тел.: +7 (495) 629-50-91. E-mail: secretary-msu-press@yandex.ru

Сайт Издательства МГУ: www.msu.ru/depts/MSUPubl2005

Как видим, в «верхушке» лексико-семантического ядра немало общего. Лексема со значением 'бить' объединяет все восточнославянские языки, 'дать' – русский и украинский языки, 'идти' – украинский и белорусский.

Поскольку лексическая семантика более устойчива, чем лексическая форма даже в родственных языках, сравнивать близость полученных ядер можно, ориентируясь на значения (семемы). В двуязычных словарях языком дефиниций является русский, что также облегчает сравнение.

Таким образом, количество совпадений по дефинициям для всех трех языков составляет 55 изосем: *дать, бить, дом, мир, быть, идти, верх, знать, ужас, поле, бой, лечь, рвать, лезть, конец, тянуть, сойти, добро, сказать, работа, слабый, плохой, маленький, дух, род, путь, ухо, есть, лад, тень, ряд, бес, круг, сила, лицо, свет, стать, горе, беда, чудо, удар, драить, брать, иметь, взять, пить, речь, грех, гроб, горб, куча, след, друг, дыра, баба*.

Для определения степени близости лексико-семантических систем пар языков можно воспользоваться следующей формулой: $R = C / (A+B) / 2$, где R – индекс связи, C – общее количество лексем, A – количество лексем в первом ядре, B – количество лексем во втором ядре.

Результаты таковы:
 $117 / (229 + 554) / 2 = 0,299$ (индекс связи пары русский – белорусский);

$72 / (229+302) / 2 = 0,271$ (индекс связи пары русский – украинский),
 $120 / (302+554) / 2 = 0,280$ (индекс связи пары украинский – белорусский).

Литература

1. Кретов А. А., Меркулова И. А., Тутов В. Т. Проблемы количественной лексикологии славянских языков // Вопросы языкознания. 2011. № 1. С. 52–65.
2. Тутов В. Т. Общая количественная лексикология романских языков. Воронеж, 2002.

Список источников

1. Белорусско-русский словарь: ок. 90000 слов / Под ред. К. К. Крапивы. М., 1962.
2. Загітко А. П., Віттонів М. О., Данилюк І. Г. Великий українсько-російський і російсько-український словник. Донецьк, 2007.
3. Клышка М. К. Слоўнік сінонімаў і блізказначных слоў / Аўт.-склад. М. К. Клышка. Мн., 2005.
4. Лепешаў І. Я. Слоўнік фразеалагізмаў: У 2 т. Мінск, 2008.
5. Ожегов С. И. Словарь русского языка [Электронная версия, используемая в компьютерной системе СтарЛинг С. А. Старостина].
7. Словарь синонимов русского языка: В 2 т. / Под ред. А. П. Евгеньевой. М., 2001.
6. Словник синонімів української мови. К., 1999–2000.
8. Фразеологічний словник Української Мови: В двох томах / В. М. Білоноженко. К., 1993.

Непрерывная диссипативная стохастическая модель развития полисемии языковых знаков и вывод на ее основе законов полисемического и возрастного-полисемического распределений знаковых ансамблей*

В. В. Поддубный, А. А. Поликарпов

Томский государственный университет, Московский государственный университет имени М. В. Ломоносова
 vvpoddubny@gmail.com, anatpoli@mail.ru

Вывод законов полисемического и возрастного распределений знаковых ансамблей.

Summary. On the basis of the proposed continuous dissipative stochastic model of signs' polysemy evolution a derivation of laws for polysemic and age-polysemic distributions of units in some natural language sign ensemble is undertaken.

Предлагаемая математическая модель основывается на предположении о диссипативном характере развития полисемии языкового знака [1]. Это предположение состоит в том, что каждый языковой знак в момент своего рождения обладает индивидуальным пределом G возможной многозначности, называемым ассоциативно-семантическим потенциалом (АСП), который постепенно реализуется-растрачивается в процессе развития полисемии знака. При этом скорость образования у знака новых значений (как правило, все более абстрактных, чем позже они образуются) принимается пропорциональной еще неизрасходованной части АСП, вследствие чего прирост новых значений постепенно замедляется. Одновременно, но с запаздыванием на время τ_0 , начинает протекать аналогичный и значительно более медленный процесс выпадения из употребления первоначальных значений знака (котгорые, как правило, относительно более конкретны и поэтому – относительно менее устойчивы). Актуальная полисемия знака в любой момент времени t его жизненного цикла выражается разностью $x(t) = x_1(t) - x_2(t)$ процессов прибавки новых значений $x_1(t)$ и потери знаком своих значений $x_2(t)$. Непрерывная модель предполагает, что эти процессы непрерывны и подчиняются линейным дифференциальным уравнениям вида:

$$x_1(t)/dt = (G - x_1(t))/\tau_1,$$

$$x_1(t \leq 0) = 0, dx_2(t)/dt = (G - x_2(t))/\tau_2, x_2(t - \tau_0 \leq 0) = 0,$$

где $\tau_1 = a_1/G$, $\tau_2 = a_2/G$ – обратные пропорциональные АСП постоянные времени роста и спада полисемии соответственно, причем $\tau_1 \ll \tau_2$.

Модель предполагает далее, что в знаковом ансамбле значения G и запаздывания τ_0 – независимые экспоненци-

ально распределенные случайные величины с параметрами (математическими ожиданиями) $\langle G \rangle$ и $\langle \tau_0 \rangle$ соответственно. Поток рождения знаков предполагается стационарным пуассоновским с интенсивностью $\lambda = 1/\langle \tau \rangle$, так что интервалы времени между появлениями в потоке соседних знаков предполагаются независимыми экспоненциально распределенными случайными величинами с параметром (математическим ожиданием) $\langle \tau \rangle$. Величина $\langle \tau \rangle$ принимается в модели за единицу времени. Моменты появления знаков в стационарном пуассоновском потоке имеют равномерное распределение. Таким образом, предлагаемая модель характеризуется четырьмя параметрами: $\langle G \rangle$, $\langle \tau_0 \rangle$, a_1 , a_2 .

Индивидуальная кривая развития полисемии $x(t, G, \tau_0)$ во времени t (кривая жизненного цикла знака), зависящая также от G и τ_0 , получается из решения приведенных выше дифференциальных уравнений. Условная плотность распределения полисемии x индивидуального знака в зависимости от времени (возраста знака) t , АСП G и запаздывания τ_0 представляется дельта-функцией Дирака (Dirac delta-function) $p(x|t, G, \tau_0) = \delta(x - x(t, G, \tau_0))$.

Усреднением этой δ -образной условной плотности по распределениям запаздывания τ_0 и G получаем условную плотность $p(x|t)$ распределения полисемии x ансамбля знаков одного возраста t . Последующим усреднением по равномерному распределению t получаем безусловную плотность $p(x)$ распределения полисемии знакового ансамбля безотносительно к возрасту знаков.

Сравнивая это распределение с эмпирическими распределениями полисемии знаков различных естественных языков, полученными из представительных толковых словарей

* Настоящее исследование выполнено при поддержке Российского фонда фундаментальных исследований (грант №11-07-00776-а по проекту «Комплексная тексто-аналитическая система „СтилеАнализатор-2“, основанная на Web-технологиях: разработка, наполнение данными и тестирование на прикладных задачах», и грант по проекту «Системные закономерности эволюции Языка Человека: онтологические основания, построение математической модели и ее эмпирическая проверка».

Литература

1. Poddubny V. V., Polikarpov A. A. Stochastic Dynamic Model of Evolution of Language Sign Ensembles // Methods and Applications of Quantitative Linguistics. Selected papers of the 8th International Conference on Quantitative Linguistics (QUALICO) / Ed. by Ivan Obradović, Emmerich Kelih and Reinhard Kohler. Belgrad, 2013. P. 69–83.

этих языков, можно определить параметры модели $\langle G \rangle$, $\langle \tau_0 \rangle$, a_1 , a_2 для каждого языка.

Релевантность модели проверялась на данных толковых, этимологических и исторических словарей русского и английского языков. Рассматриваются и объясняются случаи кажущегося расхождения теоретических прогнозов с эмпирическими данными.

Теоретические основы нового подхода к реконструкции смыслового контента зашумленной русской речи

Р. К. Потапова, В. В. Потапов

Московский государственный лингвистический университет,
Московский государственный университет имени М. В. Ломоносова

rkpotaпова@yandex.ru, kedr@philol.msu.ru

Акустический шум, перцептивно-слуховое восприятие, смысловое декодирование речи.

Summary. This paper presents an experimental study of auditory-perceptual recognition of Russian connected speech in acoustic noise. The aim of this project is development of methods for generating important linguistic information, which provides for a wide range of Russian speech processing functions.

Влияние акустических шумов и помех на восприятие устной речи изучалось достаточно глубоко в течение ряда десятилетий прошлого века как у нас в стране, так и за рубежом, что было связано в первую очередь с определением качественных характеристик тракта, передающего речевой сигнал.

Данное направление междисциплинарного характера развивалось прежде всего в вокодерной телефонии, основным предметом которой являлся компандированный речевой сигнал, реализуемый на конкретном естественном языке, в частности, русском.

Суть исследований заключалась в специфике действия процесса маскировки акустического сигнала, которая определяется изменением порога слышимости устной речи по сравнению с ее восприятием в условиях отсутствия шумов, помех и искажений. Если проследить динамику развития исследований в данной области применительно к русской речи, то можно прийти к выводу, что первоначально основная задача заключалась в оценке качества самих передающих речь трактов (см. работы М. А. Сапожкова, А. А. Пирогова и др.). В связи с этим развивались исследования, направленные во второй половине XX в. на решение задачи определения критериев разборчивости русской речи в условиях шума. С помощью лингвистов были разработаны специальные таблицы разборчивости русской речи (см. напр., работы В. Г. Михайлова). Намного позже появились попытки исследования на фоне шумов и помех определения самого языка, на котором реализуется речевое сообщение, и далее идентификации говорящего, передающего сообщение в затрудненных для восприятия условиях. Естественно, что две вышеназванные задачи оказались намного сложнее, чем задача разборчивости речи.

В начале XXI века все более интенсивно развиваются исследования, соотносящиеся не только с вышеупомянутыми задачами, но и с задачами возможности декодирования смыслового контента устного речевого сообщения, в частности, в условиях шума с учетом различных градаций «понятности» речевой передачи, не ограничиваясь при этом только лишь слоговой и словесной степенью разборчивости.

Задача подобного рода решалась нами на протяжении ряда лет в исследовании особенностей перцептивно-слухового восприятия и реконструкции смыслового контента в условиях белого шума звучащей русской речи, распределенной по видам речевой деятельности: чтения, монолога, диалога, полилога. Исследование проводилось поэтапно с учетом специфики решаемых задач:

- формирование специального корпуса экспериментальных текстов, предназначенных для чтения и пересказа;
- формирование корпуса тем, предлагаемых испытуемым в целях их использования в режиме диалога, полилога;
- проведение записи носителей русского языка в безэховой камере;

- монтаж полученных фонограмм с наложением белого шума по схеме «сигнал – шум» в режиме 0 дБ, 10 дБ, 20 дБ;
- проведение по специальной программе перцептивно-слухового анализа с привлечением auditors-носителей русского языка;
- создание базы данных реляционного типа с помощью Microsoft Excel.

Полученные данные распределялись с учетом не только типов градации, но и видов речевой деятельности.

При подобном подходе к решению задачи необходимо было определить не только число распознанных лексических единиц, их принадлежность к тому или иному классу речи (об этом см. наши более ранние работы с участием М. В. Хитиной), но также и поэтапность декодирования по распознанным фрагментам темы зашумленного текста.

Дальнейшая задача включала восстановление (своего рода реконструкцию) общего смыслового контента воспринятого в затрудненных условиях материала, что было связано с целым рядом сложностей когнитивно-семантического характера. Для решения этой задачи нами был разработан новый комплексный подход, позволяющий реализовать переход от результатов по разборчивости речи к восстановлению семантических лакун и связей в рамках реконструируемого по данным восприятия текста в условиях шума. Для этого был предложен метод, включающий иерархию итераций:

- учет просодических моделей синтактико-семантической сегментации русской речи;
- учет различных видов экспликации тема-рематической структуры отдельных высказываний в рамках воспринятых фрагментов;
- учет моделей ритмо-метрических решеток русской речи;
- восстановление по полученным фрагментарным единицам зашумленного текста вербальных структур с учетом предикативной и лексической валентности применительно к русской речи;
- применение метода анализа по непосредственным составляющим (НС);
- применение метода предикативных зависимостей;
- определение единиц, характеризующихся признаком контекстуальной детерминированности с учетом вероятностного прогнозирования;
- построение вероятностной модели синтактико-семантических связей реконструируемого звучащего текста;
- общее смысловое декодирование реконструируемого звучащего текста с требуемой степенью точности смысловой реконструкции.

В докладе обсуждаются результаты исследования с опорой на предложенную выше комплексную методику. Конечная цель проекта – разработка базы знаний для автоматизированной экспертной системы применительно к русской речи.