

ISSN 2542-2197

Вестник

МОСКОВСКОГО ГОСУДАРСТВЕННОГО ЛИНГВИСТИЧЕСКОГО УНИВЕРСИТЕТА



ГУМАНИТАРНЫЕ НАУКИ

2018 6 (797)



MINISTRY OF EDUCATION AND SCIENCE
OF THE RUSSIAN FEDERATION
FEDERAL STATE BUDGETARY
EDUCATIONAL INSTITUTION OF HIGHER EDUCATION
"MOSCOW STATE LINGUISTIC UNIVERSITY"

The year of foundation – 1940

**VESTNIK
OF MOSCOW STATE
LINGUISTIC UNIVERSITY
HUMANITARIAN SCIENCES**

Issue 6 (797)

Moscow
FSBEI HE MSLU
2018



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ЛИНГВИСТИЧЕСКИЙ УНИВЕРСИТЕТ»

Год основания издания – 1940

**ВЕСТНИК
МОСКОВСКОГО ГОСУДАРСТВЕННОГО
ЛИНГВИСТИЧЕСКОГО УНИВЕРСИТЕТА
ГУМАНИТАРНЫЕ НАУКИ**

Выпуск 6 (797)

Москва
ФГБОУ ВО МГЛУ
2018

Печатается по решению Ученого совета
Московского государственного лингвистического университета

Главный редактор
доктор филологических наук, профессор **Г. Г. Бондарчук**

РЕДАКЦИОННЫЙ СОВЕТ

Алиева Н. М., д-р филол. наук, проф. (Азербайджан)
Воронина Г. Б., канд. филол. наук, проф. (МГЛУ)
Гаспарян Г. Р., д-р филол. наук, проф. (Армения)
Голубина К. В., канд. филол. наук, проф. (МГЛУ)
Гомес М. К., проф. лингвистики (Кадис, Испания)
Дудик Н. А., канд. филол. наук (МГЛУ)
Имомзода М. С., д-р филол. наук, проф. (Таджикистан)
Ирисханова К. М., канд. филол. наук, проф. (МГЛУ)
Ирисханова О. К., д-р филол. наук, проф. (МГЛУ)
Краева И. А., канд. филол. наук, проф. (МГЛУ)
Красноженова Г. Ф., д-р социол. наук, проф. (МГЛУ)
Кунанбаева С. С., д-р филол. наук, проф. (Казахстан)
Медведева Т. В. канд. филол. наук, проф. (МГЛУ)
Моисеенко Л. В., д-р филол. наук, проф. (МГЛУ)
Мусаев А. И., д-р филол. наук, проф. (Кыргызстан)
Ноздрина Л. А., д-р филол. наук, проф. (МГЛУ)
Писанова Т. В., д-р филол. наук, проф. (МГЛУ)
Радченко О. А., д-р филол. наук, проф. (Россия)
Русецкая М. Н., д-р пед. наук, проф. (Россия)
Сорокина Т. С., д-р филол. наук, проф. (МГЛУ)
Убин И. И., д-р филол. наук, проф. (МГЛУ)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Бондарев А. П., д-р филол. наук, проф.
Васильюк И., канд. филол. наук
Воробьев В. В., д-р филол. наук, проф.
Ганин В. Н., д-р филол. наук, проф.
Гусейнова И. А., д-р филол. наук, доц.
Евдокимов А. Ю., академик РАН, д-р техн.
наук, канд. культурологии, доцент
Евтушенко О. В., д-р филол. наук, доц.
Жаринов Е. В., д-р филол. наук, доц.
Жданова Л. М., канд. филол. наук, доц.
Захари Захариев, д-р филол. наук, проф.
Карневская Е. Б., канд. филол. наук, проф.
Косиченко Е. Ф., д-р филол. наук, доц.
Кудрявцева Н. Б., д-р филол. наук, доц.
Кузнецов В. Г., д-р филол. наук, проф.
Малыгина И. В., д-р филол. наук, проф.
Осьминина Е. А. д-р филол. наук, проф.
Полетаева М. А., канд. культурол., доц.
Порохницкая Л. В., д-р филол. наук

Потапова Р. К., д-р филол. наук, проф.
Семина И. А., д-р филол. наук, доц.
Силантьев Р. А., д-р истор. наук, доц.
Собакин А. Н., д-р филол. наук, доц.
Сомова Е. В., д-р филол. наук, проф.
Сухарев Ю. А., д-р филол. наук, проф.
Тёмкин В. А., канд. истор. наук, доц.
Толкачев С. П., д-р филол. наук, проф.
Травников С. Н., д-р филол. наук, проф.
Трыков В. П., д-р филол. наук, проф.
Уралова Л. А., канд. филол. наук, доц.
Фадеева Г. М., канд. филол. наук, доц.
Харитончик Э. А., д-р филол. наук, проф.
Хитина М. В., д-р филол. наук, доц.
Цветева Е. Н., канд. филол. наук, доц.
Ченки А. Дж., д-р наук по славянским языкам
Чернозёмова Е. Н., д-р филол. наук, проф.
Янулевичене В., д-р гуманитарных наук, проф.
Голубкова Е. Е. д-р филол. наук, проф.

СОДЕРЖАНИЕ

Предисловие	9
-------------------	---

МЕЖДИСЦИПЛИНАРНО-КОНВЕРГЕНТНЫЕ ОСНОВЫ РЕЧЕВЕДЕНИЯ

Забелло Н. А.

К вопросу о создании специализированного речевого корпуса	11
---	----

Кузнецов В. Б.

СГ-коартикуляция и место образования согласного в русском языке.

Данные локус уравнений	21
------------------------------	----

Курьянова И. В.

Маркеры речевой агрессии в интернет-коммуникации
при исследовании текстов экстремистской направленности

29

Лыкова О. В., Гордеев Д. И.

Сравнительный анализ семантического поля «обман»
на материале русского и английского языков

39

Попова М. В.

Подходы к интерпретации понятия «интерференция»
в отечественной и зарубежной науке

52

Потапова Р. Г., Забелло Н. А., Бобров Н. В.

Исследование характеристик речи в состоянии наркотической
интоксикации (на основе перцептивно-слухового анализа)

65

Потапова Р. К.

Когнитивно-семиотическая модель речевой деятельности
(применительно к принципу экономии) (на немецком языке)

73

Потапова Р. К., Комалова Л. Р.

Полнотекстовая аннотированная база данных русскоязычных
полилогов социально-сетевых дискурса: подкорпус
«Политические трансформации» (на английском языке)

84

Потапова Р. К., Потапов В. В.

Речевые базы данных как часть мультимодальных корпусов в Интернете

99

АКТУАЛЬНЫЕ ПРОБЛЕМЫ АНАЛИТИЧЕСКОЙ ТЕКСТОЛОГИИ

Джунковский А. В.

Стеганография: метод трехступенчатого анализа

для письменных текстов русского языка (на английском языке)	117
---	-----

<i>Долинский В. А.</i>	
Семейство ранговых распределений в квантитативной лингвистике	124
<i>Потапов В. В.</i>	
К вопросу о современных подходах к концепции «неиконическая хронология» (по В. С. Евсееву)	156
<i>Харламов А. А., Васин Б. И.</i>	
Инструментарий для информационно-аналитической экспертной оценки научной продукции	164
<i>Хитина М. В.</i>	
Исследование идентификационных признаков лексического уровня в монологах, диалогах, полилогах	175

ИНФОРМАЦИОННО-КОММУНИКАЦИОННАЯ ЛИНГВОДИДАКТИКА

<i>Ахэрн И. К., Горожанов А. И.</i>	
Информационные и коммуникационные технологии в педагогической практике (на опыте педагогического факультета университета Комплутенсе и Центра дистанционного обучения МГЛУ) (на английском языке)	185
<i>Свердлова Н. А.</i>	
Лингвopsихологические особенности формирования языковой личности билингва	196

TABLE OF CONTENTS

Preface	9
---------------	---

INTERDISCIPLINARY-CONVERGENT FUNDAMENTALS OF SPEECH SCIENCES

Zabello N. A.

On Specialised Speech Corpus Creation	11
---	----

Kouznetsov V. B.

CV-Coarticulation and Consonant Place of Articulation in Russian. Locus Equation Data	21
--	----

Kuryanova I. V.

Speech Aggression Markers of Internet-Communication in the Study of Extremism Focused Texts	29
--	----

Lykova O. V., Gordeev D. I.

Comparative Analysis of the Semantic Field «Deception». a Case Study of Russian and American Imageboard Messages	39
---	----

Popova M. V.

Approaches to Interpretation of the Concept of «Interference» in Russian and Foreign Science	52
---	----

Potapova R. K., Zabello N. A., Bobrov N. V.

Research on Speech Characteristics under Drug Intoxication (based on auditive analysis)	65
--	----

Potapova R. K.

Cognitive-Semiotic Model of Speech-Activities (regarding the economy-principle)	73
--	----

Potapova R. K., Komalova L. R.

Russian and Full-Text Annotated Polylogues Database: “Political Transformations” Sub-Corpus	86
--	----

Potapova R. K., Potapov V. V.

Spoken Language Databases as a Part of Multimodal Corps on the Internet	99
--	----

ACTUAL PROBLEMS OF ANALYTICAL TEXTOLOGY

Dzhunkovskiy A. V.

Steganography: Three-Stage Analysis Methodology Applied
to Russian Written Texts 117

Dolinsky V. A.

Family of Rank Distributions in Quantitative Linguistics 124

Potapov V. V.

To the Question of Modern Approaches to the Concept
of «Non-Iconic Chronology» (by V. S. Yevseyev) 156

Kharlamov A. A., Vasin B. I.

Tools for Information-Analytical Expert Evaluation
of Scientific Production 164

Khitina M. V.

Research of Identification Features on Lexical Level
of a Monologue, Dialogue, and Polylogue 175

INFORMATION AND COMMUNICATION LINGUODYDACTICS

Ahern A. K., Gorozhanov A. I.

Information and Communications Technologies in Pedagogical Practice
(based on the experience of the school of education at the UCM and the
E-learning centre at the MSLU) 185

Sverdlova N. A.

Lingvopsychological Features of Formation of Linguistic Bilingual Identity 196

История кафедры охватывает период с 1987 / 1988 до 2017/2018 учебных годов.

Тридцать лет – это и много, и мало: много с позиции жизни человека, мало с позиции макрокосмоса. За это время кафедрой выпущены сотни студентов-профессионалов, десятки аспирантов и докторантов в области прикладной, математической и экспериментальной лингвистики; написаны сотни новаторских учебно-методических программ, учебных пособий, научно-исследовательских работ, имеющих большое значение для развития теоретической и прикладной лингвистики, фундаментального и прикладного речеведения. В течение всех этих лет формировался уникальный профессорско-преподавательский состав, создавались исследовательские центры, разрабатывались перспективные научные проекты, имеющие большое практическое и народно-хозяйственное значение. На протяжении истории своего существования кафедра создавала дочерние кафедры, вошедшие «в плоть и кровь» преподавательского процесса МГЛУ.

Развитие новых информационных технологий и цифровых методов получения, хранения, обработки и передачи лингвистической информации послужило мощным стимулом к дальнейшему развитию кафедры, вызвало особую востребованность ее научной продукции, позволило развить и укрепить международные связи на поприще фундаментальных и прикладных наук.

Долгих лет тебе, наша дорогая кафедра! Новых свершений и побед!

Основатель кафедры,
директор Института прикладной и математической лингвистики МГЛУ,
академик Международной академии информатизации,
доктор филологических наук, профессор

Родмонга Кондратьевна Потапова

МЕЖДИСЦИПЛИНАРНО-КОНВЕРГЕНТНЫЕ ОСНОВЫ РЕЧЕВЕДЕНИЯ

УДК 81'33

Н. А. Забелло

аспирант кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики факультета английского языка МГЛУ;
e-mail: natangens@gmail.com

К ВОПРОСУ О СОЗДАНИИ СПЕЦИАЛИЗИРОВАННОГО РЕЧЕВОГО КОРПУСА

Статья посвящена опыту создания специализированного корпуса образцов устной речи женщин – носителей русского языка, находящихся под воздействием наркотических веществ. В исследовании рассматривается разработка специальной методики аудиофиксации, рассчитанная на условия стационара. Запись фонограмм проводилась как минимум в три сессии в состоянии интоксикации (при поступлении в стационар), абстиненции (в процессе лечения) и трезвости (перед выпиской). Во время каждой сессии записывались разные виды речи: спонтанная (коммуникация), квазиспонтанная (пересказ текста) и чтение. Автор описывает трудности, возникшие при создании речевого корпуса, большинство из которых связано со следующими внешними факторами: нестабильностью эмоционального состояния дикторов, преждевременным отказом от лечения, отказом от подписания «Информированного согласия», отсутствием лечащего врача и / или психолога при записи фонограмм, отсутствием постоянного помещения для записи, появлением посторонних шумов, вызванных медицинским персоналом, на фонограмме. Также в статье предлагаются рекомендации для увеличения объема специализированного речевого корпуса и повышения качества фонограмм и коммуникации с наркозависимыми. Описываются дальнейшие действия по обработке корпуса и его применения.

Ключевые слова: речевой корпус; интоксикация; аудиофиксация.

N. A. Zabello

Postgraduate Student,
Department of Applied and Experimental Linguistics,
Institute of Applied and Mathematical Linguistics, Faculty of English, MSLU;
e-mail: natangens@gmail.com

ON SPECIALISED SPEECH CORPUS CREATION

The article is focused on creation of a corpus of audio recordings of female Russian speakers in different intoxication states. The paper presents the development of special methods designed for hospital environment. Audio recording was conducted in at least three sessions in intoxication (on admission to hospital), abstinence (during treatment) and sobriety (before release). During each session different kinds of speech were recorded: spontaneous (communication), quasipontaneous (text retelling), and reading aloud. The author describes the difficulties that have emerged during the creation of speech corpus, most of which are related to the following external factors: instability in the emotional state of speakers, premature treatment refusal, refusal to sign the "Informed consent", the absence of a attending physician and/or psychologist in the process of recording, the absence of permanent recording facilities, the appearance of extraneous noise (caused by medical staff) on the audio recording. The article also offers some recommendations for the increasing of specialised speech volume and the improving the quality of audio recordings and communication with drug addicts. In addition, further actions for corpus processing and its deployment are described.

Key words: speech corpus; intoxication; audiofixation.

Введение

Речь – явление, которое привлекает внимание лингвистов, психологов, психолингвистов и специалистов других областей знаний. В прикладном речеведении к основным направлениям в области теоретических и практических исследований относят: речь в состоянии интоксикации (алкогольной, наркотической, медикаментозной), имитируемую речь, разработку автоматических и автоматизированных систем идентификации и верификации диктора по речи; факторы, модифицирующие речевой сигнал, и т. д. [Богданов, Брухтий, Кривнова, Подрабинович, Строкин 2003, Потапова, Потапов 2006].

Корпусная лингвистика является относительно новым направлением в лингвистической науке, которое позволяет проводить исследования различного рода. На данный момент имеется большой опыт создания корпусов текстов, однако наибольший интерес и сложность представляют речевые корпуса, создание которых главным образом способствует решению задач прикладного речеведения [Кривнова, Захаров, Строкин 2001, Захаров, Богданова 2013].

Статья посвящена опыту создания специализированного речевого корпуса, в который вошли образцы устной речи женщин – носителей русского языка, находящихся под воздействием наркотических веществ.

Методика

В целях создания специализированного корпуса образцов устной речи лиц, находящихся под воздействием наркотических веществ, была разработана специальная методика аудиофиксации. Данная методика рассчитана на условия стационара – аудиозапись в специальных медицинских учреждениях под наблюдением лечащего врача, психотерапевта и с участием специалиста-лингвиста.

Аудиофиксация образцов устной речи женщин – носителей русского языка, находящихся под воздействием наркотических веществ, проводилась на базе Научно-исследовательского института наркологии – филиала ФГБУ «Федеральный медицинский исследовательский центр психиатрии и наркологии имени В. П. Сербского» Минздрава России.

Звукозапись речевой деятельности производилась после письменного согласия пациентов клиники на участие в исследовании и с разрешения лечащего врача (см. табл. 1).

Бланк информированного согласия на участие в исследовании

ИНФОРМИРОВАННОЕ СОГЛАСИЕ

Я, _____,
даю согласие на участие в исследовании, которое посвящено изучению речи женщин носителей русского языка в состоянии наркотического опьянения.

была осведомлена о следующем:

- целях данного исследования
- тестах, которые будут произведены

Мною в устном виде была получена вышеназванная информация и ответы на все мои вопросы. Я даю согласие на просмотр моей медицинской карты сотрудниками, задействованными в исследовании, при условии сохранения ими профессиональной конфиденциальности. Мною были получены гарантии того, что я имею право прекратить участие в исследовании. Я даю согласие на то, чтобы данные обо мне были занесены в компьютерный файл, а также на обработку этих данных с использованием автоматизированных систем.

Дата «__» _____ 201__ г.

Подпись _____

В связи с условиями содержания наркозависимых пациентов клиники минимальный набор образцов устной речи испытуемых ограничивается тремя сессиями.

После консультации с лечащим врачом и психологом-клиницистом, учитывая возможности испытуемых к речевой деятельности в исследуемых состояниях, для аудиозаписи предлагалось использовать связный текст и отдельные фразы для чтения.

Аудиофиксация образцов устной речи проходит в несколько сеансов (первичным состоянием испытуемой считается то, в котором она поступила в клинику). С учетом условий сбора звучащего материала состояния интоксикации фиксируются в обратном порядке по мере снижения степени опьянения до состояния трезвости. Соответственно планировалось производить запись фонограмм дикторов-женщин в трех разных состояниях, а именно:

- интоксикации;
- абстиненции;
- в естественно-нейтральном (трезвом) состоянии.

Таким образом, предполагалось проводить запись каждого диктора в три этапа:

- при поступлении в стационар;
- в процессе лечения;
- перед выпиской.

При этом на каждом этапе планировалось записывать разные виды речи: спонтанную (коммуникацию), квазиспонтанную (пересказ текста) и чтение.

В целях обеспечения статистической значимости было принято решение записывать один вид наркомании – опийной.

Во время первого сеанса коммуникация выстраивалась на основе опроса, необходимого для заполнения персональной анкеты участника исследования. Результаты тестирования и опроса заносятся в дикторскую карточку соответствующего испытуемого.

Аудиофиксация речевой деятельности испытуемых должна была проводиться в смотровом кабинете лечащего врача в присутствии клинического психолога, лечащего врача и лингвиста.

В целях снижения возможного «микрофонного эффекта» образцы звучащей речи записываются с помощью цифрового линейного РСМ диктофона марки «Olympus LS-11E с контроллером». Частотный

диапазон в режиме РСМ составляет 44,1–96 кГц. Расстояние между микрофоном диктофона и испытуемым составляет 50–70 см.

Сеанс аудиофиксации состоит из трех этапов:

1. Беседа (диалог, полилог) с испытуемой (вид речевой деятельности – коммуникация) проводится на материале анкеты испытуемого («Карточки участника исследования»). Задаются вопросы из анкеты, испытуемый отвечает на них. Затем выбирается одна из тем разговора, которая заинтересовала испытуемую, либо испытуемую спрашивают об эмоциональном и соматическом состоянии в настоящий момент. После этого предлагается перейти к выполнению заданий.

2. Чтение фраз. Испытуемому предлагалось прочитать шесть фраз вслух. Испытуемый может отказаться от выполнения задания.

Пересказ прочтенного про себя текста. Испытуемой в силу ее состояния предлагается текст и дается 5–7 минут на его освоение. Испытуемая может отказаться от выполнения задания.

Ввиду того, что некоторые испытуемые как пациенты клиники отказываются от лечения, они не продолжают участие в исследовании, и образцы устной речи этих испытуемых в определенных состояниях не удается записать.

Сеансы аудиофиксации образцов устной речи испытуемых протоколируются. Каждой фонограмме присваивается имя. В имени фонограммы учитываются следующие данные (слева направо):

- фамилия испытуемой;
- номер сеанса;
- порядковый номер фонограммы в общем объеме фонограмм.

Кроме того, для каждой фонограммы заполняется «Карточка сеанса записи», куда вносится информация о дате проведения сеанса, времени начала, кода фонограммы состояния (интоксикация, абстиненция, трезвость), физическом состоянии на момент осмотра и эмоциональном фоне испытуемой на момент осмотра.

В «Карточку участника исследования» входят следующие пункты:

- Ф.И.О.
- дата рождения
- место рождения
- рост, см
- вес, кг
- телосложение

- дата поступления в стационар
- диагноз по МКБ-10
- психоэмоциональный тип личности (согласно тесту)
- тип употребляемого наркотического вещества
- частота употребления наркотического вещества, количество раз за 100 дней
- длительность злоупотребления наркотическим веществом, количество лет
- время после последнего приема наркотического вещества, количество часов
- наличие сопутствующих заболеваний
- родной язык
- образование, профессия, род занятий
- владение иностранными языками
- социальный статус (субъективная оценка)
- дополнительная информация.

Ниже приведены фразы для чтения вслух, предлагаемые испытуемым:

1. Эти жирные сазаны ушли под палубу.
2. Папа потекает Тате и Кате.
3. Мама намыла мою малину.
4. Саша с силой кусал сало.
5. Шушера шуршала в шалаше.
6. Баба забыла бадью в бане.

Далее представлен связный текст для пересказа (так называемый фонетически сбалансированный текст), разработанный Р. К. Потаповой, который предлагался испытуемым:

«Никогда еще ленинградцы не были свидетелями таких интересных происшествий на воде. Широчайшие набережные у речных вокзалов в это июльское лето заполнены людьми: детьми, юными и пожилыми болельщиками. Двадцать лучших скутеристов из шести стран (Англии, Гвинеи, Франции, Германии, Голландии, России), вошедшие в элиту спортсменов-водников, съезжаются к берегам Невы. Трасса пробега сверхмощных скутеров растягивается ни на один километр в длину, слегка расширяясь к устью.

Как вихрь пронесся первый корабль. Не рассчитав усилий, с жутким скрежетом столкнулись английский и французский лидеры. В одно мгновение заядлые гонщики были сброшены в глубь мутных

вод. «Нужна экстренная помощь?», – подумали зрители. Помочь бы им, но без чувства страха они без задержки взбираются на угол оббитой кормы, тщетно пытаясь обмануть безжалостную судьбу.

Над землей угрожающе стучалась мгла. Небо заволокло тучами. Подтвердился прогноз обсерватории: идет циклон с дождем со Шпицбергена. В третьем пробеге участвовали лишь вятером. Что их ждет? Кто же выиграет? Смог бы худощавый шутник геологоразведчик с поэтическим именем Олег Айвазовский? Но, поддерживая бешеный темп, чтобы партнеров эффектно объехать сбоку, командир, жонглируя рулем, не сориентировался, и экипаж очутился в пучине бушующих волн. Выкарабкавшись из-под затонувшего судна, пробежчики вплавь направились к шлюпке, поджидавшей их у финишного пункта. Наконец, судья, оценив катастрофическую ситуацию, с сожалением осуществил общее пожелание: закончить эти жесточайшие испытания» [Потапова 2015, с. 312].

Трудности создания корпуса

На данный момент было зафиксировано 40 дикторов-женщин – носителей русского языка, находящихся на лечении в МНПЦ Наркологии в возрасте от 18 до 40 лет. Количество сеансов записи составило от 1 до 6 различной длительности.

В процессе записи фонограмм операторы столкнулись с рядом трудностей. Лечение в стационаре длится 21 день. Однако пациенты вправе уйти раньше указанного срока, написав отказ от лечения, что затрудняет прогнозирование второго и третьего сеанса записи фонограмм.

При поступлении пациентки не всегда находятся в состоянии интоксикации (включая остаточную). Пациенты поступают на лечение стихийно, по понятным причинам операторы не могут физически находиться в приемном отделении круглосуточно, семь дней в неделю в ожидании поступающих женщин. Врачи приемного отделения, в свою очередь, не сообщают операторам о новом поступлении, поэтому в большинстве случаев на момент записи испытуемые уже лечатся.

Учитывая добровольность участия, а также общее состояние пациенток, мы периодически сталкиваемся с негативной реакцией на «Информированное согласие», что, в свою очередь, приводит к отказу от участия в исследовании.

В большинстве случаев сеансы записи фонограмм проводились без участия лечащего врача и/или психолога ввиду их занятости. У операторов в некоторых случаях возникают трудности с определением эмоционального и общего состояния испытуемой на момент записи.

По ряду причин помещение, где проводится аудиофиксация речи испытуемых, не всегда одно и то же – кабинет психолога, столовая, коридор отделения. В некоторых фонограммах на фоне появляются посторонние звуки и голоса персонала заведения (медсестры, охранники, санитарки).

Часто сеансы записи фонограмм проходили с разрешения лечащего врача и психолога, но без их участия. Операторы не имеют богатого опыта и навыков общения с наркозависимыми, что иногда приводит к коммуникативной неудаче.

В процессе сбора материала выяснилось, что испытуемые находятся под воздействием центрально действующих препаратов в различных дозировках практически на протяжении всего лечения. Вероятно, один из наилучших моментов получить фонограмму в трезвом состоянии – проводить сеанс записи перед выпиской из стационара. Однако выписка происходит обычно в 9 утра сразу после комиссии, на которой присутствуют пациентка, лечащий врач, психолог и заведующий отделением. После этого пациентки спешат покинуть стены учреждения, они негативно настроены на общение с оператором.

В связи с небольшим количеством женщин, употребляющих наркотики и проходящих лечение в стационаре МНПЦ Наркологии, было принято решение производить запись женщин, страдающих другими видами наркомании помимо опишной, несмотря на то, что подавляющее большинство пациенток употребляют вещества, относящиеся к группе опиатов (героин, метадон, кодеин и т. д.).

В процессе работы с некоторыми испытуемыми наблюдалось, что успешность коммуникации зависит от пола человека, производящего запись фонограммы. Некоторые испытуемые более склонны идти на контакт с мужчиной.

Рекомендации

Для увеличения объема специализированного речевого корпуса и повышения качества коммуникации и фонограмм предлагаются следующие рекомендации.

1. Для получения лучшего качества фонограммы необходимо держать диктофон непосредственно у рта испытуемой, например предлагать им держать диктофон в руках, чтобы избежать возможного дискомфорта со стороны испытуемой. Также рекомендуется производить запись в одном и том же помещении небольшой площади для снижения эффекта реверберации.

2. Необходимо продолжить сбор аудиоматериала, чтобы отобрать фонограммы испытуемых только с опийной наркоманией в целях повышения репрезентативности речевого корпуса.

3. Проводить сеансы аудиофиксации оператором (участником коммуникации) желательно одного пола (мужского) во избежание влияния на речь испытуемых.

4. Необходим более тесный контакт с персоналом медицинского учреждения. Возможно, проводить сеансы записи (по крайней мере первого) следует в приемном отделении или лечащему врачу при поступлении до начала медикаментозного лечения, а не лингвисту. Данное решение также позволит точнее определять эмоциональное состояние испытуемых на момент записи.

5. Следующим этапом создания корпуса станет аннотирование имеющихся фонограмм [Harrington 2010], включая отекстовку, маркирование в записи паралингвистических явлений, сопутствующих речи (паузы, смех, бормотание, кашель и т. д.). В дальнейшем планируется проведение перцептивно-слухового и акустического анализа фонограмм с целью выявить информативные устойчивые акустические и перцептивно-слуховые характеристики, позволяющие установить идентификационные признаки речи в состоянии наркотической интоксикации. Полученные результаты будут полезны при дальнейшей разработке систем автоматического распознавания, а также экспертных систем автоматической и автоматизированной идентификации говорящего и его эмоционального состояния по голосу и речи.

СПИСОК ЛИТЕРАТУРЫ

- Богданов Д. С. [и др.]. Технология формирования речевых баз данных // «Организационное управление и искусственный интеллект». М. : Эдиториал УРСС, 2003. 448 с.
- Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн. СПб. : СПбГУ. РИО. Филологический факультет, 2013. 148 с.

Кривнова О. Ф., Захаров Л. М., Строкин Г. С. Речевые корпуса (опыт разработки и использование) // Труды международного семинара «Диалог–2001 по компьютерной лингвистике и ее приложениям» / под ред. А. С. Нариньяни (ред.). Т. 2. М. : Филол. ф-к МГУ, 2001. С. 230–236.

Потапова Р. К., Потапов В. В. Язык, речь, личность. М. : Языки славянской культуры, 2006. 496 с.

Потапова Р. К. Речь. Коммуникация, информация, кибернетика. М. : URSS, 2015. 600 с.

Harrington J. The Phonetic Analysis of Speech Corpora. Institute of Phonetics and Speech Processing, 2010. URL: www.phonetik.uni-muenchen.de/~jmh/research/pasc010808/pasc.pdf

УДК 81'342.1

В. Б. Кузнецов

кандидат филологических наук, доцент кафедры прикладной и экспериментальной лингвистики факультета английского языка МГЛУ;
e-mail: kuvlad2007@yandex.ru

СГ-КОАРТИКУЛЯЦИЯ И МЕСТО ОБРАЗОВАНИЯ СОГЛАСНОГО В РУССКОМ ЯЗЫКЕ. ДАнные ЛОКУС УРАВНЕНИЙ

В статье рассматривается сложный вопрос исследования количественных характеристик СГ-коартикуляции в русской речи методом локус уравнений.

В качестве речевого материала использовались структуры СГСГ (ударение на втором слоге). Согласные были представлены одним и тем же шумным согласным разного способа и места образования. Шесть дикторов мужчин были записаны в студийных условиях.

Показано, что в ударном слоге F_2 переход различается у твердых согласных и не зависит от способа образования согласных. В безударном слоге наблюдается последовательное противопоставление твердых велярных губным. Не различаются уравнения регрессии только у безударных и ударных велярных. F_2 переход мягких согласных, находящихся в ударном и безударном слоге, не зависит от места и способа образования согласного. Установлено, что у безударных и ударных мягких согласных уравнения регрессии значительно различаются: у безударных мягких F_2 -пересечение более низкое, а наклон линии регрессии более крутой.

Ключевые слова: форманта; формантные треки; формантные переходы; линейная регрессия; локус; локус уравнение; F_2 -пересечение; коартикуляция.

V. B. Kouznetsov

PhD, Chair of applied and experimental linguistics, department of English language, MSLU; e-mail: kuvlad2007@yandex.ru

CV-COARTICULATION AND CONSONANT PLACE OF ARTICULATION IN RUSSIAN. LOCUS EQUATION DATA

The aim of the present paper is to describe CV-coarticulation in Russian by means of locus equations. The effects of such factors as place and manner of consonant articulation, its palatalization, position in stressed or unstressed syllable on properties of F_2 transition were examined.

CVCV structures were used as speech material. The second syllable was stressed. The consonants were realized by an identical obstruent consonant having different manner and place of articulation. Six male speakers were recorded in a sound treated room.

It was shown that locus equations for nonpalatalized stressed consonants were different. The role of manner of articulation in this case was insignificant. F_2 transition of palatalized stop and fricative consonants were adequately described by two regression equations – one in stressed syllable and the other for the consonants unstressed syllable.

Key words: formant; formant transition; linear regression; locus; locus equation; intercept; slope; coarticulation; formant tracks.

Со времен первых спектрографических исследований речи [Potter, Корр G., Корр Н. 1947] сложилось представление о том, что формантные переходы гласного содержат информацию, в частности о месте образования предшествующего согласного. Анализируя спектрограммы сочетаний взрывных согласных с различными гласными, Поттер и др. [Potter, Корр G., Корр Н. 1947] высказали предположение, что в качестве инвариантного акустического коррелята места образования согласного (далее – МОБР) может служить частота F_2 ¹ того гласного, у которого после данного согласного отсутствует формантный переход. Так, для согласных [b] и [d] такими гласными-индикаторами являются гласные [u] и [æ] соответственно.

В пользу предположений Поттера и др. свидетельствовали результаты экспериментов Хаскинских лабораторий [Liberman 1954] по восприятию двухформатных гласных, синтезированных по нарисованным от руки спектрограммам. Было показано, что наилучшее восприятие [b] и [d] достигалось, если F_2 переход начинался или «указывал» на частоту 720 Гц и 1800 Гц соответственно. В случае с [g] этот звук воспринимался удовлетворительно только перед задними гласными при начале F_2 с 3000 Гц. Эти характерные частоты авторы назвали локусом (locus) согласного. Последующие эксперименты показали, что МОБР, например, [d] однозначно идентифицируется в контексте различных гласных, если F_2 переход только указывает, но не начинается с 1800 Гц [Delattre, Liberman, Cooper 1955].

Однако исследование в последующие годы формантных переходов в реальной речи [Lehiste, Peterson 1961] не подтвердило статус локуса как инвариантной характеристики МОБР согласных. Отмечалась лишь относительная стабильность F_2 переходов для альвеолярного [d]. По мнению Л. А. Чистович и В. А. Кожевникова [Чистович, Кожевников 1972], понятие виртуального локуса совершенно

¹Характерную для гласных и согласных частоту F_2 авторы называли *хаб* (HUB).

несостоятельно с точки зрения обработки речевого сигнала слуховой системой человека.

В 1990-е гг. Суссман провел серию исследований [Sussman, McCaffrey, Matthews 1991; Sussman, Shore 1996] для изучения информативности формантных переходов, используя так называемые локусные уравнения (locus equations). Идею локус уравнений предложил Б. Линдблом [Lindblom 1963]. Анализируя редукцию гласных, он заметил, что между начальным значением частоты второй форманты гласного и ее значением в момент достижения гласным целевого состояния существует линейная зависимость.

В математической статистике разработан метод парной линейной регрессии, который позволяет по исходным данным (в нашем случае пары значений F_2 , характеризующих формантный переход) подобрать прямую линию (уравнение прямой линии), которая в некотором смысле наилучшим образом описывает искомую зависимость. Начальные значения F_2 откладываются на оси Y (зависимая переменная), целевые значения F_2 – на оси X (независимая переменная, предиктор). Соответственно локус уравнение имеет следующий вид для F_2 :

$$F_2 \text{ старт} = \alpha F_2 \text{ центр} + c,$$

где α – угол наклона линии регрессии, c – точка пересечения линии регрессии с осью Y (у-пересечение). Величина угла наклона характеризует степень коартикуляции МОБР согласного с последующим гласным: чем меньше наклон, тем меньше влияет гласный на МОБР согласного.

Принципиальное отличие «виртуального локуса» Хаскинских лабораторий от метрики локус уравнения Суссман видит в том, что в первом случае речь идет об абстрактном фиксированном акустическом корреляте МОБР, не зависящим от последующего гласного, тогда как локус уравнение предназначено отразить систематический характер контекстно-обусловленной взаимосвязи между начальным и целевым значением F_2 перехода.

Безошибочное распознавание МОБР [b. d. g] на основе угла наклона и пересечения позволило Суссману [Sussman, McCaffrey, Matthews 1991] заключить, что локус уравнения обладают реляционной (relational) инвариантностью акустического коррелята МОБР взрывных согласных.

Аналитическое описание характеристик формантных переходов русских гласных применялось довольно редко. Так, М. В. Лобанов

установил расчетные формулы для задания формантных переходов в целях синтеза речи по тексту [Лобанов 1982].

В настоящей работе методика локус уравнивания используется как способ количественного описания свойств F_2 перехода в русской речи в зависимости от таких факторов, как МОБР согласного, его твердость-мягкость, способ образования и ударность-безударность последующего гласного.

Методика

Речевой материал. Характер разработанного речевого материала должен был позволить исследовать влияния на СГ-коартикуляцию таких факторов, как МОБР согласного, его твердость-мягкость, способ образования (взрывной-фрикативный), нахождение в ударном или безударном слоге. Учитывая опыт предшествующих исследований, использовалась бессмысленная структура СГСГ, в которой оба согласных и гласных реализовывались одним и тем же звуком, например [тата], а ударение приходилось на второй слог. В качестве согласных использовались твердые и мягкие взрывные и фрикативные согласные [п, п', т, т', к, к', ф, ф', с, с', х, х'] и в качестве гласных – [а, и, у]. Исследуемые структуры произносились в рамочном предложении «Вырос СГСГ сильным».

Для чтения был составлен список из 108 предложений (12 согласных x 3 гласных x 3 повторения). Предложения были объединены в 18 блоков, в которых использовался один и тот же гласный, и все согласные были твердыми или мягкими. Последовательность блоков была случайной. Подобная организация материала позволила существенно сократить число ошибок при чтении исследуемых структур.

Дикторы и запись речевого материала. Были записаны шесть дикторов мужчин без дефектов речи и диалектных черт. Запись проводилась в безэховом помещении с помощью микрофона Philips SBC MD 110, внешнего АЦП Е-МУ 0204 и компьютера. Частота дискретизации – 44100 Гц, разрядность – 16 бит.

Измерение формантных частот. До проведения измерений исходная частота дискретизации была снижена до 11025 Гц. Точки измерения частоты F_2 определялись по спектрограмме и осциллограмме. Значение F_2 на границе согласного с гласным измерялось на спектре, вычисленном на интервале первого полноценного периода колебания голосовых связок. Второе значение F_2 измерялось в точке, где траектория F_2 стабилизировалась или достигала экстремального значения.

В обеих точках длина окна анализа, задаваемая вручную, совпала с одним периодом. Спектральные разрезы вычислялись двумя методами: быстрое преобразование Фурье (далее – БПФ) и линейное предиктивное кодирование (далее – ЛПК). Порядок модели ЛПК подбирался таким образом, чтобы добиться максимального совпадения с БПФ спектром. Порядок модели мог принимать значения от 10 до 14. Частота F_2 измерялась вручную на ЛПК спектре.

Результаты

Оценивая в целом качество построенных уравнений линейной регрессии, следует сказать, что все уравнения статистически значимы при $\alpha = 0,05$; коэффициенты всех уравнений достоверно отличны от нуля, за исключением двух случаев: значений пересечения в уравнении для твердых [к, х] в ударном слоге.

Твердые взрывные и фрикативные согласные

Твердые ударные взрывные различаются либо по обоим параметрам (пересечению и углу наклона линии регрессии) соответствующих уравнений, либо по одному из них. Различие состоит в том, что 95 % доверительные интервалы значений сравниваемых параметров не пересекаются. Подобранные линии регрессии и уравнения, представленные на рисунке 1, подтверждают современные представления о СГ коартикуляции. Как известно, велярный согласный легко коартикулирует с последующим гласным. Отсюда – наибольший угол наклона линии регрессии. Наименьший угол характерен для губного согласного. Артикуляторному взаимодействию гласных с этим звуком противодействует использование разных артикуляторов: губ и языка. Уравнение регрессии для [п] способно объяснить только 60 % вариативности начального значения F_2 перехода. R^2 для [т] и [к] составляет 0,85 и 0,90 соответственно.

Способ образования твердых ударных и безударных существенной роли в формировании F_2 перехода не играет: уравнения регрессии фрикативных согласных не отличаются от соответствующих взрывных.

Среди безударных твердых последовательно противопоставляются губные велярным. Безударный переднеязычный по некоторым параметрам совпадает с губным или с губным и велярным.

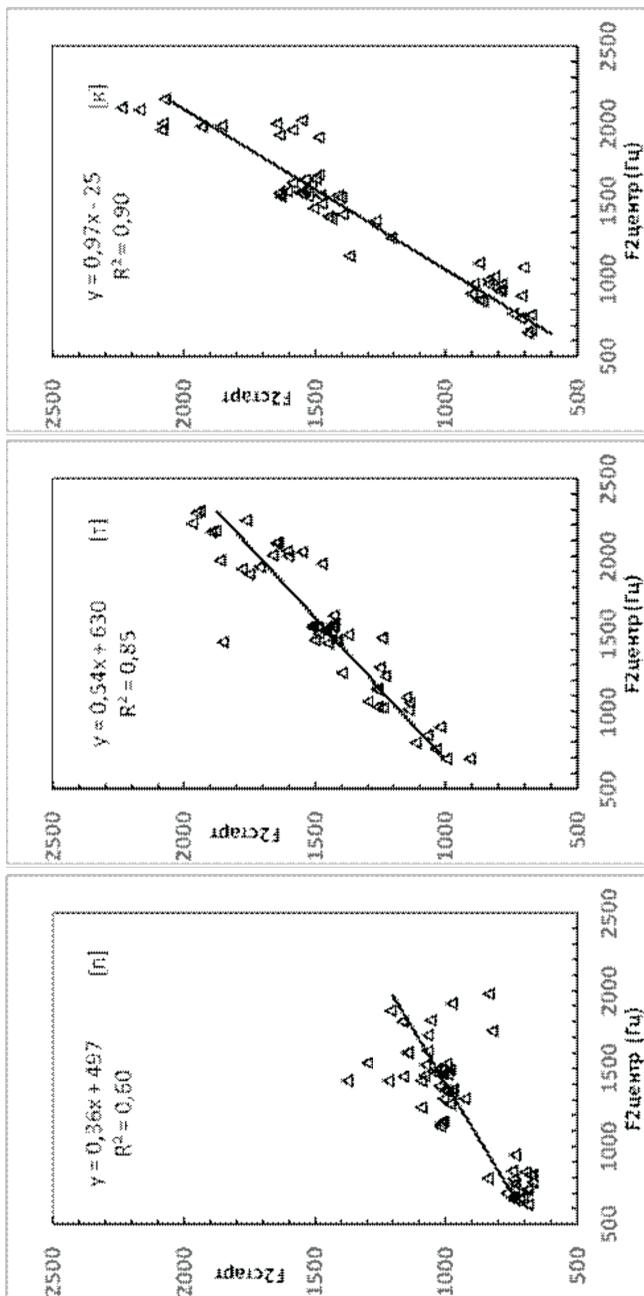


Рис. 1. Диаграммы рассеяния данных с подобранными линиями регрессии и соответствующие уравнения для твердых взрывных в ударном слоге. В уравнении регрессии для [к] значение y-пересечения не отличается от нуля при $\alpha=0,05$

Сравнение уравнений для твердых ударных и безударных показывает, что только в случае велярных наблюдается четкое совпадение. В тех случаях, когда ударные и безударные различаются, у последних отмечаются более низкие значения пересечения и более крутой наклон линии регрессии.

Мягкие взрывные и фрикативные согласные

Как можно было ожидать, ни у ударных, ни у безударных мягких согласных МОБР согласного и способ его образования не оказывают влияние на начальное значение F_2 перехода [Бондарко 1977]. Для ударных и безударных мягких локус уравнения имеет следующий вид:

ударные мягкие: $Y=0, 45x + 1118,2$ ($R^2=0, 63$);

безударные мягкие: $Y=0,74x + 511,5$ ($R^2=0,71$).

Можно заметить, что у безударного мягкого согласного начальное значение F_2 перехода в большей степени, чем у ударного, зависит от последующего гласного и существенно ниже.

Наше исследование перцептивной значимости формантных переходов гласного для идентификации МОБР глухих взрывных показало, что по переходу СГ МОБР мягкого согласного распознается на уровне случайного угадывания [Кузнецов 1999]. Более высокий результат был характерен для распознавания по ГС переходу. Спектральный анализ F_2 перехода в последовательности ГС свидетельствует о том, что МОБР согласного является вторым по значимости после ряда гласного фактором, определяющим конечное значение F_2 перехода [Кузнецов 1995; Purcell 1979].

Выводы

Основные результаты данного исследования следующие:

- продемонстрирована высокая эффективность инструмента локус уравнения при описании явлений коартикуляции;
- установлено, что способ образования согласного не влияет на характеристики F_2 перехода ни у твердых, ни у мягких согласных в ударном или безударном слоге;
- для мягких согласных подтверждено отсутствие влияния МОБР согласного на начальное значение F_2 перехода;
- впервые установлено различие F_2 переходов у мягких согласных, находящихся в ударном и безударном слогах.

СПИСОК ЛИТЕРАТУРЫ

- Бондарко Л. В.* Звуковой строй современного русского языка. М. : Просвещение. 1977. 176 с.
- Кузнецов В. Б.* Эффекты коартикуляции при переходе от гласного к мягкому согласному в русской речи // Проблемы фонетики. II. М. : ИРЯ РАН, 1995. С. 84–100.
- Кузнецов В. Б.* Роль лингвистических и перцептивных факторов в идентификации глухих смычных по вокалическому контексту // Проблемы фонетики. III. М. : ИРЯ РАН, 1999. С. 48–60.
- Лобанов Б. М., Марченков М. А.* Алгоритм синтеза формантных параметров по тексту: тезисы докладов и сообщений 12-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-12). Киев : Ин-т кибернетики им. В. М. Глушкова, 1982. С. 414–417.
- Чистович Л. А., Кожевников В. А.* Восприятие речи. // Физиология сенсорных систем. Ч. 2. Гл. 13. Ленинград : Наука. Ленинград. отд., 1972. С. 427–515.
- Delattre P. C., Liberman A. M., Cooper F. S.* Acoustic loci and transitional cues for consonants // Journal of the Acoustical Society of America. 1955. Vol. 27. P. 769–773.
- Lehiste, I., Peterson, G.* Transitions, glides, and diphthongs. // Journal of the Acoustical Society of America. 1961. Vol. 33. P. 268–277.
- Liberman A. M., Delattre P. C., Cooper F. S., Gerstman L. J.* The role of consonant-vowel transitions in the perception of the stop and nasal consonants, // Psychological // Monographs, 1954. Vol. 68. P.1–13.
- Lindblom B.* On Vowel Reduction // Royal Institute of Technology. 1963. Report # 29. Stockholm.
- Potter R. K., Kopp G. A., Green H. C.* Visible speech. New York : D. Van Nostrand Co., 1947. 441 p.
- Purcell E.T.* Formant frequency patterns in Russian VCV utterances // Journal of the Acoustical Society of America. 1979. Vol. 66 (6). P. 1691–1702.
- Sussman H. M., McCaffrey H. A. Matthews S. A.* An investigation of locus equations as a source of relational invariance for stop place categorization // Journal of the Acoustical Society of America, 1991. Vol. 90. P. 1309–1325.
- Sussman H.M., Shore J.* Locus equations as phonetic descriptors of consonantal place of articulation // Perception and Psychophysics. 1996. Vol. 58. P. 936–946.

УДК 81'33

И. В. Курьянова

заместитель начальника отдела лингвистических исследований, «Московский исследовательский центр» Департамента региональной безопасности и противодействия коррупции; e-mail: ivkuryanova@mail.ru

МАРКЕРЫ РЕЧЕВОЙ АГРЕССИИ В ИНТЕРНЕТ-КОММУНИКАЦИИ ПРИ ИССЛЕДОВАНИИ ТЕКСТОВ ЭКСТРЕМИСТСКОЙ НАПРАВЛЕННОСТИ

Статья посвящена одному из актуальных направлений судебной лингвистики – исследованию публичных текстов экстремистской направленности, размещенных в Интернет. Предметом исследования являются интернет-публикации, принадлежащие семантическому полю «агрессия». В работе рассматриваются маркеры речевой агрессии в высказываниях, аббревиатурах, цифро-буквенных обозначениях, символах, сопровождаемых или не сопровождаемых визуальной информацией. Автором описываются языковые средства речевого воздействия в процессе коммуникации. Цель статьи – выявление средств языка, используемых адресантами в процессе коммуникации, при передаче определенной коммуникативной целеустановки. Автором проводится анализ высказываний и определяются некоторые особенности оформления коммуникативного намерения, рассматриваются особенности вербальных и визуальных средств выражения состояния «агрессия» в интернет-дискурсе на лексическом, семантическом и прагматическом уровнях. Автор анализирует некоторые специфические особенности текстов экстремистской направленности (в том числе так называемых поликодовых, или креолизованных, текстов, интернет-мемов) и определяет способы их языкового оформления.

Ключевые слова: интернет-коммуникация; лингвистическая экспертиза; речевая агрессия; экстремизм; поликодовые (креолизованные) тексты; интернет-мем.

I. V. Kuryanova

Department of Regional Security and Anti-Corruption of Moscow, Moscow State Budget Institution «Moscow research center», Deputy chief of linguistic research department

SPEECH AGGRESSION MARKERS OF INTERNET-COMMUNICATION IN THE STUDY OF EXTREMISM FOCUSED TEXTS

The article deals with the important trend of forensic linguistics, i.e. public extremism focused internet texts. The focus lies on the Internet publications from the semantic field «aggression». Speech aggression markers of utterances, abbreviations, letter-digital indicators, symbols with or without visual information are considered in the present paper. The paper deals considers of language means of speech influence in the process of communications. The goal of the paper is to discover means of language which are used by addressees in the process of communications, when

transmitting certain communicative goals. The author analyzes statements and identifies some features of the speech processing effects, describes peculiarities of verbal and visual expression of the condition of aggression used in Internet discourse on lexical, semantic and pragmatic levels. The author undertakes research on some specific features of extremism focused text (so called polycode or creolized text and internet-memes) and denotes the means of their verbal representation.

Key words: Internet communication; speech aggression; linguistic forensic examination; extremism; polycode (creolized) texts; internet-meme.

В эпоху информационной глобализации понятие речевой агрессии приобретает всё более широкий контекст. В связи с ростом социально-экономической, межконфессиональной, идеологической, геополитической нестабильности особый интерес со стороны исследователей вызывает вербальная специфика феномена «агрессия» [Потапова, Комалова 2015]. Как отмечает Р. К. Потапова, расхождение между ценностными ожиданиями и ценностными возможностями отдельных личностей и сообществ является первопричиной индивидуальных и массовых конфликтов, деструктивного поведения, агрессии и террора [Речевая коммуникация... 2017, с. 17].

Модели агрессивного поведения, существующие в обществе, закрепляются в языке и транслируются из поколения в поколение как напрямую (при контактной коммуникации), так и опосредованно (через тексты). Список обозначений негативных эмоций и их проявлений, и без того достаточно разветвленный в любом языке (страх, враждебность, раздражение, отчаяние, гнев, печаль (грусть), горе, отвращение, презрение, агрессия, ненависть, тревога и др.), активно пополняется. Как отмечают исследователи, приметой нашего времени стало порождение всё новых способов сокращения позитивного поля сознания, одним из которых является образование эмотивных оксюморонов (*злобро, исчадие добра, убойная любовь, ужас восторга, достижение провала / неудачи* и т. п.), вносящих в представления о позитивных чувствах негативный компонент, порождая устойчивые *негативные коннотации* [Комалова 2017; Потапова, Комалова 2015; Речевая коммуникация... 2017].

«Питательной средой» для воспроизводства и трансляции моделей речевого агрессивного поведения являются интернет-коммуникация и тексты средств массовой информации: печатные и электронные издания газет, журналов, текстовые сообщения, отправляемые посредством Skype, Viber, WhatsApp, публикации в социальных сетях

«ВКонтакте», «Одноклассники», «Fasebook» и т. п. [Голощапова 2017; Потапова, Комалова 2015].

Ежедневно из СМИ мы узнаем о группах смерти, похищении денежных средств с электронных счетов, несанкционированном распространении персональных данных и т. п.

В научных публикациях, посвященных данной проблеме, неоднократно указывалось на явное доминирование лексики негативного содержания в современных СМИ, которую выбирают или порождают, а затем тиражируют медийные издания: *фальшивые валютчики, фальшивые лекарства, массовые вспышки отравлений (питьевой водой, лекарствами, продуктами), лицо кавказской национальности, криминалитет, бомбежка, криминальная группировка, колоться, колеса, заказное убийство, горячая точка, организованная преступность, черный вторник (пятницы, четверги и т. п.), чиновничий беспредел* и др.

Многие исследователи дискурса в СМИ и интернет-коммуникации сходятся во мнении, что телевидение и интернет-ресурс диктуют реципиенту такие установки, как «жизнь жестока», «часто насилием можно добиться своей цели», «герои могут быть жестокими и при этом остаются хорошими», «агрессия не всегда приносит вред» [Потапова, Комалова 2015; Матвеева 2016].

Данная работа продолжает изыскания по критическому осмыслению проблем, связанных с рассмотрением публичного дискурса и интернет-коммуникации как средств навязывания притязаний на власть, внедрения в общество радикальных идей и пр.

В последние годы становится всё более очевидным, что речевые проявления агрессии требуют системного и всестороннего исследования [Комалова 2017; Потапова, Комалова 2015; Речевая 2017; Щербина 2015]. Актуальность исследования феномена «агрессия» не нуждается в особом обосновании. В необходимости обращения к оценке состояния современного языка с точки зрения его «агрессивности» убеждает нас, в частности, криминогенная ситуация на территории Российской Федерации: одной из главных угроз государственной безопасности на сегодняшний день является экстремизм в различных его проявлениях. Согласно прогнозам Стратегии национальной безопасности до 2020 г., утвержденной Указом Президента Российской Федерации от 12 мая 2009 г. № 537, в современном мире всё большее развитие получают националистические настроения, ксенофобия, сепаратизм и экстремизм.

Эффективным инструментом выявления, расследования и доказывания по делам об экстремизме являются судебные экспертизы и исследования, направленные на выявление речевой агрессии в публичных дискурсах (текстах СМИ и интернет-коммуникации). Лингвистическая экспертиза одной из первых стала использоваться при доказывании по делам об экстремизме, так как в большинстве случаев позволяет установить необходимый и достаточный для последующей правовой квалификации набор признаков по материалу, предоставленному на исследование.

В связи с тем, что многие преступления экстремистского характера имеют коммуникативную природу (ст. 280 УК РФ «Публичные призывы к осуществлению экстремистской деятельности», ст. 282 УК РФ «Возбуждение ненависти либо вражды, а равно унижение человеческого достоинства», ст. 205.2 УК РФ «Публичные призывы к осуществлению террористической деятельности или публичное оправдание терроризма» и др.), объектами судебных экспертиз по делам об экстремизме, как правило, являются продукты человеческой коммуникации – устные и письменные тексты, изображения и видеозаписи.

Наиболее очевидными примерами деструктивных речевых действий являются такие, которые облакаются в резкую, категоричную, форму с использованием ненормативной (грубой и нецензурной) лексики, которая достаточно ясно маркируется в языковом сознании говорящего и адресата его речи. Использование в речи нецензурной и бранной лексики в инвективной функции с установкой на оскорбление предполагает не только общественное порицание, но и наказание по закону (например, УК РФ ст. 130).

Однако, как отмечают исследователи речевой агрессии, сложность определения данного понятия заключается в том, что речевая агрессия может выражать самые разные побуждения и намерения, возникать в разных обстоятельствах, приобретать разные формы словесного воплощения, т. е. не является единой формой поведения или деятельности [Щербинина 2015]. Замена нецензурных слов «приличными» далеко не всегда решает проблему: деструктивная роль речевых действий в коммуникации должна определяться с учетом множества факторов, как на уровне экспликаций, так и на уровне импликаций речевой агрессии.

Агрессия в терминах лингвистов – это «наступательное доминирующее речевое поведение», «вербализованное эмоциональное

реагирование», «речевой акт, заменяющий агрессивное физическое действие», «вторжение в когнитивное речевое пространство адресата», «способ создания коммуникативного дисбаланса» [Комалова 2017]. Достаточно полно отражает определение коммуникативной агрессии точка зрения Ю.В.Щербининой, согласно которой речевая агрессия – это «проявление грубости в речи, негативное речевое воздействие и взаимодействие; обидное общение; выражение отрицательных эмоций и намерений в неприемлемой в данной ситуации и оскорбительной для собеседника форме» [Щербинина 2015].

В исследовании мы, опираясь на определение, данное Р.К.Потаповой в тематическом толковом словаре, отражающем вербальную структуру коммуникативного акта агрессии, будем понимать под речевой агрессией использование языковых средств для выражения неприязни, враждебности; манеру речи, оскорбляющую чье-либо самолюбие, достоинство; побуждение (в том числе в форме призыва) к совершению насильственных (деструктивных) действий [Потапова, Комалова 2015].

Наибольшую актуальность на сегодняшний день представляют исследования, посвященные проблемам изучения речевой агрессии в интернет-коммуникации. Если язык СМИ, характеризующийся повышенной образностью, инновационностью, многозначностью, языковой игрой, подлежит цензуре, то язык интернет-коммуникации качественно трансформирован. Как отмечают исследователи, агрессивное речевое поведение пользователей интернет-ресурсов выступает для самих этих пользователей как «санкционированное насилие» [Комалова 2017]. Социальные сети предоставляют пользователям практически абсолютную свободу самовыражения, снимая ряд коммуникативных ограничений, предъявляемых в реальной коммуникации.

Большая часть составов преступлений экстремистской направленности связана с нарушением юридических норм социальной коммуникации в Интернете: закон запрещает высказывания определенного содержания и информационного статуса. Именно в интернет-коммуникации проявляется большая часть преступлений экстремистской направленности, имеющих вербальный характер.

В сферу экспертной оценки в лингвистической экспертизе по делам об экстремизме входит самое большое, по сравнению с иными категориями дел, количество самых разнообразных объектов: в силу специфики дел по расследованию преступлений экстремистской направленности

каких-либо ограничений (по жанру, стилю, способу фиксации и форме речи, виду используемых знаковых систем) не существует. Современная коммуникация характеризуется четкими тенденциями к использованию, наряду с языковыми знаками, знаков иных семиотических систем (визуальной, аудиальной). В этой связи объектами лингвистической экспертизы по делам экстремистской направленности могут являться любые вербальные и/или креолизованные компоненты информации (тексты, высказывания, слова, аббревиатуры, цифро-буквенные обозначения, символы, сопровождаемые или не сопровождаемые визуальной информацией). Каждый из видов объектов обладает своей спецификой, что, в свою очередь, влияет на выбор применяемых методов лингвистического анализа.

Сначала СМИ, а сейчас интернет-ресурс способствуют тому, что одной из тенденций развития современной коммуникации является всё большая визуализация коммуникации. Поэтому наибольшую сложность при лингвистическом исследовании представляют так называемые креолизованные (или поликодовые, семантически обогащенные) тексты.

Креолизованными называются тексты, смысл которых создается совокупностью различных знаковых систем, например с помощью слова и изображения, слова и музыки, жеста и позы и пр. [Матвеева 2016].

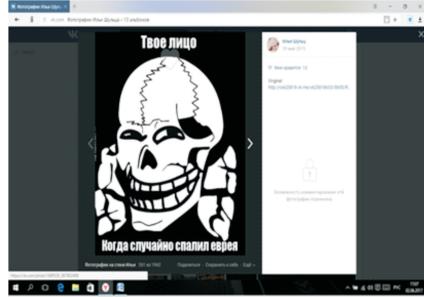
В сфере массовой коммуникации, в сфере пропаганды какой-либо идеологии креолизованные тексты имеют гораздо большую силу воздействия, чем гомогенные тексты [Матвеева 2016]. Большое распространение получают объекты, в которых изображение часто важнее вербального текста: в Интернете распространяются видеоролики, не имеющие текста

или имеющие минимальное вербальное сопровождение, назначение которого – задать вектор интерпретации [Матвеева 2016]. Часто сама вербальная информация, сопровождающая изображение, представлена символом, значение которого требует интерпретации.

Так, большую популярность в последнее время приобрели так называемые интернет-мемы. Мем – единица культурной информации,



распространяемая посредством научения < подражания, имитации> и имеющая выраженные физические и поведенческие последствия [Щербина 2015]. Мем представляет собой репликатор (подобно гену в структуре ДНК) – объект, копирующий сам себя, способный к самовоспроизведению [Потапова, Комалова 2015].



Мемами могут быть популярные мелодии, лозунги, персонажи с определенным выражением лица, перешедшие в разряд слоганов, крылатых выражений, например: *Ой, всё!, Вжух!* (сопровождающиеся изображением кота с волшебной палочкой); *Твое лицо когда...; Это фиаско, братан!* Интернет-мем сам по себе уже коннотирован определенным образом, поэтому смысловая нагрузка креолизованного текста с использованием мема, его семантика должна устанавливаться исключительно исходя из значений, входящих в семантическое поле употребления конкретного интернет-мема.

Таким образом, интерпретация объектов, рожденных в сфере интернет-коммуникации, имеет особые правила, обусловленные особенностями жанра и соответствующими конвенциями интернет-коммуникации [Матвеева 2016].

Лингвистика, будучи разделом семиотики, исследует вербальный и невербальный компоненты информации с точки зрения его смысловой составляющей. Анализ только вербального компонента в креолизованных текстах во многих случаях не позволяет адекватно интерпретировать объект исследования.

Проиллюстрируем вышесказанное примером.

Высказывание: «Встречай гостей с гор!» не выражает семантики одобрения насильственных действий по отношению к группе лиц, выделяемой по национальному (этническому) признаку (кавказцам / азиатам) – ведь встречать гостей можно и хлебом-солью, и радушием и т. п. Однако, если это высказывание представлено на фоне лица со стереотипными признаками внешности славянина, при этом вооруженного, либо на фоне топора с красными пятнами, символизирующими кровь, то смысл данной фразы становится совершенно иным.

Вербальный и визуальный компоненты следующего объекта также соотносены между собой, образуют смысловое единство.

Интернет-мем оформлен по готовой структурной модели и содержит идиоматическую формулу «*твое лицо когда...+ X*», где элемент «*X*» является вербальным обозначением каких-либо условий / описанием ситуации. Идиоматическая формула традиционно сопровождается изображением, указывающим на какую-либо оценку или эмоции относительно факторов, описанных в части «*X*».

В данном меме описываются насильственные действия в отношении евреев – *спалить – уничтожить, истребить огнем (разг.)*. Изображение белого черепа на черном фоне содержит отсылку к эмблеме элитных охранных подразделений нацистской партии в фашистской Германии (СС), в ведении которых, в частности, находились нацистские концентрационные лагеря, где в числе прочих узников целенаправленно уничтожались евреи. Данная эмблема позднее также использовалась британской нацистской организацией *Combat 18*. Автор / публикатор выражает положительное отношение к обозначенным действиям посредством визуального компонента – череп изображен с улыбкой. Таким образом, насильственные действия в отношении евреев обуславливаются идеологией нацизма, провозглашающей концепцию расовой гигиены и представляющей евреев как неполноценных людей.

Часто определение предмета речи и коммуникативной цели становится невозможным при опоре исключительно на вербальный компонент информации. Значение объекта, его коммуникативная цель заключается именно в совокупности вербального и невербального (визуального) компонентов информации.

Таким образом, в большинстве случаев невозможно адекватно исследовать креолизованный текст только в части вербальной составляющей, не анализируя его невербальных компонентов. Как отмечается в исследованиях креолизованных текстов, «изображения и слова в поликодовом сообщении не являются суммой семиотических знаков, их значения интегрируются и образуют сложно построенный смысл»¹.

Так, в отдельных случаях в креолизованных текстах используются криптограммы, которые стали знаками принадлежности к различным движениям, группировкам, субкультурам (например, знаки

¹Цит. по: [Матвеева 2016].

1488, 4/20, НН и др.). Например, знак 1488 может иметь следующие прочтения: 14 – означает 14 слов Дэвида Лэйна, члена организации белых сепаратистов:

We must secure the existence of our people and a future for white children.
– Мы должны защитить само существование нашего народа и будущее для белых детей.

88 – является закодированным приветствием «*Heil Hitler!*», поскольку буква «Н» стоит в латинском алфавите восьмой.

Таким образом, семантика изображения (в том числе цифр) как невербального знака во многом схожа с семантикой слова, она также имеет денотативное и коннотативное значения.

В лингвистической экспертизе пока не разработано четкого алгоритма исследования креолизованных текстов. Решение этой задачи требует углубления теоретических областей лингвистики, выявление тенденций и закономерностей функционирования текстовых пространств, механизмов их образования, а также методов их исследования.

Необходимым и особенно актуальным представляется создание специального словаря символов экстремистского характера. В настоящее время значения символов экстремистского характера представлены в единичных случаях и, как правило, разрозненно. Из множества существующих словарей языка преступного мира, словари символики экстремистской направленности не представлены в виде отдельных изданий. Существование словаря, содержащего максимально полную расшифровку существующих символов, значительно упростило бы задачу экспертов-лингвистов.

Актуальным направлением исследований представляется также дальнейшая разработка вопросов взаимодействия вербального и невербального компонентов в поликодовом тексте, анализ маркированного взаимодействия семиотически разнородных компонентов в разных типах поликодового текста.

Мониторинг экспертной практики по делам об экстремизме позволяет констатировать, что дальнейшая разработка алгоритма исследования речевой агрессии в публичном дискурсе и интернет-коммуникации является одной из важнейших задач судебной лингвистики, решение которой будет способствовать объективизации лингвистической и иных смежных видов экспертиз.

СПИСОК ЛИТЕРАТУРЫ

- Голощапова Т. И.* Использование специальных лингвистических знаний при противодействии рекламе и пропаганде наркотиков в сети Интернет: сб. ст. VI Всерос. науч.-практ. конф. с международным участием ГСГУ. Колмна, 2017. С. 206–212.
- Комалова Л. Р.* Агрессогенный дискурс: типология мультилингвальной вербализации агрессии. М. : Спутник+М, 2017. 275 с.
- Судебная экспертиза по делам о преступлениях экстремистской направленности [Текст] : монография / О. Н. Матвеева, Н. В. Вязигина, М. М. Градусова [и др.]. Барнаул : Новый формат, 2016. 350 с.
- Потапова Р. К., Комалова Л. Р.* Вербальная структура коммуникативного акта агрессии. Тематический толковый словарь / РАН. ИНИОН. Центр гуманит. науч.-информ. исслед. отд. языкознания ; отв. ред. Л. Р. Комалова. Вып. 1. М. : Институт научной информации по общ. наукам (ИНИОН) РАН, 2015. 146 с. (Сер. Теория и история языкознания.)
- Речевая коммуникация в информационном пространстве / отв. ред. Р. К. Потапова: колл. монография / Р. К. Потапова [и др.]. М. : ЛЕНАНД, 2017. 112 с.
- Щербинина Ю. В.* Речевая агрессия. Территория вражды. М. : Форум, 2012. 400 с.

УДК 81'322.2

О. В. Лыкова, Д. И. Гордеев

О.В. Лыкова, преподаватель в НИЯУ МИФИ, аспирант кафедры прикладной и экспериментальной лингвистики МГЛУ; e-mail: OVLykova@mephi.ru

Д.И. Гордеев, преподаватель в НИЯУ МИФИ, аспирант кафедры прикладной и экспериментальной лингвистики МГЛУ, e-mail: gordeev-d-i@yandex.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СЕМАНТИЧЕСКОГО ПОЛЯ «ОБМАН» НА МАТЕРИАЛЕ РУССКОГО И АНГЛИЙСКОГО ЯЗЫКОВ

Данная статья посвящена анализу семантического поля «обман» в английском и русском языках на материале анонимных форумов Интернета. Для проведения эксперимента была собрана база данных сообщений анонимных форумов и использовался алгоритм word2vec, позволивший проанализировать 30 млн словоупотреблений для каждого из языков. Метод word2vec был выбран для анализа, так как на данный момент он является одним из наиболее эффективных способов представления лексического пространства в векторном виде. Данный анализ выявил 10 слов с наибольшей семантической близостью для терминов из семантических полей «обман» и «deception» соответственно, что позволило сравнить данные феномены для разных языков. Семантическая близость между словами была вычислена с помощью косинусного расстояния между терминами, характеризующими семантические поля «обман» и «deception», и другими понятиями, входящими в корпус. Полученные результаты продемонстрировали склонность носителей русского языка ассоциировать понятие обмана с религиозной и духовной сферой, в то время как для пользователей американского форума характерно соотносить термин «deception» с политикой и смежными понятиями. Предложенный метод исследования может в дальнейшем быть использован для проведения межкультурного сравнительного анализа данных концептов.

Ключевые слова: word2vec; обман; ложь; нейронные сети; распознавание лжи; машинное обучение; векторы слов; извлечение данных.

O. V. Lykova

Teacher at National Research Nuclear University MEPhI, postgraduate student at the Department of Applied and Experimental Linguistics, Moscow State Linguistic University, OVLykova@mephi.ru

D. I. Gordeev

Teacher at National Research Nuclear University MEPhI, postgraduate student at the Department of Applied and Experimental Linguistics, Moscow State Linguistic University, gordeev-d-i@yandex.ru

COMPARATIVE ANALYSIS OF THE SEMANTIC FIELD “DECEPTION”. A CASE STUDY OF RUSSIAN AND AMERICAN IMAGEBOARD MESSAGES

This article uses the material of anonymous Internet forums to analyse the semantic field of deception. Two major imageboards were investigated: 2ch.hk and 4chan.org being the most popular Russian and American imageboards. For the experiment a database of messages from these forums was collected and an algorithm called word2vec was used to examine 30 million word usages for each of the languages. The word2vec method was chosen for the analysis, because at the moment it is one of the most effective ways of representing lexical space in a vector form. This analysis revealed 10 words with the highest semantic proximity for terms from the semantic field of “deception” for Russian and American English languages respectively, which allowed to compare this phenomenon in the analysed languages. The semantic proximity between words was calculated using the cosine distance between terms that characterize the semantic field under analysis and concepts that are present in the corpus. Results showed the tendency among native Russian imageboard users to link the concept of deception with religion and spiritual sphere, while American forum users tend to associate the notion of deception with politics and other similar concepts. The proposed research method can later be used to conduct cross-cultural comparative analysis of these concepts. The findings suggest new challenges for future research, since the study of lies and deception is a relatively new research field for Russian science.

Key words: word2vec; deception; lie; neural networks; lie detection; machine learning; word embeddings; data mining.

В последнее время исследование лжи и обмана приобретает всё большее значение. Известно, что данный феномен чрезвычайно важен для современной прикладной лингвистики, психологии и психолингвистики, полиграфологии и даже юриспруденции. Повышенное внимание к данному явлению обусловлено тем фактом, что люди пытались обнаружить обман годами и даже веками, но в наши дни из-за таких передовых технологий, как полиграф и искусственные нейронные сети, задача выявления реализации ложных высказываний становится более решаемой и приносит более достоверные результаты.

Речь – это способ общения с обратной связью, человек слышит то, что он говорит, поэтому контролирует речь. Легче обмануть, скрыть или исказить правду с помощью слов. В ложном высказывании фактическое состояние вещей преднамеренно описывается в искаженном виде, чтобы ввести в заблуждение партнера по коммуникации. Обман приводит к изменениям в сознании, он построен по другому принципу и «имеет свои собственные правила, свои методы» [Luria 1927, с. 92],

которые можно выявить в процессе лингвистического анализа высказываний. Таким образом, ложное утверждение представляет собой информацию, преднамеренно передаваемую в искаженной форме, сопровождающую изменение мышления и приводящее к изменению невербального поведения. Часть информации выражается явно, т. е. с помощью лингвистических средств, специально предназначенных для ее прямого выражения, другая часть передается с помощью невербальных средств коммуникации.

Ложное утверждение представляет собой сложное лингвистическое, психологическое и паралингвистическое явление. Паралингвистические параметры высказывания – это характеристики, которые сопровождают речь, такие как голос, артикуляция, выражение лица, жесты, движения тела и т. д. [Потапова 2006]. Таким образом, при анализе высказывания следует рассматривать все параметры, сопровождающие речь коммуниканта, для получения более точных результатов.

Дефиниция терминов «ложь», «обман» и «неправда» иногда может быть неточной и противоречивой из-за сложного и двусмысленного характера данных понятий. Кроме того, семантические поля этих слов являются разными для разных языков и культур. Следовательно, изучение определений этих терминов, предложенных многочисленными исследователями, может способствовать пониманию данных концептов и продвижению исследования верификации лжи. Кроме того, точное определение этих слов и понимание их семантических полей позволит нам избежать ошибок при их интерпретации и переводе.

Существует множество определений данных понятий, предложенных известными исследователями из разных стран.

Американский ученый Пол Экман является одним из ведущих экспертов, изучающих ложь и обман, и, возможно, самым известным. По его словам, ложь возникает, когда один человек специально вводит в заблуждение другого, делая это преднамеренно, без предварительного уведомления о своей цели, и без явно выраженной просьбы вести себя подобным образом от партнера по коммуникации [Экман 2009]. Согласно вышеприведенному определению, ложь должна соответствовать как минимум пяти необходимым условиям. *Во-первых*, для этого нужен лгущий коммуникант. *Во-вторых*, требуется реципиент, на которого лгущий коммуникант направляет свой обман. *В-третьих*, высказывание должно быть обманчивым. Кроме того, лгущий

коммуникант должен иметь намерение обмануть реципиента. Более того, лгущий коммуникант намерен ввести в заблуждение своего партнера по коммуникации «без предварительного уведомления о своей цели», т. е. лишить его законного права (осознаваемого лгущим коммуникантом) получить полную информацию, его «права на правду». Из данного определения следует, что ложь может рассматриваться как отказ в праве на получение информации, т. е. незаконное действие [Потапова 2016].

П. Экман выделяет две основные формы лжи: сокрытие истины и ее искажение (предоставление собеседнику ложной информации). «Если лжец скрывает информацию, он не сообщает ложную информацию. В случае искажения информации лжец предпринимает некоторые дополнительные действия – он не только скрывает правду, но также предоставляет ложную информацию, притворяясь, что она является правдой» [Ekman 2012]. Если рассматривать определение обоих видов лжи относительно нарушения закона, становится ясно, что, согласно определению П. Экмана, сокрытие истины, т. е. молчание, также можно рассматривать как незаконное действие.

Существуют и другие определения понятия «ложь», которые подтверждают, что семантическое поле данного термина связано с идеей нарушения прав человека. Ложь можно считать «преднамеренной успешной или неуспешной попыткой, без предупреждения, создать в сознании собеседника убеждение, которое лгущий коммуникант считает неверным» [Vrij 2000]. Д. Смит описывает ложь как «любую форму поведения, целью которой является предоставление собеседникам ложной информации или лишение их истинной информации» [Smith 2004]. Оба ученых подчеркивают незаконную природу лжи, полагая, что законным действием в данном случае было бы предостеречь собеседников («без предупреждения») и, таким образом, не лишать их законного права на получение информации («лишить их истинной информации»).

Значительный вклад в изучение понятий «ложь» и «обман» внес известный российский ученый В. Знаков. Его исследования были посвящены социальной, психологической и моральной природе лжи. Проведенный им эксперимент с участием 317 русских и 49 вьетнамских людей, изучение исторических источников, а также русских и зарубежных словарей способствовало изучению межкультурных различий между данными понятиями. Полученные результаты позволили ему предположить, что хотя существует много определений

термина «ложь», важно не путать этот термин с таким понятием, как «неправда» [Знаков 1999]. В. Знаков выделяет 3 разновидности понятия «неправда»: 1) вербальный эквивалент ошибки; 2) различные формы аллегории (аллегория, ирония), т. е. слова, которые в определенном контексте приобретают значение, противоположное их буквальному значению; 3) форма лжи. При этом он считает ложь не информативным, а коммуникативным явлением. Ложь – это не столько средство преднамеренного искажения фактов, сколько способ установления контакта и объединения людей. Ложь не предназначена для того, чтобы в нее верили, ее не используют, чтобы ввести в заблуждение слушателя и, таким образом, получить какую-то личную выгоду. В. Знаков полагает, что именно обман является наиболее уместным определением для ситуации, которая основана на сознательном желании одного из собеседников создать в сознании других коммуникантов ложную идею о предмете обсуждения, но, как замечает В. Знаков, в то же время автор обманчивого сообщения не искажает факты. При этом В. Знаков выделяет два основных типа обмана – обман с помощью полуправды и обман с помощью произнесения правды. Автор также считает, что хотя основная часть определений лжи, сформулированных западными мыслителями, основана на идеях нарушения прав человека, те, что даются российскими учеными и философами, подчеркивают аморальный характер лжи.

Известный русский философ В. С. Соловьев полагает, что в отличие от заблуждения и ошибки, ложь означает «сознательную и морально предосудительную противоположность истине» [Соловьев 1996]. Понимание лжи в русской культурной традиции имеет существенную особенность: в определении этого явления часто содержатся указания на морально предосудительный характер поведения лгущего субъекта. Например: «Он лжет не краснея» [Ожегов 1992]. Таким образом, можно видеть, что российские ученые склонны описывать ложь как бесчестное действие, которое является морально предосудительным. Определение лжи в русской культуре можно рассматривать с точки зрения морали, в отличие от традиций западных культур, где определение лжи рассматривается с точки зрения ее законности.

Однако со временем моральные представления могут изменяться. Всеобщая анонимность и Интернет «размывают» моральные устои [Акулич 2012]. Кроме того, само понятие лжи может изменяться в условиях анонимности в современном обществе. Тем самым на данный

момент является крайне актуальной проверка вышеприведенных соображений о сущности лжи для молодых пользователей Интернета в условиях анонимности. Одним из примеров активных сообществ с практически полной анонимностью являются имиджборды.

Сообщения пользователей имиджбордов были проанализированы с помощью алгоритма word2vec, результат выдачи которого может служить аналогом ассоциативного эксперимента [Jurgens 2012].

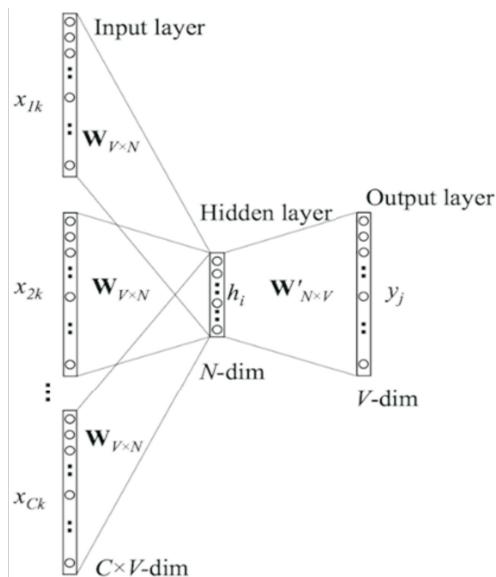


Рис 1. Иллюстрация алгоритма CBOW Word2vec

В данном исследовании было рассмотрено два имиджборда: 2ch.hk и 4chan.org, которые являются самыми популярными русскоязычным и англоязычным имиджбордами соответственно (см. рис. 1). Была использована CBOW (continuous bag of words) версия word2vec, так как она предпочтительнее для анализа коротких сообщений и дает более точные результаты для сравнения близости между словами (MSR Word Relatedness Test Set) [Mikolov 2013].

Алгоритм CBOW Word2vec основан на неглубокой нейронной сети, где $\{x_1, \dots, x_C\}$ – это унитарные векторы слов (длина вектора соответствует длине словаря V , все значения, кроме соответствующего данному слову, нулевые, значение для данного слова равно 1), количество

векторов на входе равно C – длине окна анализа. Скрытый слой – h , выход – унитарный вектор для предсказанного слова y_j . Задача Сети – предсказать на основе слов из окна анализа пропущенное слово. В результате обучения нейронной сети с помощью метода стохастического градиентного спуска на основе обратного распространения ошибок получается обученный скрытый слой h для каждого из слов модели.

В результате эксперимента полученные векторы h сравнивались для слов из семантического поля «Ложь» [Бочарова 2012]. Кроме того, было проведено сравнение полученных результатов с предпочтенными векторами для Национального корпуса русского языка [Кутузов 2017].

Таблица 1

Сравнение результатов анонимных форумов и НКРЯ для русского языка и Google News для английского (близость рассчитывалась как косинусное расстояние между векторами искомого и схожего слова)

Исходное слово	Слова корпуса анонимных форумов	Близость	Слова НКРЯ	Близость
Ложь	1. религия	0.76	1. лицемерие	0.73
	2. домыслы	0.75	2. неправда	0.71
	3. вера	0.73	3. полуправда	0.70
	4. истина	0.72	4. клевета	0.68
	5. мораль	0.72	5. обман	0.68
	6. иллюзия	0.71	6. ложь	0.67
	7. глупость	0.69	7. вранье	0.67
	8. реальность	0.69	8. вранье	0.67
	9. страдание	0.69	9. фальшь	0.63
	10. вранье	0.68	10. лживый	0.63
Обман	1. неравенство	0.64	1. обманута	0.74
	2. ветви	0.64	2. обман	0.68
	3. величайшие	0.63	3. ложь	0.68
	4. созданное	0.63	4. обманом	0.64
	5. вымысел	0.62	5. обманывать	0.62
	6. естественного	0.61	6. надувательство	0.62
	7. величина	0.61	7. самообман	0.60
	8. историческое	0.61	8. самообмана	0.60
	9. безобразие	0.61	9. притворство	0.59
	10. превозмогание	0.61	10. бездельничество	0.59

Исходное слово	Слова корпуса анонимных форумов	Близость	Слова НКРЯ	Близость
Клевета	1. несовершеннолетнего 2. совершение 3. административного 4. соединенная 5. подписание 6. правонарушения 7. заинтересованностью 8. совершившим 9. предусмотренных 10. юридического	0.85 0.83 0.82 0.81 0.81 0.8 0.8 0.8 0.79 0.79	1. клевет 2. ложь 3. инсинуация 4. клевета 5. клеветнический 6. диффамация 7. клеветник 8. злословие 9. измышление 10. навет	0.72 0.68 0.66 0.66 0.66 0.65 0.64 0.62 0.62 0.61
Лесть	1. сигнализирует 2. притворство 3. самоиронии 4. хоккинга 5. фантастическая 6. суетливый 7. инферно 8. рецензия 9. нейрофизиология 10. крымская	0.7 0.7 0.68 0.68 0.68 0.68 0.68 0.68 0.68 0.67	1. лесть 2. ласкательство 3. льстивый 4. лесть 5. предан 6. лицемерство 7. злоречие 8. двуличность 9. угодливость 10. раболепство	0.67 0.66 0.66 0.61 0.61 0.60 0.60 0.59 0.59 0.58
Лицемерие	1. феминизм 2. предательство 3. ложь 4. атеизм 5. общество 6. справедливость 7. религия 8. непонимание 9. проявление 10. злоба	0.73 0.69 0.67 0.66 0.66 0.65 0.65 0.65 0.64 0.64	1. ханжество 2. ложь 3. двоедушие 4. фарисейство 5. лицемерия 6. притворство 7. двуличность 8. двуличие 9. бесчестность 10. лживость	0.73 0.73 0.70 0.67 0.66 0.64 0.64 0.64 0.64 0.63
Коварство	1. подлых 2. у***нских 3. двачерских 4. илитариев 5. копрофилов	0.77 0.73 0.72 0.72 0.72	1. коварственный 2. хитрость 3. пронырство 4. вероломство 5. лукавство	0.68 0.63 0.63 0.62 0.61

Исходное слово	Слова корпуса анонимных форумов	Близость	Слова НКРЯ	Близость
	6. шаблонных 7. рогов 8. непорочное 9. донцова 10. слайсиков	0.72 0.72 0.71 0.71 0.71	6. ревнивость 7. коварный 8. двуличие 9. безрассудность 10. лицемерия	0.61 0.61 0.60 0.60 0.60
Измена	1. френдзона 2. дружба 3. влюбленность 4. тульпа 5. верная 6. честная 7. имеющая 8. случайность 9. сексуальная 10. любовь	0.71 0.69 0.66 0.64 0.64 0.63 0.62 0.62 0.62 0.62	1. предательство 2. измена 3. прелюбодеяние 4. неверность 5. измен 6. вероломство 7. измене 8. изменник 9. измены 10. клятвопреступление	0.73 0.64 0.64 0.62 0.61 0.60 0.60 0.59 0.59 0.59
Мошенничество	1. хулиганство 2. кражу 3. вымогательство 4. грабеж 5. умышленное 6. уголовку 7. превышение 8. правонарушение 9. вознаграждение 10. оптовая	0.74 0.71 0.71 0.7 0.69 0.68 0.68 0.68 0.67 0.66	1. вымогательство 2. мошенничества 3. хищение 4. подлог 5. кража 6. вымогательства 7. мошеннический 8. воровство 9. шулерство 10. подлога	0.69 0.67 0.67 0.63 0.63 0.62 0.62 0.61 0.60 0.60
Deception	1. rebuilds 2. hinduism 3. list_goes_on 4. dubious 5. social_constructs 6. toppled 7. propensity 8. violenceits 9. butmuh 10. grounding	0.91 0.91 0.91 0.9 0.9 0.9 0.9 0.9 0.9 0.9	1. deceit 2. deceptions 3. subterfuge 4. dishonesty 5. duplicity 6. deceitful 7. concealment 8. falsehood 9. misrepresentation 10. mendacity	0.79 0.7 0.67 0.65 0.61 0.61 0.58 0.57 0.57 0.56

Исходное слово	Слова корпуса анонимных форумов	Близость	Слова НКРЯ	Близость
Lies	1. global_warming 2. facts 3. ignorance 4. media 5. climate_change 6. mainstream_media 7. stupidity 8. false 9. free_speech 10. conspiracy	0.78 0.78 0.78 0.78 0.78 0.77 0.77 0.77 0.76 0.76	1. lie 2. rests 3. lurks 4. Lying 5. sits 6. exists 7. lying 8. falsehood 9. is 10. resides	0.69 0.57 0.51 0.49 0.47 0.46 0.46 0.45 0.44 0.44
Slander	1. left-wing 2. political_correctness 3. purges 4. identity_politics 5. medias 6. racially 7. incompetence 8. nazism 9. pro_gun 10. suppression	0.9 0.9 0.89 0.89 0.89 0.88 0.88 0.88 0.88 0.88	1. slanderous 2. slandering 3. slanders 4. libel 5. defamation 6. defaming 7. defamatory 8. libels 9. smear 10. calumnious	0.69 0.68 0.62 0.6 0.59 0.59 0.59 0.59 0.58 0.58
Flattery	1. fat_acceptance 2. illusions 3. amino_acid 4. hypnotic 5. self-awareness 6. prompting 7. psychological_warfare 8. holocaust_never_happened 9. sacrificial 10. puerto_rican	0.87 0.87 0.86 0.86 0.86 0.86 0.86 0.86 0.86 0.86	1. cajolery 2. politeness 3. ingratiation 4. blandishments 5. hyperbole 6. sycophancy 7. sarcasm 8. flatterer 9. condescension 10. bullsh**ter	0.49 0.48 0.47 0.46 0.46 0.45 0.45 0.45 0.45 0.45
Hypocrisy	1. lefties 2. bigots 3. liberalism 4. rhetoric 5. progressives 6. idiocy 7. leftists 8. hillary_supporters 9. hypocrites 10. crybabies	0.86 0.86 0.85 0.85 0.84 0.84 0.84 0.84 0.83 0.83	1. duplicity 2. disingenuousness 3. arrogance 4. hypocrisy 5. shamelessness 6. stupidity 7. absurdity 8. hypocritical 9. cowardice 10. insincerity	0.7 0.69 0.67 0.67 0.65 0.64 0.64 0.63 0.62 0.62

Исходное слово	Слова корпуса анонимных форумов	Близость	Слова НКРЯ	Близость
Cunning	1. fat_acceptance	0.91	1. devious	0.74
	2. subculture	0.91	2. wily	0.71
	3. psyop	0.91	3. canny	0.64
	4. curries	0.9	4. Machiavellian	0.64
	5. girl_queen	0.9	5. crafty	0.63
	6. russells	0.9	6. shrewd	0.63
	7. coldland	0.9	7. ruthless	0.62
	8. toppled	0.9	8. wiles	0.62
	9. indigenous_people	0.9	9. outwit	0.61
	10. ban_swords	0.9	10. clever	0.61
Adultery	1. cognizant	0.89	1. infidelity	0.67
	2. animalistic	0.89	2. Adultery	0.63
	3. idealization	0.89	3. adulterous	0.62
	4. disputes	0.89	4. apostasy	0.61
	5. engages	0.89	5. adultery	0.6
	6. infidelity	0.89	6. fornication	0.58
	7. unequal	0.89	7. unfaithful	0.57
	8. free_reign	0.88	8. zina	0.56
	9. assaults	0.88	9. adulterers	0.56
	10. equated	0.88	10. unchastity	0.56
Fraud	1. voter_fraud	0.87	1. frauds	0.77
	2. presidential	0.86	2. fraudulent	0.71
	3. legislation	0.86	3. Fraud	0.7
	4. supreme_court	0.86	4. scam	0.63
	5. senate	0.86	5. scams	0.6
	6. executive	0.85	6. embezzlement	0.59
	7. violation	0.85	7. forgery	0.58
	8. federal	0.85	8. swindle	0.56
	9. electoral	0.85	9. bribery	0.55
	10. federal_government	0.85	10. theft	0.55

Из приведенной таблицы можно заключить, что современные русскоязычные пользователи Интернета часто ассоциируют ложь с такими явлениями, как религия и феминизм. Кроме того, обобщенное представление показывает, что обман в первую очередь ассоциируется в сознании русскоговорящих пользователей форума с социальной несправедливостью и моральной стороной данного феномена («мораль», «страдание»). Американские пользователи фокусируются

больше на политической стороне семантического поля «Агрессия» («political_correctness», «nazism», «lefties», «mainstream_media»). Анализ полученных данных показал, что русскоязычные пользователи больше обращают внимание на морально предосудительный характер лжи, в то время как американские пользователи форума ассоциируют ложь и близкие к ней понятия со сферами политики и СМИ.

Заключение

Ввиду большого количества определений терминов «ложь», «обман» и «неправда», предложенных многочисленными учеными из разных стран, и выявленной разницы при анализе семантических полей данных понятий, представляется разумным полагать, что русские и западноевропейские ученые склонны интерпретировать этот термин по-разному. При анализе полученных данных были сделаны некоторые существенные выводы, касающиеся характера восприятия данных терминов в американской и русской культурах. В то время как американские пользователи анонимных форумов ассоциируют «обман» и «ложь» с политикой и правом, русскоязычные пользователи больше обращают внимание на моральный аспект данных явлений.

Полученные результаты показывают, что семантические поля данных понятий изменяются в зависимости от анализируемого языка и культуры. Чтобы избежать ошибок при использовании и переводе этих терминов, важно рассмотреть их правильное определение, их семантические поля, наиболее частый контекст их употребления и некоторые межкультурные особенности их использования и понимания.

Статистический анализ с использованием алгоритма Word2vec позволил выявить наиболее частотные термины, ассоциирующиеся с восприятием лжи для российских и американских партнеров по коммуникации, которые могут быть использованы для контент-анализа в дальнейших исследованиях.

Предложенный комплексный метод и процедура исследования в дальнейшем может быть использована для анализа семантических полей понятий «ложь», «обман» и «неправда» в различных культурах, на разных языках и для проведения межкультурного сравнительного анализа данных концептов, что способствует лучшему пониманию этих терминов и снижает вероятность ошибок при их интерпретации и переводе. Полученные данные предлагают новые задачи для

будущих исследований, поскольку изучение лжи и обмана является относительно новой и недостаточно изученной областью исследований для российской науки.

СПИСОК ЛИТЕРАТУРЫ

- Акулич М. М.* Интернет-троллинг: понятие, содержание и формы // Вестник Тюменского гос. ун-та. Социально-экономические и правовые исследования. 2012. Вып. 8. С. 47–52.
- Бочарова М. А.* Семантическое поле как способ системного описания лексики // Вестник Российского университета дружбы народов. Серия «Вопросы образования: языки и специальность». 2012. Вып. 4. С. 63–66.
- Знаков В. В.* Макиавеллизм и феномен вранья // Вопросы психологии. 1999. № 6. С. 59–69.
- Лурия А. Р.* Речь и интеллект в развитии ребенка. М. : Госиздат, 1927. 259 с.
- Ожегов С. И., Шведова Н. Ю.* Толковый словарь русского языка. М. : Азъ, 1992, 88. 257 с.
- Потапова Р. К., Потапов В. В.* Язык, речь, личность. М. : Языки славянской культуры, 2006. 329 с.
- Потапова Р. К., Лыкова О. В.* Вербальное представление лжи в русской и англо-американской культурах // Коммуникация в мультикультурном обществе: материалы I Междунар. конф. М. : МИФИ, 2016. С. 58–61.
- Соловьев В. С.* Оправдание добра. М. : Республика, 1996. 479 с.
- Экман П.* Психология лжи. СПб. : Питер, 2012. С. 31–126.
- Jurgens D. A., Turney P. D., Mohammad S. M., Holyoak K. J.* Semeval–2012: task 2: Measuring degrees of relational similarity // Proceedings of the First Joint Conference on Lexical and Computational Semantics. Vol. 1: Proceedings of the main conference and the shared task; Vol. 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, Association for Computational Linguistics, 2012. P. 356–364.
- Kutuzov A., Kuzmenko E.* WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science. Vol 661. Springer, Cham, 2017. P. 155–161.
- Mikolov T., Chen K., Corrado G., Dean J.* Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- Ekman P.* Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition). WW Norton & Company, 2009. 368 p.
- Vrij A.* Detecting Lies and Deceit. Chichester: Wiley, 2000. 276 p.
- Smith D. L.* Why We Lie: The Evolutionary Roots of Deception and the Unconscious Mind. New York : St. Martin's Press, 2004. 256 p.

УДК 81

М. В. Попова

ст. преподаватель кафедры фонетики немецкого языка факультета немецкого языка МГЛУ; аспирант кафедры фонетики немецкого языка факультета немецкого языка МГЛУ;
e-mail: neunerin@gmail.com

ПОДХОДЫ К ИНТЕРПРЕТАЦИИ ПОНЯТИЯ «ИНТЕРФЕРЕНЦИЯ» В ОТЕЧЕСТВЕННОЙ И ЗАРУБЕЖНОЙ НАУКЕ

В статье представлена история возникновения и распространения термина «интерференция» в науке, в частности его становления в лингвистике. Автор дает краткий обзор различных концепций рассмотрения данного феномена в отечественной и зарубежной лингвистике, а также предлагает вариант классификации его интерпретаций с точки зрения характера явления: негативного, позитивного или нейтрального. В последнем случае акцент делается на самом факте взаимопроникновения языков, что может отразиться на иноязычной речи билингва или мультилингва как положительно, так и отрицательно. В рамках первого подхода автор отмечает возможность дальнейшей дифференциации: анализ материалов исследования позволяет заключить, что часть лингвистов под интерференцией понимают ошибки в речи билингва, в то время как другие исследователи полагают, что отклонения от нормы являются лишь следствием интерференции. В статье приводятся примеры разночтений в понимании и употреблении термина «интерференция» и смежных понятий в зарубежных и отечественных лингвистических школах и подчеркивается факт отсутствия единого терминологического аппарата. Делается вывод о необходимости унификации терминологии, связанной с понятиями «интерференция», «трансференция» и «трансфер» – как в отечественной, так и в зарубежной лингвистике.

Ключевые слова: интерференция; трансференция; языковой трансфер; лингвистика; билингвизм; усвоение второго языка.

М. V. Popova

Senior Lecturer, Department of Phonetics, Faculty of the German Language, MSLU;
Postgraduate Student, Department of Phonetics, Faculty of the German Language,
MSLU; e-mail: neunerin@gmail.com

APPROACHES TO INTERPRETATION OF THE CONCEPT OF «INTERFERENCE» IN RUSSIAN AND FOREIGN SCIENCE

The article presents the history of the emergence and spread of the term «interference» in science, in particular - its formation in linguistics. The author presents a brief overview of various concepts of interpretations of this phenomenon in Russian and foreign linguistics, and offers a variant of classification, based on the nature of

interference: negative, positive or «neutral». In the latter case, the emphasis is on the very fact of interpenetration of languages, which can affect the foreign language of bilingual or multilingual's speech, both positively and negatively. In the framework of the first approach, the author points out the possibility of further differentiation: the analysis of research materials allows the author to conclude that some linguists understand by «interference» errors in bilingual's speech, while other researchers believe that deviations from the norm are only a consequence of interference. The article gives examples of divergences in the understanding and usage of the term «interference» and related concepts in foreign and Russian linguistic schools and states the absence of a common terminological apparatus. It is concluded that it is necessary to unify the terminology related to the concepts of «interference», «transference» and «transfer» both in Russian and in foreign linguistics.

Keywords: interference; transference; language transfer; linguistics; bilingualism; second-language acquisition.

Введение

Обращаясь к истокам возникновения данного понятия, необходимо прежде всего отметить, что изначально термин «интерференция» был введен в употребление в точных науках. Термин «интерференция» происходит от лат. *inter* – «между» и *ferens (ferentis)* – «несущий», «переносящий» и означает «содействие или противодействие двух или большего числа волн, исходящих от колебательных, периодически повторяющихся движений», что приводит при их наложении друг на друга к взаимному усилению или ослаблению [Энциклопедический словарь... 2001]. Позднее данный термин был заимствован и в другие области науки, среди которых наибольшее распространение получил в психологии, где термин обозначает «взаимодействие двух или большего числа одновременных или последовательных процессов, при котором возникает нарушение или подавление по крайней мере одного из них» [Большой психологический словарь... 2003]; в лингвистике и появившейся в англоязычных странах в рамках прикладной лингвистики сравнительно недавно дисциплине SLA (Second Language Acquisition), занимающейся проблемами усвоения иностранного языка. Необходимо также подчеркнуть, что еще задолго до появления в лингвистике термина «интерференция», играющего на сегодняшний день ключевую роль в теории языковых контактов и учении о билингвизме, мысль о взаимовлиянии языков была сформулирована основателем Казанской лингвистической школы И. А. Бодуэном де Куртене, который объяснял влияние языков друг на друга конвергентной перестройкой языков в ходе коммуникативных контактов:

«Влияние смешения языков проявляется в двух направлениях: с одной стороны, оно вносит в данный язык из чужого языка свойственные ему элементы (запас слов, синтаксические обороты, формы, произношение); с другой же стороны, оно является виновником ослабления степени и силы различаемости, свойственной отдельным частям данного языка»¹. Идеи И. А. Бодуэна де Куртенэ, которые в свою очередь независимо от русского лингвиста также сформулировал в Германии Г. Шухардт [Шухардт 2010], получили распространение в 1920–1930-х гг. в работах основателей Пражского лингвистического кружка Н. С. Трубецкого и Р. О. Якобсона. Благодаря их вкладу термин «интерференция» получил широкое распространение в лингвистике. Позднее к вопросу о взаимном влиянии языков обратились многие ученые, в числе которых следует особо отметить Л. В. Щербу, Е. Д. Поливанова, А. Мартине и У. Вайнрайха. Именно благодаря фундаментальной работе последнего были заложены основы изучения языковых контактов, а термин «лингвистическая интерференция» прочно закрепился в области лингвистики.

Концепции рассмотрения феномена интерференции в отечественной и зарубежной лингвистике

В своем труде «Языковые контакты» У. Вайнрайх пишет об интерференции как о явлении, находящем отражение на всех уровнях языка: фонетическом, морфологическом, синтаксическом и семантическом. У. Вайнрайх определял интерференцию как случаи отклонения от норм языка, которые возникают в речи билингва в результате его знакомства с более чем одним языком, т. е. в результате языкового контакта [Вайнрайх 1979]. Преобладающее большинство лингвистов, в числе которых А. Е. Карлинский, В. А. Виноградов, М. М. Михайлов, В. Н. Ярцева, Л. В. Щерба, В. Ю. Розенцвейг и М. Ю. Розенцвейг, говоря о природе интерференции, подчеркивали ее негативный характер, представляя интерференцию как «нарушение» в речи билингва. Наиболее близкую к позиции У. Вайнрайха сформулировал в отечественной лингвистике В. Ю. Розенцвейг, согласно определению которого интерференция – «это нарушение билингвом правил соотнесения контактирующих языков, которое проявляется в его речи в отклонении от нормы» [Розенцвейг 1972, с. 28]. М. Ю. Розенцвейг подчеркивал факт нарушения нормы контактирующих языков вследствие отклонения

¹Цит. по: [Алимов 1998, с. 7].

«от правил соотнесения родного и неродного языков в сторону их уподобления» [Розенцвейг 1975, с. 12]. А. Е. Карлинский пишет об интерференции как о частном случае «речевых мутаций», заключающемся в бессознательном переносе навыков из родного языка в иностранный [Карлинский 1980]. Л. В. Щерба хотя и не употребляет этот термин, в своей статье «К вопросу о двуязычии» говорит о «взаимном искажении обоих языков, на практике, при изучении иностранного языка, искажении именно этого иностранного языка под влиянием родного» [Щерба 1974].

Е. В. Шинкаренко отмечает сознательное или бессознательное смешивание систем двух языков, используемых билингвом в момент говорения, вследствие чего происходит отклонение от общепринятых норм языка [Шинкаренко 1995]. Н. Б. Мечковская характеризует интерференцию как «ошибки в речи на иностранном языке, вызванные влиянием системы родного языка» [Мечковская 2000, с. 196]. Несколько иной точки зрения придерживаются в своих работах В. В. Потапов и Р. К. Потапова, отмечая, что ошибки в речи на иностранном языке являются *следствием* интерференции со стороны родного языка (выделено авт.) [Потапов 1994; Potapov 2003; Potapov 2012; Potapova, Potapov 2011; Potapova, Potapov 2013].

Некоторые авторы, напротив, отмечают в своих работах положительное влияние интерференции, говоря о ней с точки зрения переноса знаний общетипологического свойства при взаимодействии контактирующих языков, при котором не оказывается отрицательного влияния на неродной язык билингва [Ершова 1972; Назарова 1972].

Однако трактовка понятия интерференции не ограничивается амбивалентной оценкой, подчеркивающей ее исключительно негативный или позитивный характер. Существуют также работы, посвященные нейтральному характеру этого явления, в которых интерференция понимается как двусторонний процесс [Семчинский 1974] и «отмечается сам факт проникновения одного языка в другой» [Вишневская 1993, с. 37; Любимова 1991; Maskey 1967; Clyne 1967; Мартине 1963]. У. Маккей определял интерференцию как использование черт одного языка в письменной или устной речи на другом языке [Maskey 1967]. Э. Хауген рассматривал интерференцию как лингвистическое переплетение, при котором любая лингвистическая единица может оказаться одновременно элементом двух систем [Haugen 1956]. Подобное наложение позволяет использовать билингву все языковые средства

обоих контактирующих языков, позволяя легче справиться с коммуникативной задачей. К. К. Платонов пишет об «интерференции навыков», способной как ослаблять, так и усиливать новоприобретенные навыки под влиянием уже имеющихся [Платонов 1972]. В. Н. Комиссаров в работе «Проблема интерференции в теории перевода» пишет о том, что различные виды интерференции при переводе могут оказывать как положительное, так и отрицательное влияние [Комиссаров 1989]. Л. И. Баранникова в своих работах подчеркивает многогранный характер явления интерференции, отмечая, что «проблема интерференции не является чисто лингвистической, ибо интерференция не только один из фактов языка, а явление, имеющее глубокие психологические и, в конечном счете, социально-исторические корни» [Баранникова 1972, с. 88].

В своей диссертации, посвященной интерференции и акценту, Г. М. Вишневская формулирует наиболее распространенную трактовку понятия интерференции как процесса и результата «взаимодействия языковых систем в речи билингва, из которых одна система является доминирующей, порождающей эффект воздействия со вторичной, приобретенной языковой системой», однако подчеркивает, что на сегодняшний день единого понимания данного термина и основных понятий теории интерференции не существует в связи с отсутствием единой методологии и методики изучения вышеописанных явлений: «В теории языковых контактов выработана своя терминологическая система <...>. Некоторые термины толкуются авторами исследований порой противоречиво. Именно поэтому ... мы считаем, что некоторые понятия теории языковых контактов, отраженные в терминах «билингвизм», «интерференция», «иностранноязычный акцент», нуждаются в специальном теоретическом осмыслении» [Вишневская 1993, с. 36]. Подтверждением этой точки зрения может служить определение В. А. Виноградова, в котором он, с одной стороны, опирается на общепринятую трактовку У. Вайнрайха, а с другой – дифференцирует понятия «отклонение от нормы» и «интерференция». В. А. Виноградов определяет интерференцию как «взаимодействие языковых систем в условиях двуязычия, складывающегося либо при языковом контакте, либо при индивидуальном усвоении неродного языка», объясняя отклонения от нормы в неродном языке влиянием процесса интерференции [Виноградов 1990, с. 197]. Схожей позиции придерживается В. Ю. Розенцвейг, который характеризует интерференцию

как «нарушение билингом правил соотношения контактирующих языков, которое проявляется в его речи в отклонении от нормы», объясняя это явление тем, что «уже сложившиеся навыки затрудняют образование новых, либо снижают их эффективность» [Розенцвейг 1972, с. 70].

Из работ Э.М. Ахунзянова следует, что интерференция и есть отклонение от нормы: «...интерференция не обязана всегда оставаться интерференцией, т. е. отклонением от нормы: при частом повторении она сама становится нормой» [Ахунзянов 1978, с. 83]. Однако исследователь предлагает считать интерференцию положительным или нейтральным явлением, обусловленным объективными закономерностями, а факты неосознанного, ошибочного переноса норм первого языка на второй именовать трансференцией, которая рассматривается как отрицательное явление, порождающее избыточность в речи, создающее помехи в освоении второго языка и носящее случайный характер. Прямо противоположную интерпретацию тех же терминов дает С.Н. Тимофеев, предлагая различать интерференцию как отрицательное явление и трансференцию – как положительное [Тимофеев 1988]. Подобную точку зрения в отношении трансференции как положительного переноса, в результате которого при совпадении норм контактирующих языков действуют их общетипологические свойства, не оказывающие отрицательного влияния на иноязычную речь билинга, разделяют и другие исследователи [Карлинский 1990; Любимова 1991; Наумова, Хромов 2002]. М.В. Щербакова отмечает, что при трансференции «воздействующий язык не вызывает в изучаемом языке нарушений нормы, но стимулирует уже существующие в нем закономерности» [Щербакова 2003, с. 85]. В зарубежной лингвистике термин «трансференция» часто употребляется в контексте переноса как специфических, так и общетипологических черт контактирующих языковых систем [Clyne 1967; Jenkins 2000].

Разночтения в понимании и употреблении термина «интерференция» и смежных понятий в зарубежных и отечественных лингвистических школах

В работах зарубежных лингвистов термин «интерференция» чаще всего употребляется как «(language) transfer», однако имеются и другие эквиваленты: L1 interference, linguistic interference, crosslinguistic influence, transference, interference. Анализ источников

по теме исследования позволяет предположить, что эти термины используются для разграничения понятий, отвечающих различным подходам к явлению интерференции. Наиболее употребительным является нейтральный термин «(language) transfer», который трактуется как процесс переноса элементов одного языка на другой язык, другими словами – взаимоперенос языковых черт в речи билингва или мультилингва [Grosjean & Py 1991; Jarvis & Pavlenko, Scott & Aneta 2008; Cenoz, Hufeisen, and Jessner 2001; Cook 2003; Treffers-Daller, Sakel 2012].

В настоящее время явление интерференции вызывает особый интерес у зарубежных ученых, занимающихся вопросами усвоения иностранного языка – области, которая в англоязычных странах рассматривается как самостоятельная дисциплина в рамках *прикладной лингвистики*. Наиболее распространенной формулировкой можно назвать дефиницию Р. Эллиса, в которой интерференция представляется как включение особенностей родного языка (L1) в систему знаний иностранного языка (L2), которую обучающийся пытается сформировать [Ellis 1994].

Понятие произносительной интерференции эволюционировало в зарубежной лингвистике и области усвоения второго языка в течение длительного периода, что позволило взглянуть на это явление с разных сторон. В 50–60 гг. XX в. в трудах зарубежных специалистов в области усвоения второго языка интерференция (language transfer) воспринималась исключительно как явление, негативно сказывающееся на усвоении иностранного языка и вызывающее основные трудности при его изучении. Позднее, в 1970-х гг. отношение к данному явлению было критичным. Однако со временем позиция изменилась в связи с тем, что многие ученые признали факт того, что при своей «раздражающей неуловимости», интерференция (transfer) является одним из основных факторов, определяющих межъязыковую компетенцию и эффективность обучающегося [Kohn 1986, с. 21].

Со временем границы восприятия и интерпретации понятия «интерференция» расширились. Если раньше вопрос рассматривался в рамках двух языков, где родной позиционировался как доминантный, то позднее была выдвинута гипотеза о том, что интерференция не ограничивается влиянием только родного языка, а также возможна и в случае взаимовлияния большего количества языков [Odlin 1989]. На сегодняшний день интерпретация зарубежных англоязычных

лингвистов во многом схожа с интерпретацией в отечественных исследованиях, что позволяет раскрыть сложный и многогранный характер интерференции, как явления двустороннего взаимодействия, имеющего как негативное, так и позитивное влияние в условиях искусственного билингвизма.

Французский психолингвист Ф. Грожан использует в своей статье «Попытка изолировать и затем дифференцировать понятия *transfer* и *interference*» вышеназванные термины как синонимичные, однако при этом отмечает необходимость их разграничения с учетом фактора перманентности [Grosjean 2011]. По мнению Ф. Грожана, трансференция в основном статична, поскольку она отражает постоянные «следы» одного языка в другом (например, в случае наличия иностранного акцента), тогда как интерференция динамична, поскольку она затрагивает элементы другого языка, которые случайно «проскальзывают» в письменную или устную речь [Grosjean 1998; Grosjean 2008; Grosjean 2011]. Отечественные лингвисты Р. К. Потапова и В. В. Потапов в своих работах также отмечают, что в случае переключения с одного языка на другой языковые характеристики говорящего могут иметь динамический или статический характер, и для идентификации говорящего важно выявить именно те элементы речи говорящего, которые носят постоянный характер [Potapova, Potapov 2013].

Американский лингвист и специалист в области языка жестов Дж. Дэйвис прибегает в своих работах к употреблению терминов «трансференция» (*transference*) и «интерференция» (*interference*), где первая рассматривается как стратегия, сознательный перенос норм родного языка на изучаемый во избежание недопонимания и неопределенности, а вторая – как отдельные случаи влияния родного языка, проявляющиеся в речи в виде ошибок и отклонений от нормы в речи на иностранном языке [Davis 1990]. Соотечественница Дж. Дэйвиса, лингвист С. Томасон, напротив, избегает употребления термина «*transference*», активно используя в своей работе понятие «интерференция» (*interference*) и лишь изредка прибегает к употреблению термина «*transfer*» [Thomason 2001].

Австралийский лингвист М. Клайн использует в своих работах оба термина, однако, если трансференцией (*transference*) он называет «приспособление любых элементов или черт из другого языка», то понятие «*transfer*» (перенос) М. Клайн в качестве примера трансференции (*transference*). Таким образом, согласно М. Клайну, лексическая

или морфосемантическая трансференция – это перенос лексических единиц с сохранением изначальной формы и значения, фонологическая трансференция – перенос звуков, морфологическая трансференция – перенос словоформы с изменением содержания, смоделированной по морфологической картине другого языка, а семантическая трансференция означает перенос значения без заимствования словоформы [Clyne 1967].

Заключение

Исходя из вышеизложенного, необходимо отметить, что несмотря на большой интерес к данному явлению, вопросы природы возникновения психических процессов, оказывающих на это влияние, требуют дальнейших исследований, а терминология нуждается в унификации, так как термин «interference» (интерференция) в трудах многих лингвистов имеет скорее негативную коннотацию, ассоциируясь с «тем негативным в языке, от чего необходимо избавиться»¹ [Jenkins 2000]. Вследствие этого значительно чаще используется термин «(language) transfer», который может иметь как негативный характер (negative transfer), что можно было бы соотнести с понятием «интерференция» в отечественной лингвистике, так и позитивный характер (positive transfer), что ближе по значению к понятию трансференции. Подобным образом использует в своей работе эти термины И. Шатц [Shatz 2017].

Предполагается, что в связи с необходимостью унификации терминологии, связанной с понятиями «интерференция» (interference), «трансференция» (transference) и «трансфер» (language) transfer, как в отечественной, так и в зарубежной лингвистике, необходимо проведение дальнейших исследований, направленных на возможность более четкой дифференциации данных явлений с учетом природы их возникновения и реализации в акте коммуникации. Анализ материала для исследования позволяет заключить, что наиболее уместным для описания процесса переноса как специфических, так и общетипологических языковых черт одного языка на другой при их взаимодействии, по-видимому, является в настоящее время термин «трансференция», который созвучен с англоязычным термином «(language) transfer», имеющим нейтральную коннотацию. Понятие «трансфер» (transfer) в таком случае предлагается употреблять как «перенос»,

¹Цит. по: [Колесниченко 2013, с. 184].

т. е. механический процесс заимствования элементов из одного языка (L1) в другой язык (L2). Исходя из этого, термин «интерференция» можно трактовать как нежелательное явление результата переноса («трансфера»), другими словами – влияние родного языка (L1) на иностранный язык (L2).

СПИСОК ЛИТЕРАТУРЫ

- Алимов В. В.* Явление лингвистической интерференции при изучении специального перевода: дис. ... канд. филол. наук. М., 1998. 196 с.
- Ахунзянов Е. М.* Двужычие и лексико-семантическая интерференция. Казань: Изд-во КГУ, 1978. 187 с.
- Баранникова Л. И.* Сущность интерференции и специфика ее проявления. Проблема двужычия и многоязычия. М. : Наука, 1972. 94 с.
- Большой психологический словарь / под ред. Б. Г. Мещерякова, В. П. Зинченко. М. : АСТ. СПб. : Прайм-Евразнак, 2008. 868 с.
- Вайнрайх У.* Языковые контакты: состояние и проблемы исследования / пер. с англ. и коммент. Ю. А. Жлуктенко. Киев : Вища шк., 1979. 263 с.
- Виноградов В. А.* Интерференция. Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева. М. : Сов. энциклопедия, 1990. 685 с.
- Вишневская Г. М.* Интерференция и акцент (на материале интонационных ошибок при изучении неродного языка): дис. ... д-ра филол. наук. СПб, 1993. 373 с.
- Ершова Е. Н.* О некоторых формах лингвистической интерференции и о возможности их использования в методике преподавания языков // Проблемы двужычия и многоязычия. М. : Наука, 1972. С. 130–139.
- Карлинский А. Е.* Основы теории взаимодействия языков и проблема интерференции: дис. ... д-ра филол. наук. Алма-Ата, 1980. 434 с.
- Карлинский А. Е.* Основы теории взаимодействия языков. Алма-Ата : Голым, 1990. 181 с.
- Колесниченко М. А.* К вопросу об интерференции в теории языковых контактов // Язык и культура. Новосибирск : Центр развития научного сотрудничества. 2013. № 9. С. 182–186.
- Комиссаров В. Н.* Проблема интерференции в теории перевода // *Interferenz in der Translation*. Herausgegeben von Heide Schmidt VEB VerlagEnzyklopädie. Leipzig, 1989. С. 103–108.
- Любимова Н. А.* Фонетическая интерференция и общение на неродном языке (Экспериментальное исследование на материале финско-русского двужычия): дис. ... д-ра филол. наук. Л., 1991. 375 с.
- Мартине А.* Основы общей лингвистики // Новое в лингвистике. 1963. Вып. 3. 568 с

- Мечковская Н. Б.* Социальная лингвистика. 2-е изд. М. : Аспект-Пресс, 2000. 208 с.
- Назарова М. К.* К педагогическому аспекту двуязычия. // Проблемы двуязычия и многоязычия. М. : Наука, 1972. С. 12–25.
- Наумова Е. В., Хромов С. С.* Экспериментально-фонетическое исследование просо- дико-интонационной интерференции в русской речи билингвов // Фонетика в системе языка: сб. статей. М. : РУДН, 2002. Вып. 3. Ч. 2. С. 37–46.
- Платонов К. К.* О системе психологии. М. : Мысль, 1972. 216 с.
- Потапов В. В.* Опыт лингвоконтрастного исследования фонетических систем английского и русского языков // Филологические науки. 1994. № 2. С. 88–99. (Научные доклады высшей школы.)
- Розенцвейг М. Ю.* Проблемы языковой интерференции: автореф. дис. ... д-ра филол. наук. М., 1975. 50 с.
- Розенцвейг В. Ю.* Языковые контакты: Лингвистическая проблематика. Л. : Наука, 1972. 80 с.
- Семчинский С. В.* Семантическая интерференция языков (на материале славяно-восточнороман. языковых контактов): автореф. дис. ... д-ра филол. наук. Киев, 1974. 56 с.
- Тимофеев С. Н.* Межъязыковая интерференция и установление конкретных случаев ее проявления при контакте французского и русского языков: (на уровне звуков) // Теоретические и практические проблемы прикладной лингвистики. М., 1988. С. 115–136.
- Шинкаренко Е. В.* Явление лексической интерференции в украинской речи молодежи Кировоградчины : автореф. дис. ... канд. филол. наук. Киев, 1995. 18 с.
- Шухардт Г.* Избранные статьи по языкознанию / пер. с нем. М. : Едиториал УРСС, 2010. 296 с.
- Щерба Л. В.* Языковая система и речевая деятельность. Л. : Наука, 1974. 427 с.
- Щербакова М. В.* Особенности обучения второму иностранному языку студентов языковых факультетов // Вестник ВГУ. 2003. № 2. С. 85–88. (Лингвистика и межкультурная коммуникация.)
- Энциклопедический словарь Ф. А. Брокгауза и И. А. Ефрона. URL: www.vehi.net/brokgauz/
- Cenoz J., Hufeisen, B., & Jessner, U.* Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives. Clevedon : Multilingual Matters, 2001. 197 p.
- Clyne M. G.* Transference and triggering. Observations on the language assimilation of postwar German-speaking migrants in Australia. The Hague : Nijhoff, 1967. 148 p.

- Cook V.* Effects of the L2 on the L1. Clevedon : Multilingual Matters, 2003.
- Davis Jeffrey E.* Linguistic transference and interference: Interpreting between English and ASL // C. Lucas (Ed.). Sign Language Research : Theoretical Issues. Washington, DC : Gallaudet University Press, 1990. P. 308–321.
- Ellis R.* The Study of Second Language Acquisition, Oxford : Oxford University Press, 1994.
- Grosjean F.* ‘Transfer and language mode. Commentary on N. Müller “Transfer in bilingual first language acquisition”’. Bilingualism: Language and Cognition. 1. 1998. P. 175–176.
- Grosjean F.* Studying Bilinguals, Oxford: Oxford University Press, 2008.
- Grosjean F.* An attempt to isolate, and then differentiate, transfer and interference, International Journal of Bilingualism. 16 (1). 2011. P. 11–21.
- Grosjean F., & Py B.* La restructuration d’une première langue: l’intégration de variantes de contact dans la compétence de migrants bilingues. La Linguistique. 27. 1991. P. 35–60.
- Haugen E.* Bilingualism in the Americas: A bibliography and research guide. Publications of the American dialect society. No. 36. Alabama : University of Alabama Press, 1956.
- Jarvis & Pavlenko, Scott & Aneta.* Crosslinguistic Influence in Language and Cognition. Abingdon : Routledge, 2008. 304 p.
- Jenkins J.* The Phonology of English as an International Language. Oxford : Oxford University Press, 2000. 258 p.
- Kohn K.* ‘The analysis of transfer’ // E. Kellerman & M. Sharwood Smith (eds), Crosslinguistic Influence in Second Language Acquisition, New-York : Pergamon Press, 1986. P. 21–34.
- Mackey W. F.* Bilingualism as a world problem. Montreal : Harvest House, 1967. 119 p.
- Odlin T.* Language Transfer: Cross-linguistic Influence in Language Learning. Cambridge : Cambridge University Press, 1989. 258 c.
- Potapov V. V.* Some aspects of Germans acquisition of the Russian language // Вестник МГЛУ. 2012. Вып. 13 (646). С. 136–149. (Языкознание.)
- Potapov V. V.* On language contrastive-comparative analysis of English and Russian phonetic systems // Proc. of the intern. workshop “Speech and computer” (SPECOM’2003). М., 2003. P. 276–282 ; *Potapov V. V.* The American English interference in Russian on the segmental level // SPECOM’2003. International Workshop “Speech and Computer”. MSLU, 2003. P. 283–290.
- Potapova R. K., Potapov V. V.* Akustisch-perzeptorische Analyse der gesprochenen Sprache in Experimenten mit sprechsprachlichem Code-switching «russisch-deutsch» und «deutsch-russisch» // Kommunikative Sprechfähigkeit. Russland

- und Deutschland im Vergleich. Köln–Weimar–Wien : Böhlau Verlag, 2011. S. 249–255.
- Potapova R.K., Potapov V.V.* The mechanisms of Russian-German speech communication and code-switching // *Вестник МГЛУ*. 2013. Вып. 13 (673). С. 9–18. (Языкознание.)
- Shatz I.* Native Language Influence During Second Language Acquisition: A Large-Scale Learner Corpus Analysis (PDF). Proceedings of the Pacific Second Language Research Forum (PacSLRF 2016). Hiroshima, Japan : Japan Second Language Association. Retrieved 10 September 2017. P. 175–180.
- Thomason S. G.* Language contact. An introduction. Edinburgh : Edinburgh University Press. 2001. 310 p.
- Treffers-Daller J., Sakel J.* Why transfer is a key aspect of language use and processing in bilinguals and L2-users // *International Journal of Bilingualism*, 16 (1). 2012. P. 3–10.

УДК 81'33

Р. К. Потапова, Н. А. Забелло, Н. В. Бобров

Р. К. Потапова, доктор филологических наук, профессор; действительный член Международной академии информатизации; заведующая кафедрой прикладной и экспериментальной лингвистики; директор Института прикладной и математической лингвистики МГЛУ; e-mail: rkpotarova@yandex.ru

Н. А. Забелло, аспирант кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики факультета английского языка МГЛУ; e-mail: natangens@gmail.com

Н. В. Бобров, ст. научный сотрудник НИР кафедры прикладной и экспериментальной лингвистики факультета английского языка МГЛУ; e-mail: azamath@list.ru

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК РЕЧИ В СОСТОЯНИИ НАРКОТИЧЕСКОЙ ИНТОКСИКАЦИИ (на основе перцептивно-слухового анализа)

Статья посвящена анализу информативных устойчивых перцептивно-слуховых характеристик, выполняющих функцию обнаружения идентификационных признаков речи в состоянии наркотической интоксикации. Исследование проводилось на основе перцептивно-слухового эксперимента, материалом служил корпус фонограмм речи дикторов-женщин – носителей русского языка в состоянии наркотической интоксикации и абстиненции. Методика перцептивно-слухового анализа основана на работах основателя научной школы фундаментального и прикладного речеведения Р. К. Потаповой. Из корпуса фонограмм (N=130) общей длительностью 10 ч. 14 мин. с участием 40 дикторов были отобраны 23 фонограммы (общей длительностью 250 мин. 12 сек, всего 11 дикторов) спонтанной речи в виде диалога. Далее указанные фонограммы были предложены группе испытуемых-аудиторов, участвующих в перцептивно-слуховом эксперименте. Испытуемыми стали 24 человека (носители русского языка, студенты магистратуры, обучающиеся по направлению «Теоретическая и прикладная лингвистика»). Каждый испытуемый заполнял анкету, отражающую его индивидуально-социальные характеристики. Прослушивая каждую фонограмму необходимое количество раз, испытуемым предлагалось ответить на вопросы специальных анкет, отражающие: высоту голоса, диапазон изменений высоты голоса, мелодический рисунок, темп речи, паузы, силу голоса, диапазон варьирования, изменения силы голоса, тембр, речевой ритм, различие ударных и безударных слогов во фразе, речевое дыхание, четкость произношения.

В статье также приведены результаты проведенного перцептивно-слухового эксперимента.

Ключевые слова: перцептивно-слуховой анализ; наркотическая интоксикация; фонетические характеристики говорящего.

R. K. Potapova

Sc.D., Prof., Full Member of the International Informatization Academy, Head of Department of Applied and Experimental Linguistics, Director of Institute for Applied and Mathematical Linguistics of Moscow State Linguistic University (MSLU);
e-mail: rkpotapova@yandex.ru

N. A. Zabello

Postgraduate Student, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of English, MSLU;
e-mail: natangens@gmail.com

N. V. Bobrov

Senior Research Fellow, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the Humanities and Applied Sciences, MSLU;
e-mail: azamath@list.ru

**RESEARCH ON SPEECH CHARACTERISTICS UNDER DRUG
INTOXICATION (based on auditive analysis)**

The article is devoted to the analysis of informative stable auditive characteristics enabling to detect speech in the state of narcotic intoxication. The study was conducted on the basis of a auditive experiment using the corpus of audio recordings of female native Russian speakers in the state of drug intoxication and abstinence. The method of auditive analysis is based on the works of R.K. Potapova, the founder of the scientific school of fundamental and applied speech studies. From the corpus consisting of audio recordings (N = 130) with a total duration of 10 hours and 14 minutes with the involvement of 40 female speakers, 23 audio recordings (250 minutes 12 seconds, 11 speakers in total) were selected of spontaneous speech in the form of a dialogue. Further, the audio recordings were proposed to a group of subjects participating in the auditive experiment. The subjects were 24 people (native Russian speakers, Master's Degree students studying theoretical and applied linguistics). Each subject filled in a questionnaire reflecting his/her individual social characteristics. Listening to each audio recording the required number of times, the subjects were asked to answer the questions of special questionnaires, reflecting: pitch, the range of pitch variation, melodic pattern, speech rate, pauses, voice strength, voice strength variation range, timbre, speech rhythm, differences in stressed and unstressed syllables in the phrase, speech breathing, and accuracy of pronunciation. The article also presents the results of the auditive experiment.

Key words: auditive analysis; drug intoxication; speaker's phonetic features.

Данное исследование проводилось в рамках проблемы идентификации личности, а также ее эмоционального и психофизиологического состояния [Hollien 1990; Hollien 2001; French 1994; Klasmeyer, Sendlmeyer 1997; Neuhauser 2013; Potapova, Potapov 2016].

Методика перцептивно-слухового анализа основана на работах основателя научной школы фундаментального и прикладного речеведения Р. К. Потаповой [Потапова, Потапов 2006; Междисциплинарность в исследовании 2015; Потапова, Потапов 2016] и заключается в проведении перцептивно-слухового эксперимента на аутентичном материале с привлечением групп испытуемых-аудиторов.

Отбор речевого материала проводился поэтапно. На первом этапе был разработан устно-речевой корпус фраз на русском языке, позволяющий провести перцептивно-слуховой и акустический анализ различных видов речи: спонтанной (коммуникации), квазиспонтанной (пересказа текста) и чтения в состояниях интоксикации различной природы, абстиненции и трезвости. Речевой материал был собран на базе Научно-исследовательского института наркологии – филиала ФГБУ «Федеральный медицинский исследовательский центр психиатрии и наркологии имени В. П. Сербского» Минздрава России с помощью цифрового линейного РСМ диктофона марки «Olympus LS-11E с контроллером». Частотный диапазон в режиме РСМ составляет 44,1–96 кГц. Расстояние между микрофоном диктофона и испытуемым составляет 50–70 см. В результате был получен корпус фонограмм (N = 130) общей длительностью 10 часов 14 минут с участием 40 дикторов.

На следующем этапе были отобраны 23 фонограммы (общей длительностью 250 мин. 12 сек, всего 11 дикторов) спонтанной речи в виде диалога. Далее указанные фонограммы были предложены группе испытуемых-аудиторов, участвующих в перцептивно-слуховом эксперименте.

В перцептивно-слуховом исследовании участвовали 24 человека. Формирование группы испытуемых-аудиторов проводилось с учетом следующих критериев:

- отбирались носители русского языка;
- участники эксперимента – студенты магистратуры, обучающиеся по направлению «Теоретическая и прикладная лингвистика».

Перед началом эксперимента с каждым испытуемым проводилась беседа, информирующая о процедуре исследования. Участие в исследовании являлось добровольным. Каждый испытуемый заполнял анкету, отражающую его индивидуально-социальные характеристики (см. табл. 1). Прослушивая каждую фонограмму необходимое количество раз, испытуемым предлагалось ответить

на вопросы специальных анкет (см. табл. 2). У них было право останавливаться для отдыха. Задания выполнялись строго индивидуально.

Таблица 1

Анкета аудитора

Фамилия Имя Отчество _____

Пол _____

Возраст (год рождения) _____

Место рождения _____

Образование (какую школу окончил(а) и где) _____

Специализация по образованию (например, лингвист) _____

Ваш родной язык _____

Изучали ли Вы иностранный язык ранее? _____

Какой, где, в течение какого времени? _____

Какой (какие) иностранный язык изучаете в настоящее время? _____

Имеете ли Вы музыкальное образование? Если да, то сколько времени обучались музыке (пению)? _____

Место(а) длительного проживания (от 5 лет и более) _____

Профессии родителей _____

Таблица 2

Анкета перцептивно-слухового анализа – фонетические характеристики говорящего (по Р. К. Потаповой)

Высота голоса	очень низкий
	низкий
	средний
	выше среднего
	высокий
	очень высокий
Определить высоту голоса по шкале от 0 до 10 Допускаются десятые и сотые доли	

Диапазон изменений высоты голоса	очень широкий
	широкий
	средний
	узкий
	очень узкий
Определить диапазон изменений высоты голоса по шкале от 0 до 10	
Мелодический рисунок	плавный (легато)
	скачкообразный, изрезанный (стаккато)
	монотонный
	индивидуальная специфика мелодического рисунка отдельных участков фразы (в начале, середине, конце)
Темп речи	очень медленный
	медленный
	средний
	быстрый
	очень быстрый
	вариативность темпа на разных участках фразы (в начале, середине, конце)**
	регулярность вариативности
Определить темп речи по шкале от 0 до 10. Допускаются десятые и сотые доли	
Паузы	очень длинные
	длинные
	средние
	короткие
	очень короткие
	наличие / отсутствие пауз хезитации (если есть, то указать их число)
	наличие заполненных пауз (звуки: гласные, гласноподобные, сонорные /м,н/, слоги перед паузой или словом, слова-паразиты, подкашливание, причмокивание, повтор одних и тех же слогов или слов) наличие незаполненных пауз (есть 1, нет 0)
Определить частоту пауз по шкале от 0 до 10	

Сила голоса	очень слабый
	средний
	сильный
	очень сильный
	варьирование силы голоса
Диапазон варьирования	узкий
	средний
	широкий
Изменения силы голоса	ровные, равномерные
	резкие, скачкообразные
	определение участков фразы, на которых реализуются скачки по силе голоса (начало, середина, конец)
Окраска голоса (тембр)	звонкий
	глухой
	бархатный
	хриплый
	чистый
	мягкий
	назализованный
	сиплый
	певучий
	резкий
	жесткий (голос курящего человека)
	скрипучий
	приятный
	неприятный
	сексапильный
с элементами актерского исполнения	
наличие смены разных тембров	
Насколько различаются ударные и безударные слоги во фразе	сильно различаются
	небольшая разница
	не различаются

За счет каких характеристик	изменение высоты тона голоса
	изменение громкости голоса
	изменение временных характеристик слогов
Как Вы можете охарактеризовать речевой ритм	сверхчеткий
	четкий (чеканный)
	нечеткий (смазанный)
	скандированный
Речевое дыхание	нормальное
	прерывистое
	дискомфортное
Произношение звуков и слогов	очень четкое
	четкое
	нечеткое
	совсем нечеткое

Высота голоса варьируется от низкого до выше среднего, отмечается средний либо узкий диапазон изменения высоты голоса. Мелодический рисунок преимущественно скачкообразный и изрезанный, темп речи в основном медленный или средний, паузы длинные (в том числе заполненные и паузы хезитации).

Силу голоса рассматривают как среднюю (61,9 %), как очень слабую в 20 % случаев. Значительно реже речь дикторов оценивалась как сильная и очень сильная, а также наблюдалось варьирование силы голоса. Диапазон варьирования часто узок, а изменения силы голоса плавные. Ни у одного из дикторов не был зафиксирован широкий диапазон варьирования силы голоса.

В описании тембра превалируют следующие оттенки голоса (в порядке убывания): сильный, хриплый, жесткий.

Большая часть аудиторов посчитали, что разница между ударными и безударными слогами небольшая (81 %) за счет изменения высоты тона (23,8 %), громкости (42,6 %) голоса и временных характеристик слогов (28,6 %). Речь дикторов в основном расценивалась как смазанная (71,4 %), в остальных случаях как четкая.

Речевое дыхание, по оценкам аудиторов, равномерно распределилось между нормальным и дискомфортным (по 38,1 %), в остальных же случаях (23,8 %) оценивалось как прерывистое. Произношение звуков и слогов в основном нечеткое (76,2 %).

СПИСОК ЛИТЕРАТУРЫ

- Потапова Р. К., Потапов В. В.* Язык, речь, личность. М.: Языки славянской культуры, 2006. 496 с.
- Междисциплинарность в исследовании речевой полиинформативности / *Р. К. Потапова, В. В. Потапов, Н. Н. Лебедева, Т. В. Агibalова*; под ред. Р. К. Потаповой. М.: Языки славянской культуры, 2015. 352 с.
- French P.* An overview of forensic phonetics with particular reference to speaker identification // *The International Journal Forensic Linguistics: Speech, Language and the Law.* 1994. Vol. 1. Issue 2. P. 169–182.
- Hollien H.* *The Acoustics of Crime: the New Science of Forensic Phonetics.* New York : Plenum., 1990. 370 p.
- Hollien H.* *Forensic Voice Identification.* N. Y. : Academic Press, 2001. 240 p.
- Klasmeyer G., Sendlmeyer W. F.* The Classification of Different Phonation Types in Emotional and Neutral Speech. *Forensic Linguistics.* 1997. 4 (1). P. 104–124.
- Neuhauser S.* Phonetische und linguistische Prinzipien des forensisches stimmenvergleichs (Author Michael Jessen) // *The International Journal: Speech, Language and the Law.* 2013. Vol. 20.2. P. 325–330.
- Potapova, R., Potapov, V.* On Individual Polyinformativity of Speech and Voice regarding Speakers Auditive Attribution (Forensic Phonetic Aspect)// *A. Ronzhin, R. Potapova, G. Nemeth (Eds) Speech and Computer. Proceedings of the 18th International Conference SPECOM'2016.* 2016. P. 507–514.

УДК 81'34

Р. К. Потапова

доктор филологических наук, профессор;
действительный член Международной академии информатизации;
заведующая кафедрой прикладной и экспериментальной лингвистики;
директор Института прикладной и математической лингвистики МГЛУ;
e-mail: rkpotapova@yandex.ru

КОГНИТИВНО-СЕМИОТИЧЕСКАЯ МОДЕЛЬ РЕЧЕВОЙ ДЕЯТЕЛЬНОСТИ (применительно к принципу экономии)

В данной статье в качестве основных вопросов рассматриваются формирование и природа когнитивно-семиотической модели речевой деятельности применительно к принципу экономии. Рассмотрение принципа экономии включает семиотические подсистемы речевой коммуникации, относящиеся к семиотической макросистеме порождения речевых высказываний. Разработанная семиотическая генеративная речевая модель включает 12 фаз процесса порождения устной речи, начиная от эвокативной невербальной фазы порождения речи на нейронном уровне до смыслового понимания высказывания в целом.

Начальные фазы включают нейронную программу, соотносящуюся с артикуляторной координацией, построением последовательностей сегментов и других компонентов речевого продукта.

С помощью данной модели возможно изучение всех фаз порождения речи в терминах семиотических подсистем на базе главной семиотической коммуникативной макросистемы, а также установления связи между уровнями реализации формулы «Competence – performance» применительно к речевому продукту в форме специфических (индивидуальных) высказываний носителей языка и принципом экономии, что прежде всего имеет непосредственное отношение к фоностилистике (стилистике дискурса), видам речевой активности коммуникантов и т. д. Таким образом, данная модель, построенная на базе речевой деятельности и принципа экономии, соотносится со стилистическим аспектом, включающим стилистическую вариативность речи и изменение поведенческого кода человека в целом.

Ключевые слова: принцип экономии; когнитивно-семиотическая модель; речевая коммуникация; речевая деятельность.

R. K. Potapova

Sc.D., Prof., Full Member of the International Informatization Academy;
Head of Department of Applied and Experimental Linguistics;
Director of Institute for Applied and Mathematical Linguistics
of Moscow State Linguistic University (MSLU);
e-mail: rkpotapova@yandex.ru

COGNITIVE-SEMIOTIC MODEL OF SPEECH-ACTIVITIES (REGARDING THE ECONOMY-PRINCIPLE)

The main points for consideration in this paper will be formation and nature of the cognitive-semiotic model of speech activities in regard to the economy-principle.

Under consideration economy principle concerns semiotic subsystems of the main speech communication system regarding the speech utterances. This semiotic generative speech model includes twelve phases of the speech generation process: from evocative nonverbal phase of speech generation to comprehension of speech utterance including neural program, which explains articulatory coordination, segment sequencing and other components of speech production. This model makes it possible to study all phases of speech production in terms of semiotic subsystems on the basis of the main semiotic communicative system and to connect all stages of speech competence with stages of speech performance for a set of specific utterances produced by native-speakers on the basis of the economy principle of the speech generation. First of all this viewpoint is related to phonostylistics (styles of discourse), kinds of speech activity by communicans, etc. Thus this model on the basis of speech activities and the economy principle referred to the stylistic description, the style-shifting and semiotic code-switching.

Key words: economy principle; cognitive-semiotic model; speech communication; speech activities.

DAS KOGNITIV-SEMIOTISCHE MODELL DER SPRECHSPRACHLICHEN TÄTIGKEIT (bezüglich des Ökonomieprinzips)

In diesem Beitrag werden die Hauptfragen der Formierung und des Wesens des kognitiv-semiotischen Modells der sprechsprachlichen Tätigkeit bezüglich des Ökonomieprinzips besprochen. Die Untersuchung des Ökonomieprinzips schließt semiotische Subsysteme der sprechsprachlichen Kommunikation ein, die bestimmte Teile des semiotischen Makrosystems bilden und mit der Generierung der Sprechäußerungen verbunden sind. Dieses semiotische Generationsmodell der gesprochenen Sprache schließt 12 Phasen des Generationsprozesses ein: von der evokativen non-verbalen Phase der sprechsprachlichen Produktion im neuronalen Netz bis zum Sinnverständnis der Sprechäußerung. Alle Anfangsphasen schließen auch die Artikulationskoordination, die Reihenfolge der vlautlichen Segmente und andere Komponenten des sprechsprachlichen Produktes ein.

Mit Hilfe dieses Modells gibt es die Möglichkeit, alle Phasen der Redeproduktion auf der Basis der semiotischen Subsysteme im Rahmen des Makrosystems zu untersuchen. Dazu gehört auch die Möglichkeit, die Eigenartigkeit der Formel "Competence – Performance" beim Sprechen zu analysieren, was für das Ökonomieprinzip in der sprechsprachlichen Aktualisierung der individuellen Sprechakte besonders wichtig ist. Alle diese Probleme gehören vor allem zum Aspekt "Phonostilistik" ("Diskursstilistik"). Dieses Modell, das auf der Grundlage

der sprechsprachlichen Tätigkeit und des Ökonomieprinzips erarbeitet worden ist, gehört zum stilistischen Aspekt der sprechsprachlichen Kommunikation mit allen Besonderheiten der Variationsfaktoren und der Veränderung des sprechsprachlichen Codes des menschlichen Benehmens.

Schlüsselwörter: das Ökonomieprinzip; das kognitiv-semiotische Modell; die sprechsprachliche Kommunikation; die sprechsprachliche Tätigkeit.

Einführung

In den linguistischen Untersuchungen der letzteren Jahre wird der gesprochenen Sprache in verschiedenen Formen der Sprechfähigkeit große Aufmerksamkeit geschenkt. Die Unterschiede in der gesprochenen Sprache, die von den Situationsbedingungen determiniert werden, sind zum Objekt der zahlreichen linguistischen Erforschungen geworden. In dieser Hinsicht nehmen verschiedene Probleme der Stilistik einen bedeutenden Platz ein. Der Analyse werden lexikalische, syntaktische und phonetische Besonderheiten der Sprache im Zusammenhang mit ihrem Funktionieren in verschiedenen Stilen der mündlichen Form des Verkehrs unterzogen. Das stilistische System ist ein eigenartiger Komplex von Subsystemen der phonetischen, grammatischen und lexikalischen Stile, die immer zusammenwirken, aber nicht immer übereinstimmen. Solche Auffassung der linguistischen Stilistik setzt die Existenz des stilistischen Systems der Sprache voraus, das phonetische, grammatische und lexikalische Ebenen umfaßt. Jede von diesen Ebenen kann zum Gegenstand der Erforschung im Bereich der linguistischen Stilistik werden, weil die Sprachmittel jeder Ebene ihre Funktionierensbesonderheiten besitzen [Rehrmann 2011]. Die Sprachmittel der phonetischen Ebene des stilistischen Systems der Sprache im Vergleich zu den anderen Ebenen sind bis heute am wenigsten erforscht.

Überlegung zum Thema

«Das Ökonomieprinzip in der sprechsprachlichen Tätigkeit»

Für die kommunikationsorientierte Sprechwissenschaft spielen spezielle Untersuchungen auf dem Gebiet "Stillehre" eine große Rolle. Dazu gehört vor allem der Begriff "Funktionalstilistik", der mit der funktionalbezogenen Auffassung der sprechsprachlichen Kommunikation aufs engste verbunden ist. In diesem Fall soll jede sprechsprachliche Äußerung dem Ziel der Kommunikation adäquat sein. Darin besteht die

pragmatische und kommunikative Aufgabe der Funktionalstilistik, die drei Arten von Faktoren einschließt: *Produktion* → *Rezeption* → *Bewertung*.

Unserer Meinung nach funktioniert das Ökonomieprinzip sowohl im Rahmen der Produktion, als auch bezüglich der Perzeptionsbewertung der sprechsprachlichen Äußerung. Und das Wichtigste in der Realisierung dieses Prozesses ist die zusammenhängende Verbindung zwischen dem *Ökonomieprinzip der Redeproduktion* und dem *Kompensationsgesetz*, die sowohl die Generierung des Kommunikationsproduktes als Summe der Ausdrucksmittel des menschlichen Benehmens, als auch die Auffassungsbewertung einschließt und die funktional-stilistische komprimierte sprechsprachliche Äußerung in die sinnbewertende kommunikative Einheit umwandelt.

Im Prozeß der sprechsprachlichen Kommunikation gibt es Faktoren, die verschiedene Besonderheiten des funktionalen Stils der Artikulation und Prosodie der Sprechäußerungen charakterisieren. Das sprechsprachliche semiotische Makrosystem der interpersönlichen und interkulturellen Kommunikation schließt eine Reihe von Subsystemen ein [Потапова 2013; Потапова 2014]. Dazu gehört vor allem das phonetische Subsystem, das nicht nur durch das Ökonomieprinzip, sondern auch durch alle historischen Prozesse der menschlichen Evolution beeinflusst wird. Dabei spielen alle kultur-, sozial-, persönlich-individuell- kommunikativ- und emotiv bedingte Faktoren, die den Stil der gesprochenen Sprache beeinflussen, eine große Rolle. In diesem Fall wäre es zweckmäßig, von kulturellen, sprachlichen und sprechsprachlichen Kommunikationskompetenzen nach dem Modell “Competence (Kompetenz) – Performance (Performanz)” zu sprechen [Chomsky 1957; Chomsky 1969].

“Performanz umfaßt nach Chomsky die Anwendung des inneren Regelapparats, d.h. die Erzeugung von Sätzen, und zwar unter bestimmten individuellen und situativen Bedingungen. Diese Bedingungen werden vor allem als Störfaktoren bei der Sprachproduktion aufgefaßt: beim Sprecher begrenztes Gedächtnis, Zerstretheit, Änderung seiner sprachlichen Strategien oder Intentionen, bezogen auf die Situation störende Einflüsse wie Lärmquellen, dazwischen Reden u.a. Die Sprachproduktion eines *native speakers* unterliegt also (im allg.) Performanzbedingungen, die die Anwendung des generativen Regelsystems beeinflussen, stören, verändern” [Volmert 1995, S. 17].

Die Umgangssprache wirkt vor allem auf die Dynamik der Entwicklung des semiotischen sprachlichen und sprechsprachlichen Makrosystems

[Потапова 1981; Потапова 1986; Potapova 1995; Потапова 2013; Потапова 2014]. Als Folge dieser Wirkung trifft das Funktionieren im Redefluß verschiedener phonetischen segmentalen und suprasegmentalen (prosodischen) Variabilitätsphänomäne. In diesem Fall spielt das Ökonomieprinzip bezüglich der phonatorischen und artikulatorischen Gestendynamik eine führende Rolle. Die Realisation des Ökonomieprinzips in der gesprochenen Sprache wird durch den individuellen psychophysiologischen Synkretismus der sprechenden Persönlichkeit markiert (z.B. durch segmentale und suprasegmentale Umgruppierungen, Pausierung, Reduzierung, elliptische Änderungen, Gliederungsvarianten usw.) [Krech 1987; Hirschfeld, Wallraff 2002; Meinhold, Stock 1982; Потапова, Линднер 1991; Bose 1994; Potapova, Potapov 2011; Потапова, Потапов 2012; Stock 1985; Stock 2001]. Als Folge dieser Prozesse sind für die alltägliche sprechsprachliche Kommunikation solche kognitiven Determinanten charakteristisch, wie z.B. Subjektivität, Emotionalität, Sozialität und Flexibilität [Потапов 2017; Потапова 2017].

Unserer Meinung nach werden **die sprechsprachlichen Determinanten des Ökonomieprinzips durch das semiotische, generative, sprechsprachliche Modell konzipiert**, das folgende Phasen einschließt:

- **evocative Phase** – cognitive zerebrale neuronale Bereitschaft zum Kommunikationsprozess im allgemeinen;
- **nonverbale Phase** – das Funktionieren der subjectiven und sozialbedingten Assoziationsgestalten, Einschätzung der Situation, des Kommunikationspartners usw.;
- **vorverbale Phase** – innere zerebrale neuronale Stimulierung, Motivation, Absicht, Kommunikationsvorhaben usw.;
- **sinnformierende Phase** – inhaltsbildende Phase (inneres Denken);
- **paradigmatische innere Gestaltung der komprimierten sprechsprachlichen Äußerung** (inneres Sprechen);
- **syntagmatische expandierte sprachliche und sprechsprachliche Äußerungskodierung**;
- **formulierende Phase** – lexikalische und syntaktische Strukturierung der Äußerung;
- **Produktionsphase** – Koartikulations- und Artikulationsgestik, Phonationsmodulationen;
- **akustische Phase** – artikulations-akustische Umwandlung, akustische Kodierung der Sprechsignale;

– **Perzeptionsphase A** – neuronale Dekodierung der akustischen Signale;

– **Perzeptionsphase B** – semantische (phonetisch-intonatorische) Bearbeitung und Dekodierung der akustischen Signale;

– **Verständnisauflassungsphase des verbalen Signals** – Reaktion auf den sprechsprachliche Stimulus, die sprachliche Dekodierung der akustischen Signale mit Rücksicht auf sozial- und kulturellbedingte cognitive Mechanismen der sprechsprachlichen Kommunikation.

Der Prozeß des Funktionierens der Ökonomieprinziprealisation beim Sprechen beginnt von der Anfangsphase des sprechsprachlichen Kommunikationsaktes und umfaßt alle vorverbalen Phasen der Formierung des zerebral – neuronalen kommunikativen Programms, was mit dem neuro-psycho-physiologischen Programm der interpersönlichen Kommunikation aufs engste verbunden ist. Die sprechsprachlichen Besonderheiten der kommunikativen Realisation des Ökonomieprinzips funktionieren von Anfang an in den evokativen und nonverbalen Phasen und werden durch die Idiosynkrasie der Kommunikanten, einerseits, und Besonderheiten der Kommunikationsbedingungen, andererseits, bedingt.

Das Funktionieren des generativen Modells der sprechsprachlichen Tätigkeit zeugt davon, daß das Ökonomieprinzip sowohl die segmentale (Laute, Silben, Lautfolgen), als auch die suprasegmentale prosodische (Tempo, Grundfrequenz, Dauer, Intensität usw) Ebene der gesprochenen Sprache beeinflusst. Die mannigfaltigen Formen des mündlichen Sprachverkehrs (z.B. das Auftreten in der Feierversammlung, die Mitteilung der Nachrichten per Rundfunk und per TV, Vorlesung usw.) bestimmen sowohl die Wahl der phonetischen Mittel, als auch ihre Sprechbesonderheiten. In diesem Zusammenhang könnte man noch einmal unterstreichen, daß unsere Rede nicht immer eine und dieselbe ist – sie verändert sich je nachdem an wen wir uns anwenden, in welchen Bedingungen und zu welchem Zweck wir kommunizieren usw. Der Inhalt der Rede, soziale Bedingungen, der Zweck der Mitteilung u.a. bestimmen den Stil der Sprechäußerung. Auf diesem Grundsatz – des verschiedenen Grades der Sorgfältigkeit der Aussprache (die Bewahrung des phonematischen Bestandes der Wörter, das deutliche Aussprechen oder irgendwelche Veränderung des Segments des Wortes, das Fehlen der speziellen Sorgfältigkeit bis zu der Reduktion oder dem Ausfall einzelner Segmente usw.) beruhen die vollen, neutralen und umgangssprachlichen Stile der gesprochenen Sprache.

Der volle Stil wird bei den Mitteilungen der offiziellen Nachrichten per Rundfunk, bei den feierlichen Aufrufen an die Massen, bei den Diktaten usw. realisiert. Als Beispiel des neutralen Stils dienen die Mitteilungen, Vorlesungen, Vorträge usw. Der umgangssprachliche Stil findet seine Anwendung hauptsächlich im Bereich der Alltagssprache. In allen oben aufgezählten Arten der Stile werden die sprechsprachlichen Modifikationen sowohl der einzeln genommener Wörter als auch der Wortverbindungen analysiert. Das heißt, daß alle Arten der Stile der gesprochenen Sprache auf die Modifikationen der Laute der gesprochenen Sprache beeinflussen.

Der familiäre Stil ist für die sprechsprachliche Kommunikation in der Familie, mit den Freunden, Kollegen, usw. charakteristisch. Er widerspiegelt den spontanen Charakter der Rede und das Fehlen der Selbstkontrolle. In diesem Stil wird das Streben nach der Kompression und Expressivität beobachtet. Der auserlesene Stil zeichnet sich durch das Streben nach besonderer Klarheit und auch Expressivität aus. Alle orthoepischen Regeln sind beachtet.

In die russische experimentelle Phonetik wurde der Begriff der Sprachstile am Anfang des 20. Jahrhunderts eingeführt. Man behauptete, daß man im Zusammenhang mit der Mannigfaltigkeit der sozialen Bedingungen eine Menge von Stilen in der Sprache unterscheiden könnte. Es lohnt sich aber, nur zwei Hauptstile zu differenzieren: den einen, der für ein ruhiges Gespräch charakteristisch ist, und den anderen, den wir gebrauchen, wenn wir unter irgendwelchen Umständen unsere Rede besonders klar machen wollen.

Es ist zu betonen, daß die Veränderung des Sprechtempos für die Differenzierung der Sprechstile als markiertes Merkmal betrachtet werden könnte [5; 1]. In diesem Fall werden die Begriffe des **effektiven** und **fiktiven** Tempos in erster Linie genannt. Das effektive Tempo charakterisiert die Geschwindigkeit der Artikulation in bezug auf einen konkreten Sprechakt. Das fiktive Tempo charakterisiert die ideale Norm des Sprechens, die der schriftlichen Sprache entspricht. Von besonderem Interesse sind die Werte, die die Abweichung von dem idealen vollen Typ des Sprechens zeigen. Zu diesem Zweck dient der Koeffizient der Lautzahlverringering (K_i), der durch die folgende Formel dargestellt wird:

$$K_i = 10 * (V_{fik} - V_{eff}) / V_{fik}$$

Die Daten K_i zeugen davon, daß bei dem Vergleich verschiedener Stile bei dem Zusammenfall der Zahl der ideal und real artikulierten

Laute die insgesamte Dauer, die mittellautliche Dauer, der Koeffizient der Lautverringernung K_i verschiedener Stile nicht gleich sind. Wenn das Tempo wächst, vergrößert sich auch K_i und wenn das Tempo etwa langsamer wird, so verringert sich auch K_i .

Im Prozeß der Erforschung der Dauerwerte ist es möglich, sich an die verschiedenen Parameter, deren Zahl und Charakter von konkreten Aufgaben der Erforschung abhängen, anzuwenden. Dazu gehören absolute und relative Dauerwerte des Syntagmas; Durchschnittsdauer des Syntagmas; absolute und relative Dauerwerte für Vokale, Konsonanten und Silben in einem Syntagma usw.

Es wurde die Zahl der real artikulierten Laute in jedem Syntagma gerechnet und die Vergleichsanalyse in Bezug auf die Zahl der ideal zu artikulierenden Laute durchgeführt. Zu diesem Zweck wurden experimentelle Syntagmas der Norm der Hochlautung entsprechend transkribiert. Real ausgesprochene Laute wurden zweifacherweise analysiert: a) als Ergebnisse der Höranalyse; b) als Resultate der Dekodierung des Sprechsignals mit Hilfe der Anwendung des speziellen hard- und software-Programms. Danach wurde der Koeffizient der Lautverringernung K_i für alle Sprechstile berechnet. Und in der nächsten Etappe wurde der Tempovergleich zwischen den verschiedenen Sprechstilen beschrieben.

Die festgestellten Werte K_i zeigen anschaulich, daß der Koeffizient der Lautzahlverringernung für alle Sprechstile unterschiedlich ist. Am wenigsten ist der Koeffizient der Lautzahlverringernung K_i für die Sprechrealisation des feierlichen pathetischen Stils. Dann folgen der Nachrichtenstil, der wissenschaftlich-populäre Stil und der umgangssprachliche Stil, die im Vergleich zu den anderen Stilen das schnellere Tempo und folglich den größten Koeffizient der Lautverringernung K_i haben. Wenn man die Ergebnisse der durchgeführten Analyse in bezug auf Sprechstilarten zusammenfasst, so könnte man feststellen, daß es eine bestimmte Tendenz zur Differenzierung der Stilarten durch das Tempo und die entsprechenden K_i – Werte gibt. Alle diese Unterschiede werden durch K_i – Skala charakterisiert, was für weitere Sprechstiluntersuchungen von großer Bedeutung wäre. Die Ausnutzung K_i gibt die Möglichkeit, alle individuellen Artikulationsbesonderheiten des Sprechers zu eliminieren und relative Hauptunterschiede in bezug auf Sprechstile korrekt zu beschreiben, was sehr perspektiv für die die Erforschung des Problems des funktionierens des Ökonomieprinzips beim Sprechen zu sein scheint.

Zusammenfassung

Unsere Untersuchungen zeugen davon, daß in diesem Fall zwei Subprinzipien des Ökonomieprinzips funktionieren: Informationskompression und Informationskompensation. Im Fall des Funktionierens der neuronalen Informationskompression werden verschiedene akustische Werte (z.B., F_o , I_n , t_n) im sprechsprachlichen "Mechanismus" des Menschen komprimiert. Und das Wichtigste besteht darin, daß im Output dieses Blocksystems der sprechsprachlichen Tätigkeit diese komprimierten Werte nicht expandiert werden und constant bleiben: z.B. die Dauerwerte der Pausen werden entweder komprimiert oder zu 0 umgewandelt, die Tonhöhe- und Intensitätswerte werden auch komprimiert, z.B., für F_o ; Umfang-, Register-, Kontur Werte usw anbeht. Die Prosodie ändert sich in diesem Fall linear / syntagmatisch: eine Gruppe von komprimierten prosodischen Charakteristiken wird im Allegro-Redekontinuum durch die andere Gruppe von prosodischen Charakteristiken kompensiert. Und diese Kompensation der Prosodiemerkmale dient zum Verstehen der ganzen Äußerung: z.B. die Frequenzzahl der Pausen sind mit dem Tempo, die Dauer der Pausen – mit der Dauerkompression der Vokale, die Intensitätswerte – mit der Tonhöhe usw verbunden. In allen Fällen funktioniert das Kompensationsgesetz in der reziproken Umgestaltung der prosodischen Werte, was zum Verstehen der ganzen Äußerung führt [Потапова 1981; Потапова 1986; Potapova 1995; Потапова 2013; Потапова 2014]. Zum Zweck der Kompensation dienen auch verschiedene nonverbale Mittel (z.B. Mimik, Gestik, Proxemik usw.).

Unsere Hauptidee beht die Hypothese, daß das Ökonomieprinzip eine *kommunikativ bedingte neuro-psycho-motorische Universalie* ist, die mit allen semiotischen kommunikativ-bedingten semiotischen Subsystemen des menschlichen Makrosystems aufs engste verbunden ist und die auf der Grundlage des psychomotorischen Kompensationsgesetzes des menschlichen Benehmens existiert.

LITERATURANGABEN

- Потапов В. В. Разработка проблем вариативности английской речи в трудах российских лингвистов: XXI век // Норма и вариативность в языке и речи: сб. науч. тр. / отв. ред. Е. А. Казак. М. : РАН ИНИОН, 2017. С. 9–39.
- Потапова Р. К. О «типах» примыкания в германских языках // Теория языка, методы его исследования и преподавания. К 100-летию со дня рождения Л. В. Щербы. Ленинград: Наука, 1981. С. 200–207.

- Потапова Р. К.* Слоговая фонетика германских языков. М.: Высшая школа, 1986. 144 с.
- Потапова Р. К.* Семиотическая гетерогенность устно-речевой коммуникации // *Нейрокомпьютеры: разработка, применение*. М.: Радиотехника, 2013. № 1. С. 4–6.
- Потапова Р. К.* Речь как подсистема общей семиотической системы межличностной коммуникации // *Вестник МГЛУ*. 2014. Вып. 1 (687). С. 189–196. (Лингвистика.)
- Потапова Р. К.* Современная немецкая речь: произносительная норма и вариативность // *Норма и вариативность в языке и речи: сб. науч. тр. / отв. ред. Е. А. Казак*. М.: РАН ИНИОН, 2017. С. 91–122.
- Потапова Р. К., Линднер Г.* Особенности немецкого произношения. М.: Высшая школа, 1991. 319 с.
- Потапова Р. К., Потапов В. В.* Речевая коммуникация: От звука к высказыванию. М.: Языки славянских культур, 2012. 462 с.
- Bose I.* Zur temporalen Struktur frei gesprochenen Texte // *Forum Phonetikum*. 58. Frankfurt am Main: Hector, 1994. 176 S.
- Chomsky N.* Syntactic Structures. The Hague–Paris: Mouton, 1957. 117 p.
- Chomsky N.* Aspekte der Syntax-Theorie. Frankfurt am Main: Suhrkamp, 1969. 313 S.
- Hirschfeld U., Wallraff U.* Untersuchungen zum Schwa im Deutschen // *A. Braun, H. R. Masthoff (eds.): Phonetics and its applications. Festschrift for J.-P. Köster on the occasion of his 60th birthday*. Wiesbaden–Stuttgart: Franz Steiner Verlag, 2002. S. 493–505.
- Krech E.-M.* Vortragskunst. Leipzig: VEB Bibliographisches Institut, 1987. 175 S.
- Meinhold G., Stock E.* Satzphonologie der deutschen Sprache // *Phonologie der deutschen Gegenwartssprache*. 2., durchges. Aufl. Leipzig: VEB Bibliographisches Institut, 1982. 256 S.
- Potapova R. K.* Phonetische Besonderheiten der segmentalen Sprechereinheiten des Deutschen (in bezug auf eine Vergleichsanalyse der Dauerwerte für deutsche lange und kurze Vokale im Redekontinuum) // *A. Leonhargt, D. Mehnert (Hg.): Begegnungen. Festschrift zum 70. Geburtstag von G. Lindner. Hörgeschädigten Pädagogik*. Heidelberg: Jilius Groos Verlag, 1995. Beiheft 36. S. 158–171.
- Potapova R. K., Potapov V. V.* Kommunikative Sprechfähigkeit: Russland und Deutschland im Vergleich. Köln–Weimar–Wien: Böhlau Verlag, 2011. 312 S.
- Rehrmann N.* (Hg.) Schlechter Stil: Sprachkritik aus fünf Jahrhunderten. Darmstadt: Lambert Schneider, 2011. 207 S.
- Stock E.* Phonetische und rhetorische Determination der Wirkung gesprochener Sprache // *Th. Höhle, E.-M. Krech, G. Lerchner, D. Sommer (Hrsg.): Hallesche Studien zur Wirkung von Sprache und Literatur*. Halle / Saale, 1985. Bd. 10. S. 35–52.

- Stock E.* Sprech- und sprachwissenschaftliche Ansätze zur Erkrärung der Artikulationsabläufe // S. Lemke (Hg.): Sprechwissenschaftler/in und Sprecherzieher/in. Eignung und Qualifikation. München–Basel, 2001. S. 62–67.
- Volmert J.* (Hg.). Grundkurs Sprachwissenschaft. München: Wilhelm Fink Verlag, 1995. 270 S.

Р. К. Потапова, Л. Р. Комалова

Р. К. Потапова, доктор филологических наук, профессор; действительный член Международной академии информатизации; заведующая кафедрой прикладной и экспериментальной лингвистики; директор Института прикладной и математической лингвистики МГЛУ; e-mail: rkpotapova@yandex.ru

Л. Р. Комалова, доктор филологических наук, ст. научный сотрудник отдела языкознания; Институт научной информации по общественным наукам РАН; профессор кафедры прикладной и экспериментальной лингвистики МГЛУ; e-mail: GenuinePR@yandex.ru

**ПОЛНОТЕКСТОВАЯ АННОТИРОВАННАЯ БАЗА ДАННЫХ
РУССКОЯЗЫЧНЫХ ПОЛИЛОГОВ СОЦИАЛЬНО-СЕТЕВОГО
ДИСКУРСА: ПОДКОРПУС «ПОЛИТИЧЕСКИЕ ТРАНСФОРМАЦИИ»¹**

Описываемые в статье результаты являются частью более обширного исследования антисоциального речевого поведения путем лингвистической диагностики социальных сетей в российском сегменте Интернета с тематической ориентацией на социально-когнитивные факторы депривации. Описывается подкорпус аннотированных полилогических единств. При формировании корпуса особое внимание уделялось речи / дискурсу / текстам с отрицательной (деструктивной) прагматической ориентацией, соотносящейся с проявлением многофакторной депривации в письменной речи. Подкорпус состоит из набора аннотированных дискурсивных единиц (полилогов), представляющих тематические кластеры смыслового поля «политические трансформации». Подкорпус составлялся путем сплошной выборки из социально-сетевого дискурса с октября 2015 г. по март 2016 г. Полилоги (в авторской орфографии) представлены на русском языке с речевыми элементами на других языках. Участниками полилогов являются русскоязычные пользователи таких социальных сетей, как «ВКонтакте» и «Живой журнал», а также пользователи, комментирующие новости на сайтах информационных агентств и сетевых СМИ. Специальная аннотация позволяет выявить прагматическую составляющую каждого полилогического единства, эмоциональное состояние коммуникантов, возможный перлокутивный эффект, оказываемый на коммуникантов и наблюдателей (невовлеченных в полилог). Сформированный подкорпус является многоцелевым и может использоваться при описании интегративного «портрета» пользователей, участвующих в социально-сетевом дискурсе в исследуемом сегменте Интернета.

Ключевые слова: лингвистический корпус; социально-сетевой дискурс; аннотация; письменная речь; Интернет; отрицательное эмоциональное состояние; полилог; русский язык; прагматика.

R. K. Potapova

Sc.D., Prof., Full Member of the International Informatization Academy, Head of Department of Applied and Experimental Linguistics, Director of Institute for Applied and Mathematical Linguistics of Moscow State Linguistic University (MSLU); e-mail: rkpotapova@yandex.ru

¹Исследование поддержано Российским Научным Фондом (РНФ). Проект № 18-18-00477.

L. R. Komalova

Sc.D., Senior Research Fellow at the Department of Linguistics, Institute of Scientific Information for Social Sciences of the Russian Academy of Science, Professor at the Department of Applied and Experimental Linguistics, Moscow State Linguistic University, e-mail: GenuinePR@yandex.ru

**RUSSIAN SND FULL-TEXT ANNOTATED
POLYLOGUES DATABASE:
“POLITICAL TRANSFORMATIONS” SUB-CORPUS**

The paper represents a part of the global research on antisocial speech behavior by linguistic diagnosis of social network interaction on the Russian segment of the Internet with thematic focus on socio-cognitive deprivation factors. A sub-corpus of annotated polylogic items is described. The formation of the corpuses was aimed at the acquisition of objective experimental data on the current state of social network discourse existing at present on Russian segment of the Internet. Particular attention was paid to the speech / discourse / texts with negative (destructive) pragmatic orientation related to the manifestation of multifactor deprivation in written speech activity. The described sub-corpus is a collection of specially annotated independent discursive items that are polylogic items (polylogues) with isolation of reference-thematic clusters related to the semantic field “political transformations”. The sub-corpus was collected by continuous sampling of the social network discourse over some time period: October 2015 – March 2016. The polylogues (in their authors’ wording) are presented in Russian with speech elements in other languages. The communicants are Russian-speaking users of such public social networks as “VKontakte” and “Livejournal”, as well as users commenting on news reports on the websites of news agencies and online media. The annotation of the sub-corpus gives possibility to reveal the pragmatic component of each polylogic item, communicants’ emotional state, possible perlocutionary effect on communicants and on bystanders (not involved readers of the polylogue). The created sub-corpus is a multi-purpose one and can be used in the formation of an integrative “portrait” of participants involved in the social network discourse in this segment of the Internet.

Key words: Corpus linguistics; social network discourse; annotation; written speech; Internet; negative emotional state; polylogue; Russian language; pragmatics.

Introduction

Against the backdrop of intense political and geopolitical transformations of the late XX – early XXI centuries there is a need in development of effective mechanisms for recognition of antisocial (including speech) behavior (see, e.g. [Cole 2001; Gawda 2010; Mceachern, Snyder 2012; Ротарова 2015; Потапова, Потапов 2004, 2006; Rinaldi 2003; Речевая коммуникация 2017]).

The methodology of the proposed scientific research is oriented to solve the following problems:

- determination of the role of linguistic means in shaping the idea of extremism, stirring up hatred, hostility, discord, etc. on the basis of deprivation phenomenon;
- determination of linguistic and verbal means used to manipulate the communicants' conscience;
- systemization of lexical and prosodic means of shaping the conceptual categories, criteria of concepts and frame structures of knowledge for the semantic field “deprivation”;
- classification of linguistic and verbal means of representation and conceptual metaphor formation in social network discourse¹;
- development of key criteria for carrying out linguistic examination of speech / discourse / texts with destructive orientation including predictive reconstruction of deprivation types.

Within the proposed concept [Потапова 2014; Potapova 2015; Potapova 2017] an integral part of the research is focused on determination of the “portrait” of the speaker (writer) based on the indices extracted of the pre-formed databases including the following communicant's Internet identity parameters: sex, approximate age, national affiliation, residence, native and / or non-native languages, social status, economic status, type of deprivation, political views, etc.

One of the destructive strategies of coping with deprivation factors is response / instrumental / initial types of aggression. In the study of the semantic field “aggression” the focus of attention includes means and mechanisms for the implementation and motivation of negative emotions of the recipient, including those produced in the interaction in social networks on the Internet. Previous results of SND content analysis (by the method described in [Potapova 2014]) confirm the assumption on the expansion of the semantic field “aggression” and reduction of the perception criticality threshold (see, e.g. [Кошкарлова 2009; Шаховский 2016]).

The availability of texts on the Internet and modern software developments enable monitoring the Internet content and identification of discourses that incite aggressive behavior. Very relevant and promising is the development of databases and knowledge bases that enable clusterization

¹Social Network Discourse (SND) paradigm was suggested by professor Rodmonga Potapova in works [Потапова 2014; Potapova 2015; Потапова 2017].

for verbal means of actualization of destructive (including speech) behavior (see, e.g. [Комалова 2017]) together with the capability to assess their effect strength, as well as the potential of respondent's motivation to various forms of aggression [Blaya, Fartoukh 2016; Erekovia 2014; Runions 2013; Spitzberg Gawron 2016; Wah, Min 2014; Werner, Bumpus, Rock 2010].

Corpus description

As part of the global research of destructive speech behavior under the influence of socio-cognitive deprivation factors we formed various annotated speech / discourse / texts corpuses containing language material from blogs, microblogs, vlogs, social networks, forums and news websites with public domain discussions. The formation of the corpuses was aimed at the acquisition of objective experimental data on the current state of social network discourse existing at present on Russian segment of the Internet. Particular attention was paid to the speech / discourse / texts with negative (destructive) pragmatic orientation that related to the manifestation of multifactor deprivation in written speech activity.

All the data collected is classified according to their predominant topic areas mapping to respective predominant types of deprivation. The topic areas include politics, geopolitics, ecology, economy, interethnic and interreligious conflicts, and interpersonal relationships.

Russian SND annotated polylogue corpus forms part of the variety of corpuses created and consists of Russian electronic media discourses. The corpus includes samples of written and spoken monologues, dialogues, and macropolylogues classified with respect to their form, content, function, influence types, emotional and modal colouring. Total amount of the data is more than 16 million word-forms and 26 hours of audiovisual material.

The described sub-corpus is a collection of specially annotated independent discursive items that are polylogic items (polylogues) with isolation of reference-thematic clusters related to the semantic field "political transformations".

A specific feature of polylogic items in the described sub-corpus is their pragmatic component representing a negative pragmatic attitude of the communicants (communication participants) to each other, to the environment, to events related to political transformations in the country of their residence, in other countries, as well as in the whole world. In this

case, the main markers of the influencing discourse are evaluative lexical items and neologisms which are tag-marked within the polylogues.

The sub-corpus was collected by continuous sampling of the social network discourse over some time period: October 2015 – March 2016. The polylogues (in their authors' wording) are presented in Russian with speech elements in other languages. The communicants are Russian-speaking users of such public social networks as "VKontakte" and "Livejournal", as well as users commenting on news reports on the websites of news agencies and online media.

The annotated items were formed with due regard to the definition of negative pragmatic attitude of the communicants during speech interaction, with the participation of 30 subjects (annotators) – Russian native speakers differentiated by gender and age (21 women, 9 men; aged 20–23 years) who were asked to answer the special questionnaire, lately transformed into annotation parameters of the items.

The sub-corpus is being created as an open system, which can be replenished, with the possibility to expand the list of annotation parameters as well. The sub-corpus currently includes 937 annotated polylogic items in Russian on political and geopolitical transformation issues.

Sub-corpus Annotation

The starting sub-corpus file was created in the Microsoft Office Excel 2007 as a table, which presents an extended annotation of each polylogic item. The annotation parameters were developed on the basis of previous experience in forming annotated full-text corpuses with the text material in Russian, English and Spanish from online media, the content of which include examples of the semantic field "aggression" [Потапова, Камолова 2014a; Potapova, Komalova 2014b].

Changing the sub-corpus item type from the author monologic text (prepared written discourse compositionally organized in the complete whole – online media monotexts) to the polylogic item of the social network discourse (SND) required transformation of the annotation structure (table 1). Despite the fact that an initial post (conditional utterance) sets the topic of discussion, the large number of participants (communicants) blurs the scope of the original topic and adds some "multidisciplinarity" to the polylogue. At the same time, depending on the activity of certain participants of the discussion, the discussion focus can be transferred from one topic to another one. Therefore, a cluster that offers selection from

the list of stereotypic topics of communication (formulated during the previous experimental studies) was removed from the annotation item, and two windows were introduced. Now the operator formulates a) the main topic of the polylogue and, if it's necessary, b) associated topic(s). The annotation structure is now also missing a cluster of quantitative data on the number of word forms and graphemes in polylogues and has a cluster of quantitative data on participants involved in polylogues.

Compared with the annotation for monotexts of online media, the SND-polylogue annotation is characterized by the representation of features of the speech production of the so-called collective author and the appeal to the two suggestion foci: the internal one – the effect on the direct participants of the communication (the included recipient), and the external one – the effect on the “casual” observer (the bystanding recipient). In this regard, the polylogic item annotation focuses on evaluation of the polylogue effect strength (affective and pragmatic components) on the recipient, evaluation of direct participants of the polylogue by their speech activity. The annotation also includes a cluster of data on bystanders, the role of which is taken on by the subjects.

The annotation of each polylogic item contains the following data obtained from the formal analysis of the available data and the subjective evaluation by the subjects:

I. Information about the source of the verbal (speech) material: social network name; time period of the polylogue creation (discussion “thread”: first message date – last message date); link to the polylogue electronic address on the Internet (table 2).

Table 1

Annotation clusters (in comparison)

Annotation blocks	Written media text annotation	SND annotation
Date-line	headline	–
	name of the text-source	name of SND source
	date on web-site publication	time period of SND creation
	name of the text's author	<i>this cluster was extended to a separate block (V)</i>
	web-site address of the text	link to the SND-source

Annotation blocks	Written media text annotation	SND annotation
Linguistic "passport"	topic of discussion	topic(s) of discussion
	text's genre	–
	pragmatic effect on a text reader	<i>extended to a separate block (III)</i>
	–	information about the SND-participants (block IV)
Special metrics	total amount of notional and non-notional words in the text	–
	total amount of graphemes without blanks	–
	local dictionary of the semantic field "aggression" (a list of words from the text which belongs to the semantic field "aggression")	–
	density of words from the semantic field "aggression"	–
Content	original full text	original full SND fragment

Table 2

Examples of "Block I" representation in the sub-corpus structure

Block I (formal SND-data)		
SND-name	Time period of the polylogue creation	Polylogue web-address
Debatepolitics	07.02.2016 – 17.02.2016	debatepolitics.ru/showthread.php?t=11327
Bolshoyforum	09.02.2016, 18:36:51 – 16.02.2016, 21:16:36	bolshoyforum.com/forum/index.php?page=915
Nastej	19.01.2016 – 08.02.2016	nastej.ru/threads/Дмитрий-Глуховский-Без-порток.1686/
Newkalininingrad	18.02.2016	www.newkalininingrad.ru/forum/topic/286820-ukraina-namerena-pereimenovat-rossiiu/
Newsland	17.02.2016 – 17.02.2016	newsland.com/user/4296647985/content/5051552
Politforums	24.11.2015 – 24.11.2015	www.politforums.net/foreign/1448374795.html

Block I (formal SND-data)		
SND-name	Time period of the polylogue creation	Polylogue web-address
Twitter	14.02.2016	twitter.com/EchoMskRu/ status/698576431920578560
Ukrpolitforum	19.01.2016, 16:49 – 24.01.2016, 18:52	ukrpolitforum.com/topic/51172- politika-rossii-vrag-svoego-naroda-ili- narodov/
VK	16.02.2016	vk.com/wall-67991642_284277
ВКонтакте	27.01.2016 – 29.01.2016	vk.com/politicsclub1?w=wall- 71631912_818635
НТВ	09.02.2016 20:03:59 – 10.02.2016 07:12:32	www.ntv.ru/novosti/1604264
РИА	11.01.2016, 19:28 – 14.01.2016, 11:09	ria.ru/ analytics/20160111/1358068519.html

II. Determination of the polylogue topic: main topic of the polylogue; associated topic(s) within the main topic (subtopics) (table 3).

Table 3

Examples of “Block II” representation in the sub-corpus structure

Block II (Polylogue topic(s))
Участники обсуждают проблему сирийского конфликта. Они поднимают вопрос о роли России в этом конфликте. Также говорят о заявлении британского министра.
Участники обсуждают строительство СП2. В этом полилоге они затрагивают отношения между Россией и Америкой. Также они поднимают вопрос о влиянии Америки на Европу и Украину.
Внешняя политика: на Мюнхенской конференции по безопасности Дмитрий Медведев констатировал начало новой холодной войны между НАТО и Россией.
Внешняя политика: в Великобритании пройдет референдум о выходе из Евросоюза
Внутренняя политика: глава Южной Осетии Леонид Тибилов объявил, что Южная Осетия собирается провести референдум о вхождении в состав РФ. В связи с осложнением международной обстановки многие граждане Южной Осетии высказываются за присоединение к России.

Взаимоотношения Турции, России и курдов. Внутренняя политика России
Отношения Турции и России. Военная мощь РФ. Отношения России и Украины. Личность Пескова

III. Pragmatic component of the polylogue (the *subjects* ' evaluation of the effect strength of the polylogic item contents) (table 4):

a) emotional component noted by the subjects (not by the participants of the polylogue analyzed) who just got familiarized with the polylogue contents regarding the following evaluations: concern, anxiety, fear, stress, dissatisfaction, indignation, irritation, disgust, other (the subject was to indicate, which one);

b) forecasting the potential perlocutionary effect experienced by the participant of the polylogue (addressee), namely: what can the communicant be made to do in real life after participating in such discussions;

c) forecasting the potential perlocutionary effect experienced by the bystander when familiarizing with the contents of this polylogue.

Table 4

Examples of “Block III” representation in the sub-corpus structure

Block III (pragmatic component of SND-fragment)		
a	b	c
Полилог несет отрицательную окраску. Заметно возрастает ненависть к США, Турции и Сирии	Участники полилога настроены очень категорично по отношению к США. В реальной жизни такой полилог мог бы проходить еще более агрессивно	Реципиент постарался бы избежать участия в данном полилоге и не стал бы дочитывать его до конца
Напугана тем, что некоторые люди так легко говорят об уничтожении других стран	К дискуссии о том, как должна поступить Россия с Турцией	К рассуждению о том, почему Турция на самом деле сбита российский самолет
Недовольна, что Турция косвенно обеспечивает ИГИЛ	К смене отношения к Турции, к недовольству ею	К смене отношения к Турции, к недовольству ею
Расстроилась, так как поняла, что торговля для России посредством грузовых перевозок стала почти невозможна	К дискуссии о политике Польши, о бездействии властей Украины	К участию в обсуждении

Block III (pragmatic component of SND-fragment)		
а	б	с
Забеспокоилась, что есть возможность начала новой «холодной войны» между Россией и США	К дискуссии на тему «холодной войны» 1960–1990-х гг. XX в. и на тему стремлений США	К рассуждению на тему полилога
Не изменилось	К дискуссии о намерениях Обамы	К участию в данном полилоге
Разозлило, что некоторые участники полилога пишут только с целью поругаться с кем-нибудь	К размышлению о том, почему Великобритания хочет выйти из ЕС	К размышлению о том, почему Великобритания хочет выйти из ЕС. Может вызвать у реципиента желание высказать свое мнение

IV. Information on the *subject* involved in the experiment: gender; age as of the experiment time; place of birth and places of long residence; native language and studied language(s); social background; educational level (table 5).

Table 5

Examples of “Block IV” representation in the sub-corpus structure

Block IV (subjects' portrait)					
Gender	Age (years)	Place of birth / long resistance	Native / studied languages	Social background	Education
Ж	20	Москва	русский / английский, немецкий	из семьи служащих	НВ ¹
М	20	Новосибирск, Москва	русский / английский, немецкий	из семьи рабочих	НВ
Ж	20	РФ	русский / английский, немецкий, французский	интеллигенция	НВ
Ж	20	Лимас-сол, Кипр; Россия, Москва	русский / английский, немецкий	интеллигенция	НВ

V. Information on the *participants* of the polylogue (hypothetical data based on the perception by the subjects of the polylogue content and parametric data indicated by the communicants in their social network account): communicant’s nickname, gender, age, “conditional” native language, nationality, education level, occupation, social status (table 6).

Examples of “Block V” representation in the sub-corpus structure

Block V (communicants' portrait)							
Nick name	Sex	Age	Native language	Nationality	Education	Occupation	Social status
A ²	М	32	рус.	рус.	высшее	типография	работающий
Б	Ж	32	рус.	рус.	высшее	филолог	интеллигенция
В	М	35	рус.	рус.	высшее	фотограф	служащий
Г	М	36	рус.	рус.	высшее	электрик	работающий
Д	М	55	рус.	рус.	высшее	преподаватель	служащий
Е	М	43	рус.	рус.	среднее	кожевенное дело	работающий
Ж	М	37	рус.	рус.	высшее	электромонтер	работающий
З	Ж	51	рус.	рус.	высшее	юрист	служащий
И	Ж	64	рус.	рус.	среднее	домохозяйка	пенсионер
К	Ж	73	рус.	русс.	высшее	анестезиолог	служащий
Л	М	49	рус.	рус.	высшее	редактор	служащий
М	М	44	рус.	рус.	высшее	директор по производству	служащий

Each annotated polylogic item is assigned a special identification number, which is specified in the annotation. When clicking the mouse button on the number, a new window opens containing a word-file with the polylogic item (table 7), which in addition to a brief annotation (providing formal data – blocks I and II) includes the full original discourse fragment of the message “thread” for the period specified.

¹ Неоконченное высшее.

Table 7

Example of a fragment of the annotated polylogic item with markup – representation in the sub-corpus

SND-name	TACC «ВКонтакте»
Time period of the polylogue creation	12.02.2016 6:31-10:07
Polylogue web-address	vk.com/tassagency
Polylogue topic(s)	Участники полилога обсуждают проблему сирийского кризиса, роль России и Америки в данном конфликте. Также они ставят вопрос о том, какое будущее ждет Россию
<p>Original full SND fragment¹:</p> <p>A: США тоже хочет присоединиться к России и направлять гуманитарную помощь. Правда она пока исподтишка бомбит освободившие районы и как всегда сваливает это на Россию. Поэтому сирийцам нужно быть с ними поосторожней. Не дай Бог подложат в мешки с продуктами бомбу.</p> <p>Б: Забегало _Н пендосиво *_Н, завыло, когда хвост их _Н выкормышам *_Н защемило.</p> <p>В: США - _Н примазище! *_Н</p> <p>Г: Радик, Россия – не Великая Держава, а великая территория от проклятого царизма и _Л хвалёного социализма.. *_Л. ешё не всю разбазарили путёмы...</p> <p>Д: Ну тут единственный комментарий может быть – изначально нефть качалась террористами для США и готовили их специалисты из США.</p> <p>Е: Как только объявят о перемирии – террористы вновь заполняют всю Сирию и будут торговать нефтью. Сорняк с грядки убирают под корень. etc.</p>	

The polylogic item is a multi-communicants discourse fragment marked with special tags that allow identification of discourse fragments that were evaluated by the subjects as a representation of communicants' pragmatic and emotional attitude aimed at confrontation. Tags mark verbal

¹ Private information is decoded.

fragments of the polylogue, where the communicants use evaluative lexical items (_Л _*Л) and neologisms (_Н _*Н). See examples¹ below:

- (1) Ах, как удобно все свалить на _Л “кривые руки” _*Л арабов.
- (2) Ну да это как водиться во всем виноват _Л ВВП. _*Л
- (3) А зачем лайкаете эту фотку? Типа хорошо _Н каклятские _*Н каратели _Л разъепашили _*Л школу, вам нравится.
- (4) Предавший раз, предаст и дважды. Тем более, если речи идет о маргиналах, поддерживавших _Н дыныры _*Н
- (5) Ну так он же _Н крысчанин _*Н, что с него взять.
- (6) _Н рашисты _*Н, одним словом.
- (7) Судись по _Л карательной операции _*Л _Н русликов _*Н во время двух гражданских войн в Чечне?
- (8) С каких пряников Львов вдруг “польский”?) На _Н сколенвстане _*Н историю уже не учат?))
- (9) Да вообще _Л пи...ц _*Л, сколько уже в _Н укрупии _*Н наших полегло))) в сирии вообще мясорубка... Уже в каждую семью похоронки приходят
- (10) Подтянулось убогое. Давай блесни. Повысь градус _Л тупости _*Л этому посту))))

The operator can optionally make adjustments to the markup and annotation of each polylogic item and save these corrections.

Navigating and search in the described sub-corpus and a concrete polylogic item is provided in accordance with search capabilities using filters embedded in “Microsoft Office Excel 2007” and higher, “Microsoft Office Word 2007” and higher programs.

Sub-corpus research capabilities

The “political transformations” sub-corpus is practically applicable for researching the evaluative lexis and syntax aimed at influencing, taking into account the emotional state of both a communication participant and a bystander (an observer) by means of messages in social network

¹All examples are given in original wording. Authors of the paper are not responsible for the SND-communicants and their point of view on other communicants and social events represented in the examples.

discourse on the Russian segment of the Internet. The sub-corpus can be used for topic analysis to identify the structure of discourse generating a negative emotional state of the subject in such research domains as linguaconflictology, text expertise, forensic linguistics, fundamental and applied speechology, sociolinguistics, and new information technologies and linguistics etc.

The data obtained during the research allows to conclude that individuals whose actual life situation is determined by prolonged multifactor deprivation (in social, economic, political, cognitive, psycho-emotional spheres of life) manifest socio-cognitive deprivation in their written speech in the form of frequent use of speech genres of aggressive type when actualizing themes associated with facts of deprivation.

REFERENCES

- Комалова Л. Р.* Агрессогенный дискурс: типология мультилингвальной вербализации агрессии. М. : Спутник+, 2017. 275 с.
- Кошкарлова Н. Н.* Конфликтный дискурс: психологические и лингвистические аспекты. Челябинск : ЮУрГУ, Губерния, 2009. 175 с.
- Потапова Р. К.* Социально-сетевой дискурс как объект междисциплинарного исследования: II Междунар. науч. конф. «Дискурс как социальная деятельность: приоритеты и перспективы», 16–17 октября 2014. Ч. 1. М. : Рема, 2014. С. 20–22.
- Потапова Р. К.* Депривация как базовый механизм вербального и паравербального поведения человека (на основе социально-сетевого дискурса) // Речевая коммуникация в информационном пространстве. М. : Ленанд, 2017. С. 17–36.
- Потапова Р. К., Комалова Л. Р.* База данных русскоязычных текстов, содержащих единицы семантического поля «агрессия» // Вестник МГЛУ. 2014. Вып. 19 (705). С. 112–121. URL: www.vestnik-mslu.ru/Vest-2014/Vest-19z.pdf.
- Потапова Р. К., Потапов В. В.* Семантическое поле «наркотики»: дискурс как объект прикладной лингвистики. М. : URSS, 2004. 192 с.
- Потапова Р. К., Потапов В. В.* Язык, речь, личность. М. : Языки славянской культуры, 2006. 496 с.
- Речевая коммуникация в информационном пространстве. М. : Ленанд, 2017. 112 с.
- Шаховский В. И.* Диссонанс экологичности в коммуникативном круге: человек, язык, эмоции. Волгоград : ИП Поликарпов И. Л., 2016. 504 с.
- Blaya C. and Fartoukh M.* Digital uses, victimization and online aggression: a comparative study between primary school and lower secondary school

- students in France // *European Journal on Criminal Policy and Research*. 22.2. 2016. P. 285–300. URL: [dx.doi.org/10.1007/s10610-015-9293-7](https://doi.org/10.1007/s10610-015-9293-7).
- Cole D.* Narrative development in aggressive boys. *Behavioral Disorders*. 26.4 (Aug. 2001). 2001. P. 332.
- Erekovice A.* An insight into multimodal databases for social signal processing: acquisition, efforts, and directions. *The Artificial Intelligence Review*. 42.4 (Dec. 2014), 2014. P. 663–692. URL: [dx.doi.org/10.1007/s10462-012-9334-2](https://doi.org/10.1007/s10462-012-9334-2).
- Gawda B.* Syntax of emotional narratives of persons diagnosed with antisocial personality // *Journal of Psycholinguistic Research*. 39.4 (Aug. 2010), 2010. P. 273–283. URL: [dx.doi.org/10.1007/s10936-009-9140-4](https://doi.org/10.1007/s10936-009-9140-4).
- Mceachern A.D. and Snyder J.* Gender differences in predicting antisocial behaviors: developmental consequences of physical and relational aggression // *Journal of Abnormal Child Psychology*. 40.4 (May 2012), 2012. 501-512. URL: [dx.doi.org/10.1007/s10802-011-9589-0](https://doi.org/10.1007/s10802-011-9589-0).
- Potapova R.K.* From deprivation to aggression: verbal and non-verbal social network communication // VIth International Scientific Conference on “Global Science and Innovation” (18–19, November, 2015). Vol. 1. Chicago. 2015. P. 129–137.
- Potapova R.K. and Komalova L.R.* On principles of annotated databases of the semantic field “aggression” // A. Ronzhin, R. Potapova, V. Delic (eds.) *SPECOM 2014*. LNAI. 8773, Springer, Heidelberg, 2014. P. 322–328. URL: [dx.doi.org/10.1007/978-3-319-11581-8_40](https://doi.org/10.1007/978-3-319-11581-8_40).
- Rinaldi C.* Language competence and social behavior of students with emotional or behavioral disorders. *Behavioral Disorders*. 29.1 (Nov. 2003). 2003. P. 34–42.
- Runions K.* Toward a conceptual model of motive and self-control in cyber-aggression: rage, revenge, reward, and recreation // *Journal of Youth and Adolescence*. 42.5 (May 2013), 2013. P. 751–771. URL: [dx.doi.org/10.1007/s10964-013-9936-2](https://doi.org/10.1007/s10964-013-9936-2).
- Spitzberg B. and Gawron J. M.* Toward online linguistic surveillance of threatening messages // *The Journal of Digital Forensics, Security and Law*. 11.3. 2016. P. 43–77.
- Wah L.L. and Min L.H.* The development and application of an online Malaya language corpus-based lexical database. *Kajian Malaysia*. 32.1. 2014. P. 151–166.
- Werner N.E., Bumpus M.F., and Rock D.* Involvement in Internet aggression during early adolescence // *Journal of Youth and Adolescence*. 39.6 (Jun. 2010). 2010. P. 607–619. URL: [dx.doi.org/10.1007/s10964-009-9419-7](https://doi.org/10.1007/s10964-009-9419-7)

Р. К. Потапова, В. В. Потапов

Р. К. Потапова, доктор филологических наук, профессор; академик Международной академии информатизации; заведующая кафедрой прикладной и экспериментальной лингвистики; директор Института прикладной и математической лингвистики ФАЯ МГЛУ; e-mail: rkpotapova@yandex.ru

В. В. Потапов, доктор филологических наук, главный научный сотрудник НИР кафедры прикладной и экспериментальной лингвистики факультета английского языка МГЛУ; ст. научный сотрудник филологического факультета МГУ; e-mail: volikpotapov@gmail.com

**РЕЧЕВЫЕ БАЗЫ ДАННЫХ
КАК ЧАСТЬ МУЛЬТИМОДАЛЬНЫХ КОРПУСОВ В ИНТЕРНЕТЕ¹**

В статье представлена история развития *речевых баз данных* (РБД), приведены принципы формирования подобных баз, даны примеры РБД, в том числе и базы, разрабатываемые на кафедре прикладной и экспериментальной лингвистики МГЛУ. В связи с развитием мультимодальных баз данных, включающих, наряду с речью, зрительные и слуховые образы, указанные направления приобретают особое значение в аспекте изучения мультимодальных средств передачи информации в Интернете и их воздействия на пользователей Интернета. Речевые базы данных – широкий инструментарий, используемый как в классических дидактических целях, так и в целях разработки новых информационных технологий. Существуют самые разнообразные цели формирования баз данных. Сформированные базы могут иметь как узкое, так и широкое применение. Сегодня спектр речевых БД насчитывает огромное количество узкоспециальных баз. В настоящее время одним из перспективных направлений в развитии различных баз данных является формирование мультимодальных баз данных, отражающих информационное поле Интернета в целом и специфику социально-сетевого дискурса в Интернете. Особое внимание привлекают формы передачи мультимодальной информации, включающей вербальный, паравербальный, невербальный и экстравербальный каналы коммуникации в *социально-сетевом дискурсе* (далее – ССД). Именно данное направление позволит раскрыть специфику как реализации намерений коммуникантов в социально-сетевом дискурсе, так и их индивидуальную оценку происходящих событий, что непосредственно связано с проблемами социального прогнозирования и акмеологической динамики общества в целом.

Ключевые слова: речевая база данных; мультимодальность; Интернет; автоматическое распознавание речи.

¹ Исследование поддержано Российским Научным Фондом (РНФ). Проект № 18-18-00477.

R. K. Potapova

Sc.D., Prof., Full Member of the International Informatization Academy, Head of Department of Applied and Experimental Linguistics, Director of Institute for Applied and Mathematical Linguistics of Moscow State Linguistic University (MSLU); e-mail: rkpotapova@yandex.ru

V. V. Potapov

D. Sc. Chief Researcher, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the Humanities and Applied Sciences of Moscow State Linguistic University (MSLU); Senior Researcher, Faculty of Philology, Lomonosov Moscow State University; e-mail: volikpotapov@gmail.com

SPOKEN LANGUAGE DATABASES AS A PART OF MULTIMODAL CORPUS ON THE INTERNET¹

The article presents the history of development of speech databases (RDB), provides principles for the formation of such databases, gives examples of RDB, including the bases developed at the Department of Applied and Experimental Linguistics at the Moscow State Linguistic University. In connection with the development of multimodal databases, including visual and auditory images along with speech, these directions take on special significance in the aspect of studying multimodal means of information transmission on the Internet and their impact on Internet users. Speech databases are a broad toolkit used both for classical didactic purposes and for the development of new information technologies. There are a variety of purposes for the formation of databases. Formed bases can have both narrow and wide application. Today the spectrum of speech databases has a huge number of highly specialized databases. Currently, one of the promising areas in the development of various databases is the formation of multimodal databases reflecting the information field of the Internet in general and the specificity of social-network discourse on the Internet. Particular attention is attracted to the forms of transmission of multimodal information, including verbal, paraverbal, nonverbal and extraverbal communication channels in *social-network discourse* (SDS). It is this direction that will make it possible to disclose the specifics of both the implementation of the intentions of communicants in the social-network discourse and their individual assessment of current events, which is directly related to the problems of social forecasting and the acmeological dynamics of society as a whole.

Keywords: spoken language data base; multimodality; automated speech recognition.

Введение

Интерес к созданию корпусов звучащей речи был в значительной степени инициирован разработками в области автоматического

¹This research is supported by the Russian Science Foundation, Project № 18-18-00477.

распознавания речи, где исследователям приходится сталкиваться с огромной акустической вариативностью звуковых единиц языка, которая имеет весьма разнообразные источники – от системной контекстной вариативности, обусловленной коартикуляцией, до психофизиологического состояния говорящего или технических характеристик микрофона, который используется при записи речевого материала¹.

Согласно результатам ряда исследований [Потапова 1992; Потапова 2009а; Потапова 2009б; Потапова 2017; Potapova 2011], при создании систем распознавания речи используют два подхода: глобальный и аналитический. При первом подходе осуществляется общее распознавание изолированного слова (отделяемые паузой примерно в 200 с) или связного текста, но с сокращенным словарем и произнесенного одним диктором. Второй подход позволяет приступить к проблеме распознавания слитной речи, произнесенной большим количеством дикторов. Он осуществляется в два этапа: в первую очередь распознаются звуки (речь идет о фонетическом, или аналитическом распознавании), затем обрабатывается информация «высшего» уровня (лексика используемых слов, структуры произносимых фраз).

Процесс распознавания речи (глобальный, или аналитический) начинается с анализа речевого сигнала, предварительно преобразованного в электрический сигнал с помощью микрофона. Электрический сигнал преобразовывается в дискретный, т. е. в последовательность чисел с изменением амплитуды сигнала с соответствующими интервалами с помощью вычислительных машин. Однако чаще предпочитают работать со спектральным представлением сигнала с более продолжительными временными интервалами (около 10 мс).

В системах общего распознавания необходима фаза обучения, во время которой пользователь произносит список слов с нужной лексикой. Для каждого произнесенного слова осуществляется акустический анализ и его результат вводится в запоминающее устройство ЭВМ.

Во время фазы распознавания при произнесении слова пользователем акустический образ распознаваемого слова сравнивается со словами-эталоном. Слово, наиболее схожее с произнесенным, соответствующим образом распознается и воспроизводится.

Именно на речевых корпусах, или речевых базах данных, и проводится обучение систем автоматического распознавания речи (далее –

¹ Цит. по: [Кривнова, Захаров, Строкин 2001].

APP). Первые речевые корпуса появились в середине 1980-х гг. в США, где их разработка финансировалась прежде всего Министерством обороны. При поддержке этого ведомства были созданы: TI-DIGITS корпус (1984) для тестирования систем распознавания изолированных цифр и цифровых последовательностей; Road Rally для анализа и распознавания ключевых слов (word spotting) и King Corpus для систем идентификации говорящего (speaker recognition). В рамках государственной программы развития лингвистических технологий, известной как ARPA / DARPA (the Advanced Research Projects Agency), это же Министерство финансировало создание корпуса TIMIT, который послужил прототипом для многих других речевых баз данных. При этой же финансовой поддержке были разработаны специализированные речевые корпуса Resource Management (RM) и Wall Street Journal (WSJ) для исследований в области распознавания слитной речи, а также Air Travel Information Service (ATIS) для исследования спонтанной речи и понимания естественного языка в диалоговых системах [Кривнова, Захаров, Строкин 2001].

Накопленный к концу 1980-х гг. в. опыт показал, что создание представительных речевых корпусов требует совместных усилий исследовательских институтов, индустриальных компаний и государственных спонсоров. Финансовые и временные затраты на разработку высококачественных ресурсов оказались очень велики. Эксперты отметили, что дорогостоящие, но необходимые для развития информационных технологий ресурсы не должны разрабатываться для какой-то одной специальной системы или задачи [Cole 1997]. Разработчики пришли к выводу, что ресурсы должны обеспечивать возможность их многократного использования разными пользователями, т. е. быть общедоступными, и более чем для одной цели, т. е. быть многофункциональными. В связи с этими требованиями возникла проблема стандартизации лингвистических описаний, согласования форматов представления информации в разных видах лингвистических ресурсов и их типологии [Cole 1997].

Следует, однако, отметить, что чаще РБД разрабатываются именно для конкретной задачи, так как круг возможных применений велик, а конкретная задача задает непосредственные характеристики базы. В настоящее время одним из таких стандартов можно назвать фонетический алфавит Sampa (Speech Assessment Methods Phonetic Alphabet). Он представляет собой Международный фонетический алфавит, записанный символами ASCII с рядом изменений под конкретный язык.

В 1991 г. в США был создан лингвистический консорциум (LDC)¹, который поддерживает создание новых языковых корпусов и распространяет ресурсы, полученные из разных источников. В частности, в настоящее время LDC предлагает более 50 речевых корпусов, включая упомянутые выше, которые в совокупности содержат многие сотни часов звучащей речи. Технологический Центр в штате Орегон (CSLU Center for Spoken Language Understanding) коллекционирует, аннотирует и распространяет телефонные речевые корпуса. Активность Центра поддерживается промышленными спонсорами, но собранные корпуса доступны университетам по всему миру. Этот Центр располагает также многоязычным корпусом для оценки алгоритмов идентификации языка, который состоит из фрагментов спонтанной речи на одиннадцати разных языках мира. В 1995 г. координационный центр лингвистических ресурсов (ELRA European Language Resources Assosiation) был образован и в Европе. В распоряжении этого Центра находятся речевые корпуса для большинства официальных языков Европейского союза: для британского и шотландского вариантов английского языка, голландского, датского, шведского, немецкого, французского, итальянского, испанского, а также несколько многоязычных корпусов. В настоящее время в результате осуществления программы Sorernicus ELRA распространяет также речевые корпуса для языков Восточной Европы (польский, болгарский, эстонский, румынский, венгерский и др.). На сайте Европейской Ассоциации в Интернете можно найти предложения и в области речевых корпусов для русского языка. В их разработке принимала участие Санкт-Петербургская компания «Одитек» [Викторов, Викторова, Воронцова ... 1999].

Принципы создания речевых баз данных (РБД)

Сравнение различных акустических баз данных позволяет сформулировать некоторые обязательные требования к современной фонетической базе данных, предназначенной для фундаментальных и прикладных исследований. Базы данных (далее – БД) для прикладных исследований, в частности в области синтеза и распознавания речи, должны обеспечивать решение следующих задач [Скрелин, Щербаков 2003]:

¹LDC – Linguistic Data Consortium.

- внесение в БД звуковых эталонов – оцифрованных записей речи нормативных дикторов в разных стилях речи, от спонтанной речи и чтения текстов, полученных на основе ее расшифровок, до чтения списка слов. Другими словами, в БД необходимо включить звуковой материал, представляющий максимальную вариативность реализации языковых единиц (фонем и интонационных конструкций) в различных условиях речевой деятельности человека;

- внесение сегментной информации и подробного фонетического описания включаемых звуковых образцов, поскольку необходимо снабдить этот материал подробным описанием: адресами границ звуков и интонационных единиц, словоформ и слогов, так как существуют различные методики распознавания и синтеза речи с точки зрения базовых единиц; а также фонемной и подробной фонетической транскрипцией;

- обеспечение эффективного выполнения запросов к содержимому БД для поиска нужных звуковых фрагментов по их транскрипционным описаниям и указанным в описаниях признакам.

Недостаточная проработка реализации любого из вышеперечисленных пунктов существенно снижает ценность БД в целом [Скрелин, Щербаков 2003].

Запись речевого материала проводится как в лабораторных условиях, так и из доступных источников, например, теле- и радиотрансляции, вещаний в Интернете и т. п. Для записи интернет-трансляций понадобится программа-soft, которая может выполнять функции магнитофона, и программа, поддерживающая и воспроизводящая формат потокового аудио из Интернета. Например, CoolEdit 1.0, который, выполняя функции записи, является одновременно и звуковым редактором; и Windows Media Player, RealPlayer, и т. д.

Чтобы облегчить сбор и хранение данных РБД, разрабатывается специальная оболочка. Она представляет собой отдельную программу-приложение, которая обладает возможностью:

- записи / воспроизведения фрагментов;
- хранения информации о фрагменте;
- хранения информации о дикторе (если такая информация нужна);
- осуществления поиска информации по различным параметрам.

После записи речевого материала, ввода речевого материала в компьютер (оцифровки) и сохранения его, эксперт-фонетист производит транскрибирование материала; файл транскрипции имеет, как

правило, формат .txt. Затем эксперт-акустик проводит сегментацию материала, сохраненного в файлах форматов .wav и .txt, с его последующим сохранением в двух папках, обозначенных соответственно WAVE и TEXT. Эксперт-фонетист создает правила перехода звук–буква, причем звуки представлены специальным алфавитом Sampra, варианты которого создаются для каждого языка. Эксперты создают файлы-примечания, если в их создании есть необходимость (при особенностях отдельных частей материала). База записывается на CD или DVD диски.

3. Примеры баз данных

3.1. Краткое описание репрезентативного речевого корпуса TIMIT (американский вариант английского языка)

Американский корпус TIMIT до сих является прототипическим образцом корпуса репрезентативного типа. Поэтому полезно рассмотреть, каким образом при его создании были решены вопросы содержательного характера. Акустико-фонетический корпус TIMIT предназначался для широких фонетических исследований, а также для разработки и тестирования автоматических систем распознавания слитной речи в рамках американского варианта английского языка. В его разработке принимало участие несколько широко известных организаций и исследовательских центров: MIT¹, SRI² and TI³ и NIST⁴. Это один из первых речевых корпусов, который начал распространяться на CD-дисках. В записи корпуса принимали участие 630 дикторов из 8 региональных диалектных зон США, причем разработчики стремились к одинаковому процентному представительству диалектов, хотя это удалось сделать не для всех зон. Соотношение дикторов по полу также выдерживалось – каждый записанный диалект представляют около 70 % дикторов-мужчин и 30 % женщин. Среди прочих признаков при подборе и записи дикторов учитывались: возраст, рост, расовая принадлежность, образовательный уровень, время записи речи. Текстовый материал TIMIT включает 2342 отдельных предложения. Из них два предложения представляют со-

¹ The Massachusetts Institute of Technology.

² Stanford Research Institute.

³ Texas Instruments.

⁴ National Institute of Standards and Technology.

бой специально сконструированные фразы, насыщенные контекстами, в которых можно ожидать максимального проявления диалектной принадлежности диктора. Остальные 2340 предложений делятся на две группы следующим образом. Первая состоит из 450 специальных фонетически сбалансированных предложений, которые обеспечивают полное покрытие фонемного инвентаря и встречаемость фонем в особых «проблемных» контекстах. 1890 предложений второй группы отбирались из имевшихся текстовых корпусов, критерий отбора – увеличение разнообразия типов предложений и фонетических контекстов употребления фонем [Кривнова, Захаров, Строкин 2001]. В ТИМТ корпусе звуковые файлы, полученные от разных дикторов, разделены на тренировочную и тестовую части. При этом разработчики руководствовались следующими соображениями [Кривнова, Захаров, Строкин 2001]. Тестовая часть корпуса должна составлять от 20 до 30 % всего корпуса, ни один из дикторов не должен участвовать одновременно в обеих частях, в каждой части должны быть представители всех диалектов разного пола, тренировочный и тестовый наборы не должны содержать одинаковых предложений, тестовый набор должен обеспечивать полное покрытие фонемного инвентаря, достаточное разнообразие их фонетических контекстов и частоту встречаемости.

С каждым предложением, как в тренировочном, так и тестовом наборе, ассоциированы четыре разных файла, которые отличаются только расширением и содержат разную информацию о произнесенном предложении. Один из файлов звуковой, а остальные – текстовые. Структура ассоциированных текстовых файлов одинакова и отражает временную соотнесенность (time-alignment) разных языковых объектов с сигналом, т. е. разные уровни его разметки [Кривнова, Захаров, Строкин 2001]. Корпус RuSpeech (www.mdi.ru/aspnews/body/03.12.2001_39303.html).

Это речевая база данных, которая содержит фрагменты непрерывной русской речи с соответствующим текстом, фонетической транскрипцией и дополнительной информацией о дикторах. Cognitive Technologies ставила перед собой цель – создать дикторонезависимую систему распознавания непрерывной речи.

В настоящее время в состав RuSpeech входит более 50 тыс. предложений с фонетической разметкой каждого произнесенного предложения. Для создания корпуса были приглашены 220 дикторов, каждый из которых наговорил в среднем по 250 предложений. RuSpeech

содержит около 50 часов непрерывной речи объемом 15 Гб, которые размещаются более чем на 30 компакт-дисках, что превышает объемы аналогичных речевых баз английского языка WSJ Speech и TIMIT. Речевой интерфейс состоит из системы сценария диалога, синтеза речи по тексту и системы распознавания речевых команд.

База данных для систем верификации / идентификации диктора
(www.auditech.ru)

Речевой материал

Речевой материал состоит из четырех фонетически сбалансированных фраз и 20 слов (названия 10 штатов США и названия 10 российских городов). Ниже приведена таблица фраз и слов и их фонетическая транскрипция (см. табл. 1).

Таблица 1

Пример речевого материала

Фразы, слова	Фонетическая транскрипция
Мама мыла Маню	м«амам»ылам«ан'ю
Саша кусал сало	с«ашакус«алс«ала
Это жирные фазаны ушли под палубу	этаж«ырныйифаз«аны ушл'«ипатп«алубу
Не видели мы такого невода	н'ив'ид«ал'имытак«ова н'«евада
Арканзас	арканз«ас
Алабама	алаб«ама
Калифорния	кал'иф«орн'ийа
Колорадо	калар«ада
Аризона	ар'из«она
Монтана	мант«ана
Джорджия	дж«орджыйа
Миссури	м'ис«ур'и
Вашингтон	вашынкт«он
Оклахома	аклах«ома
Ленинград	л'ин'ингр«ат
Петербург	п'ит'ирб«урк
Москва	маскв«а
Вильнюс	в'«ил'н'ус

Фразы, слова	Фонетическая транскрипция
Владимир	влад'«им'ир
Ярославль	й и р а с л «а в л'
Иваново	и в «а н а в а
Суздаль	с «у з д а л'
Рязань	р'и з «а н'
Новгород	н «о в г а р а т

В фонетической транскрипции знак [«] обозначает ударение и ставится перед ударным гласным, знак ['] обозначает мягкость согласного. Редукция предударных и заударных гласных приведена в соответствии с правилами русской орфографии (московско-петербургского произношения).

Дикторы

Информантами (дикторами) были 39 мужчин и 29 женщин в возрасте от 25 до 55 лет, не имеющие профессиональных навыков чтения в микрофон. (Предполагаемые «наивные» пользователи систем автоматического распознавания речи). Среди них – научные сотрудники, программисты, учителя, врачи, сотрудники радиомастерской, участники хоровой студии.

Записи производились с использованием сменяемых двух электретных и трех угольных микрофонов, вмонтированных в трубку телефонного аппарата, нагруженного на цепь, эквивалентную телефонной линии. Для каждого диктора речевой материал был записан по 10 раз в течение 1 – 1.5 мес. Акустические условия записи: офис, учебный класс, лаборатория. Для 46 дикторов (31 мужчина и 15 женщин) записи были пропущены дополнительно через реальную телефонную сеть. Речевой материал вводился в компьютер через Sound Blaster. Частота дискретизации 16 кГц. Определение начала и конца производилось вручную с помощью программы Marker.exe

Файловая структура:

SPBASE

README.DOC

DESIGN_R.DOC

DESIGN_E.DOC

Структура имени файла:

Файлы записаны в формате Windows PCM (*.wav).

DDYZNNSX

DD	Y	Z	NN	S	X
N диктора	N сессии	N подхода	N слова	Пол	Тип
[00-68]	[1-2]	[0-4]	[0-23]	[M, F]	[E, C, T]

Базы данных в рамках европейских программ SPEECHDAT (II) и SPEECHDAT (E) (www.auditech.ru)

Речевые базы данных для русского языка создавались в рамках европейских проектов SpeechDat(II) и SpeechDat(E). Целью проектов, объединенных названием SpeechDat, является создание речевых баз данных в странах Европы посредством записи речи в реальных условиях через телефонный канал стандарта ISDN. Базы данных призваны служить общим ресурсом для 20-и европейских языков и диалектов и способствовать разработке общих систем телесервиса. В проектах SpeechDat, профинансированных Европейским союзом, были представлены крупнейшие промышленные и академические организации. Все базы данных, созданные в рамках этих проектов, имеют стандартный дизайн и прошли все этапы валидации. Созданные в рамках проектов SpeechDat речевые базы данных удовлетворяют следующим требованиям:

- охватывают фонетически насыщенные слова, слова-команды, словосочетания, числа, цифры, числовые последовательности, фонетически богатые предложения;
- представляют различные типы произношения и стили произношения (команды, речь-чтение и спонтанная речь);
- фиксируют окружающую акустическую обстановку;
- пригодны для разработки и обучения надежных систем распознавания для телесервисов.

В речевой базе данных SpeechDat(II) представлено 48 (в базе данных SpeechDat(E) – 50) слов и выражений, как спонтанно произнесенных, так и прочитанных. Продолжительность записи (диалога между диктором и компьютером) составляла 8–10 минут в зависимости от темпа речи. Исходный словарь базы данных содержит списки наиболее употребительных слов и команд из компьютерной лексики, цифр и цифровых последовательностей, названий крупных городов и фирм,

обозначающих время фраз, дат, денежных единиц, телефонных номеров, номеров кредитных карт, сочетаний «имя–фамилия», фонетически богатых слов и предложений, спеллинг (побуквенное произнесение) слов. Технические характеристики записывающей установки были стандартизированы для всех речевых баз данных. Записи производились в автоматическом режиме через реальный цифровой телефонный канал европейского стандарта ISDN. Сигнал имеет формат: 8 бит, 8 кГц, А-закон. Качество соединения и линии связи характеризовалось отношением сигнал / шум. Непригодные по зашумленности записи исключались.

Обработка речевого материала выполнялась экспертами по речевой акустике. Она заключалась в многократном прослушивании всех звуковых файлов и их аннотации, которая производилась в соответствии со спецификацией, разработанной для участников проекта SpeechDat(II). Аннотация подразумевала внесение следующей информации в файл-метку:

- орфографическая запись высказывания;
- специальные пометки, указывающие на наличие возможных шумов, оговорок, обрывов записи;
- оценка качества записи;
- данные о дикторе (возраст, пол, региональный акцент);
- тип телефонного аппарата;
- тип акустического окружения.

Из всех слов, произнесенных дикторами разборчиво и без оговорок, был составлен лексикон (файл LEXICON) с указанием частоты встречаемости каждого слова и его фонематической транскрипции. Часть слов приведена с вариантами произнесения (разговорный вариант). Полученный лексикон насчитывает около 16500 единиц. Фонематическая транскрипция лексикона выполнена в соответствии с системой символов Russian SAMPA (машинно-ориентированного языка). Кроме того, содержится файл акустического качества каждого речевого сигнала, файл информации о респонденте (пол, возраст, регионально-диалектальная принадлежность), файл содержимого базы данных. Файл DISIGN содержит полное описание базы, ее словаря, записывающей платформы, полную информацию о лексиконе (особенности произношения, частота встречаемости фонем и др.).

Поддержание стандартов качества баз данных, которые были созданы, обеспечено двумя ступенями валидации, которая выполняется

фирмой SPEX (Speech Processing Expertise Centre), созданной в рамках проекта SpeechDat для проверки качества и соответствия стандартам созданных баз данных.

3.2. Речевая база данных французского языка

Данная база данных разрабатывалась на кафедре прикладной и экспериментальной лингвистики МГЛУ по проекту «Корпусная лингвистика многоцелевого назначения». В задачу входило формирование фонетической базы данных французского языка, представленной звучащими текстами. Первой задачей разработки РБД была разработка корпуса правил соотношения «фонема–буква» для французского языка. Правила были сведены в таблицы, состоявшие из четырех классов: фонема, буква / буквосочетание, примеры, примечание. В примечании указывались исключения из правил и различного рода дополнительная информация (см. табл. 2).

Таблица 2

Пример из таблицы «Соотношение “фонема-буква” (носовые гласные)»

Фонема	Буква / буквосочетание	Примеры	Примечание
/W/	In	le vin/lq 'vW /	in /in/ в заимствованных словах :
			le kronprinz
			/lq-krOn-'prints/
			le mackintosh
			/lq-ma-kin-'tOs/
			le meeting
			/lq-mi-'tiŋ/
			la ring/la-'riŋ/
			le smoking
			/lq-smO-'kiŋ/
	Im	le timbre/lq-'tW :br/	
	Ain	le pain/lq-'pW /	
	Aim	la faim /la-'fW /	
	Ein	le peintre	
		/lq-'pW :tr/	

Для оформления таблицы выбран формат «Простой 3». Текст набран шрифтом Times New Roman, 14 кеглем; при транскрибировании фонем использовались фонетические шрифты: NewtonPhoneticNt, Phonetic TM, Phonetic TMUniv, WP Phonetic. Помимо этих правил в этот же корпус входили таблицы французских гласных, согласных, полугласных, носовых гласных; таблица используемых в базе транскрипционных значков международного фонетического алфавита. Информация сохранена в файлах .doc (см. табл. 3).

Таблица 3

Пример базы исследуемых транскрипционных знаков

Международные транскрипционные знаки			
	/ i /	/ m /	
	/ e /	/ ŋ /	
	/ E /	/ k /	
	/ α /	/ g /	
	/ O /	/ O /	
	/ o /	/ Z /	
	/ u /	/ j /	
	/ y /	/ w /	
	/ Q /		
	/ oe /		
	/ q /		
	/ A /		
	/ W /		
	/ P /		
	/ H /		

Второй корпус базы представлен фрагментами французской речи, подлежащей фрагментации до фразового и фонемного уровней, в зависимости от внутренней спецификации задачи. Записи производились программами Cool Edit 2000 и Real Player Plus 8.0. Ряд записей представляет собой оцифрованные записи текстов разного характера с аудиокассет. Условия оцифровки: 22050 Гц, 16 бит, моно. Тексты включают монологи, диалоги, полилоги, театральные постановки и др. в исполнении 25 мужчин и 20 женщин. Другие записи представляют собой сообщения новостей, взятые с разных порталов Интернета, в прочтении 25 мужчин и 20 женщин. Источником служили файлы форматов .m3, .rm.

При записи новостей он-лайн возникли некоторые трудности: при загрузке файлом реального времени .gm происходили изменения бит-рейта, которые отразились на качестве звучания речи. Последующая стадия обработки звука позволяет компенсировать этот недостаток. После записи речь в файлах подлежит сегментации и созданию в отдельных файлах, соответствующих определенному сегменту. Далее проводится аннотирование. Результаты сохраняются на CD-ROM. По ряду текстов имеются и видеозаписи, что существенно расширяет базу данных.

Особенности базы:

- 45 дикторов;
- общее время звучания – 785 минут;
- записи производились в разных условиях.
- РБД для арабского языка

База данных разрабатывалась по проекту «Корпусная лингвистика многоцелевого назначения» («Разработка транскрибированных баз данных для арабского языка» [Потапова, Лебедев, Бобров 2011] по заказу ГПКБ «Спецвузавтоматика»).

Таблица 4

Фонема	Буква / буквосочетание	Примеры		
/h/	ه (ха) лунная	إله	[?ila:h]	божество
		رهوان	[rahwa:n]	иноходец
		مدهش	[mudhiS]	удивительный
		ذهب	[Dahaba]	уходить, отправляться
		زهذ	[zahada]	отказываться
		وهب	[wahaba]	одарять

В задачу входило формирование фонетической базы данных арабского языка, представленной звучащими текстами. Первой задачей разработки РБД была разработка корпуса правил соотношения «фонема–буква» для арабского языка.

Помимо правил база содержала:

- папку с файлами текста (.txt);
- папку со звуковыми файлами (.wav);
- папку со звуковыми файлами неотсегментированного материала (тренировочный комплекс).

Исходным материалом служили файлы, представленные в следующих двух таблицах.

Таблица 5

**Данные о числе файлов, времени их звучания
и источниках звучащей речи
(общее время звучания – 8,5 часов; дикторы: М – 41; Ж – 16)¹**

№	Файл	Время звучания**	Дикторы	Источники звучащей речи
1	1 a	45:17	7m,1f	Aljazeera
2	1 b	22:46	5m	Aljazeera
3	2 a	30:10	8m,3f	Aljazeera
4	2 b	14:16	7m,	Aljazeera
5*	3 a	45:34	4m,3f	London course of Arabic
6*	3 b	15:31	4m,3f	London course of Arabic
7*	4 a	45:05	4m,3f	London course of Arabic
8*	4 b	11:28	4m,3f	London course of Arabic
9*	5 a	45:32	4m,3f	London course of Arabic
10*	5 b	13:04	4m,3f	London course of Arabic
11*	6 a	46:09	4m,3f	London course of Arabic
12*	6 b	12:37	4m,3f	London course of Arabic
13***	7 a	45:21	3f	Alarabia
14***	7 b	45:09	3f	Alarabia
15****	8 a	46:08	10 m,6f	Aljazeera
16****	8 b	28:16	10m,6f	Aljazeera

Ниже представлены оцифрованные записи (22050Hz, 16 bit, mono) с аудиокассет. Все файлы типа Windows PCM (.wav) (см. табл. 6).

Таблица 6

**Данные по оцифрованным файлам
(время звучания отсегментированного материала 6,5 часов;
представлена на 2 CD; дикторы: 59 мужчин и 20 женщин).**

№	Файл	Время звучания**	Дикторы	Источники звучащей речи
1	Aljazeera1	13:42	5m	Aljazeera
2	Bbc1	9:48	2m, 1f	BBC
3	Dw2	30:19	7m,1f	Deutsche Welle
4	Jaber Ibn Hayan	27:26	4m,2f	VOA

¹ **Примечание:** * – в файлах с 5 по 12 представлены одни и те же дикторы. В данном наборе файлов представлен лондонский курс арабского языка.

** – время звучания представлено в формате минуты: секунды (мм: сс)

*** – в файлах с 13-14 представлены одни и те же дикторы.

**** – в файлах 15 и 16 представлены одни и те же дикторы.

Заключение

Речевые базы данных – широкий инструментарий, используемый как в классических дидактических целях, так и в целях разработки новых информационных технологий. Существуют самые разнообразные цели формирования баз данных. Сформированные базы могут иметь как узкое, так и широкое применение. Сегодня спектр речевых БД насчитывает огромное количество узкоспециальных баз. Базы для широкого спектра решаемых задач требуют хорошего финансирования, значительной доли участия в их создании человеческого фактора. Успешность созданных баз определяется многими факторами – как внутренними, так и внешними. Например, для русского языка наиболее проблематично создать речевую базу из-за сложности грамматики и фонетики. Так или иначе, создание речевых БД имеет большое будущее в свете развития новых информационных технологий, о чем свидетельствует большое наличие узкоспециальных БД, которые, возможно, будучи обобщенными, сформируют в будущем полифункциональную базу данных специального назначения. Так, например, в настоящее время одним из перспективных направлений в развитии различных баз данных является формирование мультимодальных баз данных, отражающих информационное поле Интернета в целом и специфику социально-сетевому дискурса в Интернете, в частности [Потапова 2017; Potapova, Bobrov 2015; Potapova, Potapov 2016; Potapova, Potapov, Vazhenova 2015 и др.]. Особое внимание привлекают формы передачи мультимодальной информации, включающей вербальный, паравербальный, невербальный и экстравербальный каналы коммуникации в социально-сетевом дискурсе (ССД). Именно данное направление позволит раскрыть специфику как реализации намерений коммуникантов в ССД, так и их индивидуальную оценку происходящих событий, что непосредственно связано с проблемами социального прогнозирования и акмеологической динамики [Potapova, Potapov 2017 и др.] общества в целом.

СПИСОК ЛИТЕРАТУРЫ

Викторов А. Б. [и др.]. Речевые базы данных для задач автоматического распознавания речи и верификации говорящего // *Современные речевые технологии: сб. тр. IX Сессии Российского акустического общества.* М. : Геос, 1999. С. 142–145.

- Кривнова О. Ф., Захаров Л. М., Строкин Г. С.* Речевые корпуса (опыт разработки и использование): сб. 2001 (Dialog). М.: МГУ, 2001. URL: www.dialog-21.ru/digest/2001/articles/krivnova/
- Потапова Р. К.* Тайна современного кентавра. М.: Радио и связь, 1992. 248 с.
- Потапова Р. К.* Основные тенденции развития многоязычной корпусной лингвистики (часть 1) // Речевые технологии. 2009а. № 2. С. 92–114.
- Потапова Р. К.* Основные тенденции развития многоязычной корпусной лингвистики (часть 1) // Речевые технологии. 2009б. № 3. С. 93–112.
- Потапова Р. К.* Речевая коммуникация в информационном пространстве. М.: Ленанд, 2017. 112 с.
- Потапова Р. К., Потапов В. В.* Основы многоаспектного исследования «электронной личности» по голосу и речи в информационно-коммуникационной среде Интернета // Человек: Образ и сущность. Гуманитарные аспекты. М., 2017. № 1–2 (28–29). С. 87–111.
- Скрелин П. А., Щербаков П. П.* Требования к современной фонетической базе данных для фундаментальных и прикладных исследований // Технологии информационного общества – Интернет и современное общество: тр. VI Всероссийской объединенной конференции. Санкт-Петербург, 3–6 ноября 2003 г. СПб.: Изд-во филол. ф-та СПбГУ, 2003. С. 62–63.
- Cole R.* (ed. in chief). Survey of the state of the art in human language technology. Cambridge: Cambridge University Press and Giardini, 1997. 526 p.
- Potapova R.* Multilingual spoken language databases in Russia // R. Potapova, A. Ronzhin (eds.) Speech and Computer 2011 (SPECOM'2011). Kazan, 2011. P. 13–17.
- Potapova R., Bobrov N.* Versatile linguistic database annotation: practical issues and a new flexible approach // Fakotakis N., Ronzhin A., Potapova R. (eds.) Proc. of the 17th Intern. conference “Speech and Computer 2015” (Athens, Greece September 20–24, 2015). University of Patras. Vol. II. P. 41–53.
- Potapova R., Potapov V.* Polybasic attribution of social network discourse // Ronzhin A., Potapova R., Németh G. (eds.) SPECOM 2016. LNCS. Vol. 9811. Heidelberg: Springer, 2016. P. 539–546.
- Potapova R., Potapov V.* Human as acmeologic entity in social network discourse (multidimensional approach) // A. Karpov, R. Potapova, I. Mporas (eds.) SPECOM 2017. LNAI. Vol. 10458. Cham: Springer, 2017. P. 407–416.
- Potapova R., Potapov V., Bazhenova I.* Development of the research cloud technology stand-alone system (regarding integrated speech databases) // N. Fakotakis, A. Ronzhin, R. Potapova (eds.) Proc. of the 17th Intern. conference “Speech and Computer 2015” (Athens, Greece September 20–24, 2015). University of Patras. Vol. II. P. 1–7.

АКТУАЛЬНЫЕ ПРОБЛЕМЫ АНАЛИТИЧЕСКОЙ ТЕКСТОЛОГИИ

УДК 81'33

А. В. Джунковский

аспирант кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики факультета английского языка МГЛУ;
e-mail: Vetinari01@gmail.com

СТЕГАНОГРАФИЯ: МЕТОД ТРЕХСТУПЕНЧАТОГО АНАЛИЗА ДЛЯ ПИСЬМЕННЫХ ТЕКСТОВ РУССКОГО ЯЗЫКА

В статье рассматривается стеганоаналитический метод раскрытия смысловой информации с помощью лингвистических методов в письменных текстах русского языка на основе выявленной типологии стеганографических методов. Наш трехступенчатый метод стеганоанализа включает стадии мета-анализа текста, непосредственную стадию лингвистического анализа, а также стадию контекстуального экстралингвистического анализа семантических элементов текста. Метод позволяет добиться высокой вероятности выявления признаков сокрытия смысловой информации лингвистическими методами в письменных текстах русского языка. Метод трехступенчатого стеганоанализа был выработан нами на основе синтеза и совершенствования результатов типологического анализа методов стеганологии отечественных и зарубежных ученых. Выработанный метод на данный момент используется в наших исследованиях по экспериментальному выявлению эффективности отдельных лингвистических методов стеганографии для письменных текстов русского языка. Актуальность статьи обусловлена необходимостью выработки эффективных методов стеганоанализа для письменных текстов русского языка. В настоящее время существует некоторое количество работ по стеганоанализу звучащей речи, однако исследование стеганоаналитических методов для письменных текстов русского языка остается темой, требующей освещения.

Ключевые слова: стеганография; стеганоанализ; контейнер; шифр; уязвимость стеганографических методов; восприятие человеком; методология стеганоанализа.

Dzhunkovskiy A.V.

Postgraduate Student, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the English Language, MSLU;
e-mail: Vetinari01@gmail.com

STEGANOGRAPHY: THREE-STAGE ANALYSIS METHODOLOGY APPLIED TO RUSSIAN WRITTEN TEXTS

In this paper we present our findings concerning the typology of various steganalytical means applicable for the purpose of uncovering information in Russian written texts hidden through means of linguistic steganography. Our threefold steganalysis method includes three major stages: meta-analysis of the text, linguistic analysis proper, and contextual extralinguistic analysis of the semantic elements of the text. The results the method yields allow for high probability of detecting signs of concealment of semantic information through linguistic means in written Russian texts. We have created and conceptualized the three-stage analysis methodology on the bases of analyzing, synthesizing and improving upon the results of typological analysis of steganological methods carried out by foreign and domestic scientists. The incepted method is currently used in our further experimental research with the goal of determining the efficiency of various linguistic steganographic methods for the purpose of concealing information in Russian written texts. The relevance of the present paper is based on the need of developing effective steganoanalytical methods for written Russian texts. Presently, only such studies of steganalysis exist that deal with spoken language, and there is little research that deals with steganalysis and steganography in written Russian texts. We present this methodology as basis for investigating the efficiency of various methods of steganography and steganalysis and simultaneously a tool for conducting practical work on analyzing and concealing information in written Russian texts using linguistic methods.

Key words: steganography; steganalysis; container; cipher; steganographic method vulnerability; human perception; steganalysis methodology.

I. Introduction

Contemporary advancements in computational technologies brought about a noteworthy phenomenon. Cryptography quickly gained relevance and is now slowly becoming suboptimal as a means of achieving the goal of confidential information protection. The reason for this is the fact that cryptography and cryptanalysis, the practice of unauthorized decryption, have come to depend on automation. The very nature of cryptography dictates that encryption must follow strict, if often obscure and unintuitive, mathematical patterns. Thus, both encryption and decryption of information by crypto-means is nowadays a question of computational power and algorithmization. While there are many practical applications to crypto methods the usefulness of which cannot be denied, the evolution of these methods has made them nigh unusable for those who do not possess the necessary technology and are still in need of discreetly transferring confidential information through unsecure channels [Potapova 2010, p. 120].

The aforementioned information is crucial for the purpose of understanding the significance of steganography and steganalysis, which we collectively refer to as steganology. Steganography is by its very nature highly protected from automatic and computer-aided attempts of successful attacks (term used for steganalytical and cryptanalytical attempts of gaining unauthorized access to protected information) and requires a human expert [Алферов и др. 2012, p. 67]. Steganalysis of large volumes of information is extremely prohibitive as it involves high costs to maintain a large number of highly skilled experts. This makes steganology a promising field of research.

We focus our efforts on linguistic (or text) methods of steganography in Russian language. While the field of steganology is underdeveloped as of now, this statement is even truer for linguistic steganography [Bennet 2013, p. 15]. When it comes to research limited to text steganography on the materials of Russian language, no papers or studies of note are to be found as of yet. That being said, a sufficient, albeit small number of papers dedicated to steganographic means in other languages [Wayner 2002, p. 78], as well as such means in spoken Russian texts exist, granting us an opportunity to develop a steganographic and steganalytical methodology for Russian written texts.

Furthermore, having established that steganographic methods are heavily resistant against automatic attacks [Лукьянов 2005, p. 7], we undertook to focus on analyzing the efficiency of various steganographic methods against visual attacks carried out by human experts.

II. Methods

The main methods of developing the three-stage steganalysis methodology were analysis and synthesis using the scarce contemporary achievements of experts in steganography and steganalysis working with various languages [Xiang, Sun, Luo 2014, p. 1894] to develop classifications of steganalytical [Katzenbeisser 2000, p. 87] and steganographic [Wayner 2005, p. 163] methods. No such work has been undertaken for Russian steganography applied to written text before now. Furthermore, the existing stego methodologies based on these typologies are, as a rule, untested experimentally [Encyclopedia of Cryptography... 2014] and vary vastly from each other. In view of this, we have synthesized the existing typologies in order to create both steganalytical and steganographic methodologies for encrypting and decrypting information in Russian written texts.

III. Results

We discovered that no classification of methods of visual steganalysis and methods of steganography aimed at counteracting visual attacks existed. Thus, we have created such classifications.

In general, their idea is as follows: for visual steganalysis, we separate the existing methods into three types, which correspond to the steps in the methodology.

The first is the metanalytical step during which the format and design elements of the text are examined. Step two is the main linguistic analysis. Finally, step three is the extralinguistic analysis, during which the contents of the message are checked against objective reality.

The classification of methods of steganography applicable to written Russian text is symmetrical to the steganalytical classification. The first category is the alteration of metaelements, the second includes linguistic alterations, and, finally the third – extralinguistic alterations.

In this manner, our classification efforts have yielded sufficient basis for having developed two distinct methodologies for steganalysis and steganography correspondingly.

The three-stage steganalytical method we have developed is as follows:

1. Text design and formatting analysis (The meta-analysis stage)
 - a. Letter, word, and paragraph line spacing size analysis;
 - b. Vertical alignment analysis of words and letters on the same line;
 - c. Font size, color and type analysis;
 - d. Line spacing analysis;
 - e. Text and individual text elements inclination analysis;
 - f. Individual grapheme feature analysis (especially relevant to handwritten text);
 - g. Grapheme connections analysis (When dealing with handwritten text).
2. Text linguistic features analysis (Linguistic analysis proper)
 - a. Spelling peculiarities analysis;
 - b. Peculiarities on the morpheme level;
 - c. Peculiarities on the lexical level;
 - d. Peculiarities on the syntactic level;
 - e. Usus errors analysis on all language levels;
 - f. Stylistic features analysis;
 - g. Diachronic usus consistency analysis;
 - h. Content-analysis.

3. Extra-linguistic text peculiarities analysis (context and semantic analysis stage)
 - a. Comparative analysis of the given text with other text by the same author;
 - b. Analysis of the consistency of the message with the objective reality;
 - c. Analysis of the consistency of particular minutiae in the text to objective reality;
 - d. Analysis of the adherence of the text style to the social hierarchy status (when cognizant of the identities of communicating parties, their social statuses and peculiarities of interpersonal communication).

This three-stage steganalytical method provides human experts with strong methodological basis for conducting their work. We shall now move forward to the discussion of the presented methodology.

IV. Discussion

The three-stage steganalytical method allows for a thorough analysis of written text in Russian when searching for signs of alterations indicative of information hidden within. It allows one to consider the text in three distinct ways. First, as a physical object in the stage of meta-analysis, the text as a linguistic product in the second stage of linguistic analysis proper, and, finally, as a part of the larger context and in connection with it.

Of course, the presented methodology based on our typology can hardly be considered universal and final at this stage, as we foresee necessity for alteration and further development. That being said, the present results have proven to be rather useful in conducting experiments aimed at assessing the efficiency of various steganographic containers.

The first stage of the analysis deals with the meta-elements and is the only stage where human eye may fail to see certain peculiarities evident during automatic steganalysis. With this in mind, we conclude that at the first stage, it is prudent to combine manual and automated means of analysis. A question may arise as to why the involvements of the human eye can be considered beneficial in analyzing these aspects. Result proofing aside, at the present stage automated steganalysis of handwritten text is a question of future advancements. The state of OCR, let alone steanalysis of handwritten text, is insufficient to be used as a standard procedure. Therefore, the need for visual human expert analysis arises.

The second stage of the analysis is something that linguists, medical and criminal experts have been working on for a considerable amount of time. As such, our main contribution here is simply synthesizing all of the distinct elements of the analysis that already exist into one comprehensive linguistic analysis technique.

The final stage of the analysis deals with the connection of the text and its contexts to objective reality, the identities of the author and the addressee, and of general and specific validity of presented information.

The main downside of the present methodology is the same that all steganalytical methods share: in best case scenarios, the expert may only draw conclusions on the possibility that certain text peculiarities may contain a hidden message within. Another downside is the fact that this methodology has not been tested in other languages or in spoken Russian texts. We suspect that large alterations are required for our methodology to function in these cases.

Nevertheless, we have grounds to believe our methodology to be one of the most promising existing methods of steganalysis usable for Russian written texts.

V. Conclusions

The presented three-stage analysis method creates basis for analyzing the efficiency of individual means of steganalysis. The incepted method is currently used in our further research where we employ the presented methodology as a means to classify various steganalysis techniques during experimental work on determining viability of various alterations in written Russian texts as containers for concealing information.

While we remain cautious about the prospects and ready to improve upon the methodology, it has already proven to be fruitful as a tool for conducting further research internally.

The core of the methodology allows one to view the text as three different entities simultaneously: as a physical object, as a self-contained language product, and, finally, within a larger text nexus. As such, it allows to regard a text at a meta-object consisting of a plethora of potential containers where additional information could be concealed.

The nature of our research not only improves the theoretical basis of steganology, but has practical value for both concealing and revealing such information.

REFERENCES

- Алферов А. П.* [и др.]. Основы криптографии. М. : Гелиос АРВ, 2012. 480 с.
- Лукьянов Г. В.* Основы кодирования и криптографического преобразования информации. М. : Сарма, 2005. 128 с.
- Потапова Р. К.* Речь: коммуникация, информация, кибернетика. М. : УРСС, 2010. 600 с.
- Bennet K.* Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text // CERIAS Tech Report. 2013. № 4. P. 1–30.
- Encyclopedia of Cryptography and Security. Second Edition // van Tilborg H.C.A., Jajodia S. Berlin : Springer Science & Business Media, 2014. 1416 p.
- Katzenbeisser S.* Principles of Steganography. Boston : Artech House, 2000. 186p.
- Wayner P.* Disappearing Cryptography: Information Hiding. San Francisco : Morgan Kaufmann, 2002. 221 p.
- Wayner P.* Strong Theoretical Steganography. Berlin : Cryptologia, 2005. 410 p.
- Xiang L., Sun X., Luo G.* Linguistic Steganalysis Using the Features Derived from Synonym Frequency // Multimedia Tools and Applications. 2014. № 3. P. 1893–1911.

УДК 81'32

В. А. Долинский

доктор филологических наук, профессор кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики факультета английского языка МГЛУ; e-mail: vdolinsky@yandex.ru

СЕМЕЙСТВО РАНГОВЫХ РАСПРЕДЕЛЕНИЙ В КВАНТИТАТИВНОЙ ЛИНГВИСТИКЕ

В статье рассматривается семейство ранговых распределений в квантитативной лингвистике: частотные характеристики слов, полисемия, ассоциативные поля слов, стилеметрия, языковые различия, распределения языков и другие объекты изучения в квантитативном аспекте.

Анализируются языковые модели, развиваемые в русле лингвистического ассоцианизма, синергетики и квантитативной лингвистики.

Структурообразующим компонентом значения слова является частотность вызываемых им ассоциативных реакций. Ассоциативные поля слов раскрывают психологически адекватную картину мира носителей языка.

Эффект концентрации и рассеяния, характерный для частотной структуры текстов, имеет своим источником, по-видимому, такой механизм порождения текста, который обусловлен ассоциативной природой языкового сознания.

В качестве иллюстраций к статье приведены таблицы и графики различных ранговых распределений в квантитативной лингвистике.

Ключевые слова: квантитативная лингвистика; корпусная лингвистика; частотно-ранговые распределения; закон Ципфа; ассоциативные поля слов; лексика; семантика; стилеметрия; языковые различия; распределения языков.

V. A. Dolinsky

Doctor of Philology, Professor, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the English Language, MSLU; e-mail: vdolinsky@yandex.ru

FAMILY OF RANK DISTRIBUTIONS IN QUANTITATIVE LINGUISTICS

The article considers the family of rank distributions in quantitative linguistics: word frequency characteristics, polysemy, word association fields, stylometrics, language distinctions, language distributions and other objects of study in quantitative aspect.

The language models being developed within the framework of linguistic associationism, synergetics, and quantitative linguistics are analyzed.

The frequency of association responses evoked by a word is a structural component of the word's meaning. Word association fields reveal the psychologically relevant worldview of native speakers.

The concentration and diffusion effect that is inherent in the frequency structure of texts appears to derive from a text generation mechanism that is driven by the associative nature of language consciousness.

Tables and graphs of different rank distributions in quantitative linguistics are given as illustrations to the article.

Keywords: quantitative linguistics; corpus linguistics; rank frequency distributions; Zipf's law; word association fields; vocabulary; semantics; stylometrics; language distinctions; language distributions.

Распределение в статистике определяется как перечисление значений случайной величины и их вероятностей. Распределение – это разбиение совокупности лингвистических явлений на классы. В широком смысле распределение можно понимать как: «упорядоченную совокупность количественно выраженных значений, т. е. результатов измерения объекта, обычно с указанием значимости (частоты, вероятности, ранга) этих значений в данной совокупности» [Тулдава 1987, с. 41].

Распределение является и процессом, и результатом такого разбиения, группировки, упорядочения. Распределение – это ряд значений признака лингвистического явления и численностей, частот этих явлений [Алексеев 2001, с. 26–27].

Ранговым называется распределение, в котором значения случайной величины расположены в порядке их убывания.

Лексическая система языка, состоящая из необозримого множества слов, распределяет в речи свои единицы по принципу «согласованного оптимума» (*илл. 1*). Первым, кто обратил внимание на удивительную гармоничность (*parsimony, equilibrium* – упорядоченность в частотном распределении) в речевой реализации лексического «инвентаря» разных языков, казавшегося ранее «свалкой» неупорядоченного словесного материала, был выдающийся американский филолог, психолог, математик Джордж Кингсли Ципф (1902–1950). Основоположник квантитативной лингвистики и пионер естественнонаучного подхода к языковым явлениям и процессам языковой саморегуляции, Дж. К. Ципф писал:

Речь не может и не должна быть перманентно отвлечена от живых организмов и их поведения. Речевые явления не могут быть изолированы ни от содержания речи, ни от индивидуального, социального и культурного фона говорящего. Продукт речи должен рассматриваться как естественное психологическое и биологическое явление и изучаться

в объективном духе точных наук и их методами. Наш главный метод исследования – это применение статистических принципов к наблюдаемым явлениям потока речи. <...> Задача, не менее стоящая того, чтобы за нее взяться, так как ... ее решение может пролить желанный свет на многие философские вопросы о природе жизни и смерти вообще. <...> Тенденция, которую отражают наши формулы, – на мой взгляд, результат «привычки» или «накопленной силы случая», – может рассматриваться как «стремление к устойчивости», понимая под устойчивостью (*perseveration*), следуя Скиннеру, психологический термин, описывающий часто наблюдаемый феномен. Я уверен, что уравнивающие силы, подсказанные нашими формулами, – шире и глубже, и что они действуют, чтобы вызвать равновесие ради равновесия. Ни в каком отношении нельзя абстрагировать поток речи и всё то, что по традиции и фактически является частью речевого поведения, от всего остального нашего бытия. В самом деле, если кто-нибудь склонен верить, что его поток речи не связан с любым данным органом его анатомии, дайте ему только уколоться иголкой, раскаленной докрасна, в данный орган и наблюдайте возможный отзвук оттуда в его речепроизводстве <...>. Законы жизненного процесса хорошо могут быть обоснованы как изоморфные законам речи. Это убеждение приобретает прочность, если посмотреть на поток речи, как на равнодействующую множества сил, на которые мы реагируем. Линейность речевых привычек поразительна, даже в латентной речи и ассоциативно связанных словах» [Zipf 1949].

Ни одна квантитативная модель в лингвистике не вызывает столь разнообразных интерпретаций и оценок, порой противоположных, как модель Ципфа [Арапов 1988; Altmann 1997; Herdan 1966; Köhler, Altmann, Piotrowski 2005; Долинский 2012]. Ряд аналогий из смежных наук (закон Виллиса в биологии, закон Парето в экономике и др.) позволяет предположить, что закон Ципфа является одним из первых, математически сформулированных, законов в гуманитарном знании: обратно пропорциональная связь между частотой и рангом слова. Аналогично ранговому распределению слов в тексте, характеризуются встречаемость химических элементов в земной коре, распределение обитателей мирового океана по весу, космических тел по массе, распределения городов по численности населения, фирм по количеству служащих, что заставляет рассматривать этот тип распределения как универсальный «закон предпочтения», «закон иерархии значимостей» или (в терминах Ципфа) – «закон наименьшего усилия». В природе, обществе и языке действует некий универсальный механизм,

который абсолютно разнородные объекты выстраивает по иерархии (или ранжирует) сходным образом. Причины этого не вполне ясны, несмотря на обилие гипотез [Орлов 1980; Тулдава 1987; Чайковский 2001]. Представляется обоснованным вывод о том, что закон Ципфа является не одним из многих эмпирических распределений, но теоретическим законом, имеющим математическую базу в теории устойчивых негауссовых распределений.

Одним из реальных оснований применения количественных методов в изучении языка и речи (текста) следует признать объективную присущность языку количественных признаков, количественных характеристик [Алексеев 2001; Арапов 1988; Долинский 1988; Мартыненко 2009; Пиотровский 2006; Köhler, Altmann, Piotrowski 2005]. Повторяемость (рекуррентность, периодичность) языковых, в том числе лексических единиц, их воспроизведение в различных текстах является наиболее важным условием квантификации языкового материала и применения методов количественной математики для его анализа. Количественный подход способен охватить лишь определенный аспект языка и речи. Но это важный аспект, отражающий многие стороны речевой деятельности, которые невозможно обнаружить чисто качественным анализом. Количественный анализ иногда упрощает языковую реальность (например, когда не учитываются языковой полиморфизм, поливалентность и многообразие оттенков значений). Но при этом подходе возможен и более дифференцированный анализ полисемичности и других свойств языковых единиц. Часто оказывается, однако, что «переплетение» ассоциаций элементов языка настолько сложно и бесконечно, что оно не может поддаться в полной мере не только количественному, но и качественному анализу. Кроме того, следует указать на неизбежный недостаток качественного анализа, который нередко остается на уровне субъективных, произвольных, интерпретаций.

Выявление частотных свойств текстов и построенных на них словарей, установление связей между словами в парадигматике и синтагматике, изучение факторов порождения текста – должны вести к обобщению, упорядочению и осмыслению эмпирического материала на более высоком теоретическом уровне. Конечной целью является синтетический, интегральный, подход к изучению (описанию и объяснению) количественных свойств системы речевой деятельности и языкового механизма в целом – в неразрывном единстве количественного анализа с качественной интерпретацией.

Квантитативная лингвистика придает математическую форму языковым механизмам, порождающим структуры и зависимости. Поскольку язык принадлежит к классу самоорганизующихся систем, в нем всё взаимосвязано, но не всегда напрямую. Связи, обнаруживаемые в статистических распределениях, формируют цепи зависимостей. Чем больше расстояние между двумя элементами цепи, тем слабее зависимость. Изучение зависимостей (часто взаимозависимостей), установление закономерностей – важный аспект квантитативных исследований.

Одной из задач квантитативной лингвистики является составление *частотных словарей*, необходимость использования которых для решения прикладных и теоретических задач постоянно возрастает. Частотный словарь – это модель особым образом преобразованного текста, модель рангового распределения частот употребления единиц в тексте. Частотный словарь включает в себя упорядоченный список слов или других языковых единиц (словоформы, словосочетания), которые зарегистрированы составителем в обследованном им тексте или корпусе текстов и снабжены данными о частоте их употребления. (В конкордансе к этим данным добавляются и данные об их *адресе* и *контекстах*). С помощью частотных словарей можно ответить на вопросы: каков вокабуляр (лексический состав) языка (текста), с какой интенсивностью слова используются в речи, какие из них употребительнее (предпочтительнее) в той или иной сфере коммуникации. Частотные словари находят применение в лингводидактике, психолингвистике, терминографии, когнитивистике, культурологии, медицине, юриспруденции, криминологии и т.д. [Алексеев 2001; Частотный 1977; Ляшевская. Шаров 2009].

Хорошо известно, что слова по частоте их употребления (хотя и весьма неустойчивой для каждого отдельного слова) распределяются в любом крупном наборе (корпусе) текстов в виде «гиперболической лестницы». Например, в «Частотном словаре русского языка» [Частотный словарь 1977] почти 10 % словоупотреблений приходится на первые три слова (*в, и, не*); еще 10 % – на следующие 8 слов (*на, я, быть, что, он, с, а, как*); а 50 % всех словоупотреблений составляют всего 213 лексем (в этом словаре, построенном на текстах объемом 1 056 тыс. словоупотреблений, учтено более 39 тыс. разных слов). При этом более 13 тыс. слов (33,7 %) употреблено по одному разу.

Сравнение частотных словарей указывает, что нет никакой сходимости частот даже для самых употребительных слов, и их дисперсия превышает нормальную в сотни раз. На илл. 2–4 приведены данные разных частотных словарей и подвыборок одного словаря.

Если из подсчетов, выполненных на языковом материале, удастся сделать качественные выводы, то в них сами числа не играют существенной роли, важны исключительно определяющиеся этими числами отношения *порядка*. На шкале порядка различаются степени интенсивности признака или свойства, происходит ранжировка единиц по степени проявленности. Порядковая шкала допускает операции больше / меньше, равенство / неравенство. Содержательный смысл имеет не сама по себе частота употребления слова, число производных от него или число значений, а найденное по этим величинам место данного слова среди других слов, его относительная *ценность*.

Количественным измерениям подвергаются и такие характеристики языковых единиц, как их «объем» в плане выражения и в плане содержания. Это длина слова (морфемы, словосочетания, предложения) и число его значений (выраженное, например, в словарных ЛСВ). Положение слова на ценностной шкале связано с его количественными характеристиками. Ранг слова, отражающий его употребительность в тексте (корпусе текстов), содержит информацию о положении слова в линейно упорядоченном словаре. Все данные о связи между рангом и другими количественными характеристиками единиц имеют вид нелинейных уравнений регрессии. Эти зависимости как общие тенденции следует рассматривать не как точно установленные факты, а, скорее, как «полезные метафоры». Таковы зависимости между рангом i слова (определенном на основании его частоты F_i), длиной l_i , числом значений k_i . В общем случае, например, чем выше ранг слова, тем оно, в среднем, короче и многозначней [Арапов 1988; Altmann 1997].

Квантитативные параметры *полисемии*, измеренные по данным толковых словарей, варьируют в зависимости от объема словаря (большой, средний, малый). Как правило (но далеко не всегда) увеличение объема толковательной статьи сопровождается более «дробным» членением семантики слова (илл. 5).

Толкование дает лишь приближенное представление о значении слова. Число частей в нем – только косвенная оценка сложности значения, как число ломтей, на которые можно разрезать пирог, – косвенная характеристика его размеров [Арапов 1988, с. 137].

Реализацией свойства слова иметь некоторое число значений является *семантический объем* слова, который подлежит количественному измерению. Многозначное слово (лексема) как единица системы языка представляет собой единое смысловое целое, объединяющее в плане содержания ряд лексико-семантических вариантов, или отдельных «значений» слова. Опыт изучения полисемии показывает, что в рамках одного толкового словаря разграничение значений слов проводится достаточно последовательно, о чем свидетельствуют удачные попытки изучения общих квантитативных закономерностей полисемии на основе таких словарей [Поликарпов 1987; Тулдава 1987]. Данные различных языков показали, что семантический объем слова и количество слов с данным семантическим объемом в словаре статистически связаны: наибольшую долю в словаре составляют однозначные слова, затем следуют (в порядке убывания численности) группы слов с двумя, тремя и т. д. значениями (*илл. 6*). Такой вид распределения представляет собой, по-видимому, универсальную квантитативно-системную характеристику полисемии естественных языков. Здесь, как и во многих других ранговых распределениях, проявляется известный принцип *концентрации и рассеяния* языковых единиц.

Ряд актуальных проблем стилистики выделяется с позиции квантитативного анализа текстов (*стилеметрия*): лексическое «богатство» автора, классификация текстов на основе квантитативных параметров, лексическая близость и атрибуция текстов. Частота слова в авторских текстах – важная характеристика значимости этого слова. В первом приближении можно считать, что наиболее частые слова являются и наиболее важными. Однако при оценке авторских текстов, из принципа «значимое – часто» не следует, что «частое – значимо». В противном случае самым значимым словом для большинства языков служил бы союз *И* (или предлог *В*). Для того чтобы оценить ценность, значимость того или иного слова, следует сравнить его ранг в частотном словаре данного автора с рангом в частотном словаре, составленном по соответствующему классу текстов исследуемого периода.

На *илл. 7* приведено распределение персонажей романа М. А. Булгакова «Мастер и Маргарита» по частоте их упоминаний [Чайковский 2001]. На *илл. 8* дано распределение лексем с семантикой цветообозначения (цветовой спектр, в %) в текстах Н. В. Гоголя [Белый 1934].

Наряду со словарным толкованием слова, дающим логически непротиворечивое представление о его семантике в виде упорядоченного и структурированного текста, *ассоциативное поле* слова, полученное в эксперименте с группой носителей языка, также является своего рода текстом, манифестирующим смысл стимульного слова, раскрывающим содержание языкового сознания во всем его явном и имплицитном преломлении [Долинский 1988; 2012].

Важной областью использования данных психолингвистических экспериментов стало изучение языкового сознания и культурноязыковых архетипов. Слово в сознании человека всегда связано бесчисленными нитями с языковой картиной мира и, соответственно, осмысление каждой языковой единицы есть установление связи с лексической (и, шире, – культурноязыковой) системой в целом. В разных языках не существует «полных» лексических эквивалентов, как не существует их у носителей одного языка разных возрастов, профессий и т. д. [Norms 1970].

Слова – «черпаки смысла». Они пригодны для «зачерпывания» смыслового содержания, единого для всех, однако разные люди черпают далеко не одинаково. Так, для дальтоники *красный* и *зеленый* означают одно и то же, а для героев Дж. Оруэлла «Мир есть война». Зачерпывание глубокого и емкого содержания, проникновение в смысл – есть выход в пространство, полное индивидуальных ассоциаций, связываемых в узел психическим единством слова. Представление об атомах смысла, восходящее к Декарту, Лейбницу и авторам грамматики «Пор-Рояль», столь необходимое для построения логической семантики, в психологическом плане не более, чем дискретная составляющая континуального в своей основе сознания человека. Отнесение к слову фиксированного значения, описываемого в терминах признаков и компонентов, эквивалентно представлению о некоем усредненном и редуцированном его смысле. Такое значение, «очищенное» от спонтанных и непредвиденных ассоциаций, неизбежно оказывается бедным, жестким, неглубоким.

Представить психологически адекватный смысловой облик слова как конечную величину или как совокупность дискретных значений или признаков – то же, что представить себе трехмерный образ шара, построенного из кубиков. Какие образы вызывают в памяти такие, например, словосочетания и обороты речи: *глубокий смысл, вместить смысл, неисчерпаемый смысл; смысл, не лежащий на поверхности,*

углубиться в смысл, сокровенный смысл, уловить смысл? Не является ли ассоциативное поле слова ответом на вопрос: «Что *стоит* за словом X?» и «Какой смысл вы *вкладываете* в слово X?»

Наличие ассоциативного поля проявляется в широкоизвестных явлениях трудностей припоминания слова, состояниях, при которых искомое слово как бы находится «на кончике языка» («tip of the tongue phenomenon»), когда искомое слово замещается другим, взятым из общего ассоциативного смыслового поля.

Когда человек не может вспомнить какое-то слово, хотя оно «вертится на кончике языка», или когда опознание ранее предъявленных знаков более успешно, чем их воспроизведение, – это означает, что человек помнит несколько больше, чем может вспомнить, а значит, что-то хранится в памяти в каком-то туманном виде [Аллахвердов 2003, с. 97–98].

Два главнейших свойства ассоциативного ряда – неопределенность порядка и безграничность количества. Отличительной чертой ассоциативного эксперимента является эмпирический и процедурный характер лингвистической информации, содержащейся в его данных, возможность их квантификации, а также возможность повторения эксперимента, оценки и верификации (фальсификации) его результатов. Это выгодно отличает данный подход, близкий к естественнонаучному, освобождая его от ограничений, присущих методу интроспекции.

Главными квантитативными характеристиками ассоциативных полей слов являются частота реакций (ассоциатов) и их репертуар (ассоциативный словарь). Частота F появления данной ассоциации в групповом эксперименте отражает общеизвестность, распространенность, стереотипность, «обкатанность» (или уникальность) этой ассоциации, свидетельствующие о наличии (отсутствии) таких же связей между словом-стимулом и словом-реакцией у данной группы (или свойственные языковому сообществу в целом). Репертуар (инвентарь) L ассоциаций, полученных от группы испытуемых, свидетельствует о широте или узости ассоциативных связей слова-стимула с другими единицами языка. Качественный состав ассоциативного поля неизмерим с помощью квантитативных параметров без привлечения дополнительных логико-дедуктивных гипотез или аксиоматических построений.

В процессе ассоциирования акцентирование отдельных аспектов содержания, отдельных областей смыслового поля не исключает одновременный учет всего ассоциируемого с данным словом. В условиях речепроизводства адекватность выбранного слова замыслу говорящего и ожиданиям слушающего обеспечивается подсознательным учетом всего так или иначе соотносимого с данным словом, всего того, что, оставаясь «за кадром», не получая выхода в «окно сознания», в то же время всегда может быть актуализовано через смену «фокусного расстояния», изменение «ракурса», редупликацию, «размывание» деталей и т. п. Прежде чем стать абстрактным элементом, знаком, подчиненным логике высказывания, слово должно быть проинтерпретировано на уровне неосознанной языковой символики. Дж. К. Ципф подчеркивал

Особые возможности количественного изучения вербальных ассоциаций личности, которые более чего-либо другого характеризуют специфику индивидуальности [Zipf 1949].

Ассоциативные поля слов раскрывают психологически адекватную картину мира носителей языка. Структурообразующим компонентом значения слова является частотность вызываемых им ассоциативных реакций. График функции распределения вербальных ассоциаций слова оказывается тем зримым образом, который дает нам возможность представить семантический мир языковой личности и лингвокультурного сообщества.

На илл. 9 приводится график ранговых распределений ассоциативных полей русских существительных СМЫСЛ, РАДОСТЬ, ЛЮБОВЬ [Долинский 2012]. На илл. 10 даны графики и полные списки ассоциаций американцев и русских на слова-стимулы PREFIX [Norms 1970] и ПРЕФИКС [Долинский 2012].

В последние десятилетия исследование количественных характеристик языковых единиц и структур привлекло внимание представителей *синергетической лингвистики* (G. Altmann, R. Köhler, Р. Г. Пиотровский). Синергетический подход опирается на идеи «сил» диверсификации и унификации, коммуникативные «потребности» носителей языка и ориентирован на изучение языковых процессов и языковой саморегуляции [Пиотровский 2006; Altmann 1992; 1997; Köhler, Altmann, Piotrowski 2005]. Р. Г. Пиотровский отмечал:

Своеобразие общей методологии синергетики состоит также в том, что она пытается описать диалектику конфликта между неподвижным бытием и становлением, статичным законом и динамичной креативностью, а что касается творческих личностей, то борьбу рутинных правил и новаторства. <...> Если обратиться к системной лингвистике, использующей теоретико-множественный и логико-математический аппарат, то она ориентирована на описание статического (синхронного) строения языка, но не на его динамику (диахронию). И всё же, несмотря на все своеобразие путей развития современной науки о языке, идеи синергетики в связке с концепциями саморегуляции систем и теории катастроф стали проникать в лингвистический обиход [Пиотровский 2006, с. 8–9].

Согласно модели синергетической лингвистики, в сфере семантики слушающий старается исключить полисемию, не допуская существования у каждого слова более, чем одного значения. Подобная мера дает ему возможность получать сообщения с наименьшими усилиями (least efforts) по их декодированию. Говорящий, напротив, старается придать каждому слову максимум значений, что потребует наименьших усилий с его стороны по кодированию сообщений. Победа в коммуникации говорящего привела бы к унификации: исключению всех слов, кроме одного, передающего все значения. Победа слушающего привела бы к диверсификации: превращению всех слов в однозначные. Эмпирический результат этих противоположных тенденций – компромисс, направленный к определенному равновесному вероятностному распределению значений. По отношению к ассоциациям слова тенденция та же, что и со значениями: слушающий стремится игнорировать скрытые (неявные, имплицитные) ассоциации слова, непривычные смыслы; говорящий постоянно порождает новые ассоциации на слово, новые коннотации. Их число и частота могут различаться для каждого слова. Это означает, что семантика слова постоянно подвергается процессам диверсификации и унификации. Результатом этого будет компромисс между коммуникативными потребностями (communication needs) говорящего и слушающего (эти роли попеременно играет каждый коммуникант), компромисс, порождающий асимметричное распределение, в котором одна или несколько ассоциаций (значений) появляются очень часто, а остальные многочисленные ассоциации образуют периферию частотного распределения.

Синергетическая метафора построена на принципе дискретности. Однако в акте коммуникации можно наблюдать не только борьбу

говорящего и слушающего за «победу», но и стремление к универсальной «ранговой шкале», общему смыслу, требующему не войны «интересов» и «коммуникативных потребностей», а сотрудничества и компромисса.

Корпусный подход в лингвистике основан на анализе больших баз данных языковых массивов, хранящихся на компьютере. Во всех случаях, когда мы имеем дело с корпусными исследованиями (текстами естественного языка, данными психолингвистических экспериментов и т. п.), проявляется так называемый эффект *концентрации и рассеяния*, который состоит в том, что имеется небольшая группа очень частых единиц («ядро», или «голова» рангового распределения), и большая группа редких единиц («периферия», или «хвост», распределения).

Для аналитического выражения зависимости между частотой и рангом слова (квазигипербола), предлагалось множество формул, которые представляют собой разновидности классического «закона Ципфа»: самая простая из них ($F_i \cdot i = \text{const}$) не всегда надежно аппроксимирует эмпирические частоты, а самая изошренная – страдает избытком плохо интерпретируемых параметров и отсутствием универсальности применения. Неоднократно отмечалось, что распределение имеет слишком длинный «хвост» и, как результат, слишком медленную сходимость из-за бесконечной дисперсии, а также «патологическую» концентрацию в начале и резкую асимметрию, что позволяет сделать вывод о негауссовости данных распределений.

Известный лингвостатистик Густав Хердан писал:

«Математики верят в него [закон Ципфа], потому что лингвисты постановили считать его лингвистическим законом, а лингвисты со своей стороны верят в него потому, что математики постановили считать его математическим законом» [Herdan 1966].

Попыткам аналитического выражения данной функции посвящен ряд работ, так или иначе модифицирующих формулы Ципфа.

Для ранговых распределений ассоциатов, полученных в эксперименте, наилучшие результаты дает формула Ципфа-Долинского [Долинский 1988; 2012; Altmann 1992]:

$$F_i = F_1 \cdot i^{-(a+b \cdot \ln i)}$$

или после логарифмирования:

$$\ln F_i = \ln F_1 - (a + b \cdot \ln i) \ln i$$

где a и b – параметры.

Язык служит как раскрытию и сохранению смыслов, так и их сокрытию и утрате. Противоречивый «скрывающе-раскрывающий» статус слова по отношению к смыслам определяет количественную структуру ассоциативных полей – совмещение гомогенности и гетерогенности распределений. Эффект концентрации и рассеяния, характерный для частотной структуры текстов, имеет своим источником, по-видимому, такой механизм порождения текста, который обусловлен ассоциативной природой языкового сознания.

С одной стороны, одна или несколько особо прочных ассоциаций играют роль социальных «сторожей» и «ограничителей» смыслов; с другой – множество уникальных непрочных ассоциаций выступают в роли индивидуальных «первопроходцев» и «расшатывателей» смыслов. Фактами языка являются и «голова», и «хвост» распределения, независимо от их употребительности и отношения к «норме».

Первая тенденция *сохранения смыслов* проявляется в гомогенности распределения ассоциаций, сужении их поля, доминировании принципа наименьших усилий; она служит сохранению языка от естественных шумов, разрушающих его строй. Вторая тенденция *раскрытия смыслов* проявляется в гетерогенности распределения, расширении поля ассоциирования, доминировании принципа наибольших усилий; она служит сохранению языка от преднамеренных попыток людей выхолостить его смысл.

В сфере *социального функционирования* языки также распределены крайне неравномерно – от нескольких самых распространенных до сотен находящихся на грани вымирания [Потапов 1997]. Из 6–7 тыс. языков мира (социально признанных – от 3 до 4 тыс.) только 23 языка (от 50 млн носителей каждый) являются родными для 4,1 млрд человек (из 7,2 млрд жителей Земли). 83 % языков распространены в пределах одной страны. 90 % языков имеют менее 10 тыс носителей каждый. По прогнозам, к концу XXI в. исчезнет от 65 до 90 % языков мира.

На илл. 11 дано многомерное ранговое распределение языков по числу стран, в которых на них говорят, в сопоставлении с числом носителей, числом изучающих, числом интернет-пользователей и долей веб-сайтов.

Асимметричность рангового распределения языков приводит (в диахронии) стихийно развивающиеся языки к форме литературных языков, характеризующихся социальной регламентированностью и нормированностью, ведущей к консервации языкового состояния. Общая лингвистическая теория, в которой по-прежнему нет места для «малых» языков, до сих пор страдает крайним «англо- и европоцентризмом» [Кибрик 2011].

На илл. 12 представлено распределение литературных языков по числу лауреатов Нобелевской премии по литературе [Лауреаты 2016]. В 1901–2016 гг. премия вручалась 109 раз (в 1904, 1917, 1966, 1974 по 2 лауреата). Премия не вручалась в 1914, 1918, 1935, 1940–1943 гг. 113 Нобелевских лауреатов по литературе писали на 25 языках: Рабиндранат Тагор (Нобелевская премия по литературе 1913) писал на бенгальском и английском языке, Сэмюэль Беккет (1969) писал на французском и английском языке, Иосиф Бродский (1987) писал поэзию на русском языке и прозу на английском языке. Эти три лауреата были маркированы бенгальским, французским и русским языком, соответственно.

Альфред Нобель в своем завещании отклонял любые соображения о национальности кандидатов: должен быть выбран самый достойный, «скандинав он или нет». Проблема представленности языков всего мира стала ведущей, и в течение долгого времени шведская Академия, выбиравшая Нобелевских лауреатов, справедливо подвергалась критике за то, что она делала присуждение награды делом Европы. В 1984 г. постоянный секретарь Академии объявил, что внимание к неевропейским авторам «постепенно возрастало», и были сделаны усилия, чтобы «достичь глобального распределения».

Экспериментальное изучение ранговых распределений методами, разрабатываемыми в рамках количественной лингвистики, позволяет выявить многообразные связи качественного и количественного характера между элементами языковой системы, ее многоуровневое иерархическое устройство, пружины ее функционирования и эволюции. Многие аспекты языка, связанные с количественными параметрами распределений его единиц, их ассоциативным потенциалом, не позволяют нам теперь разделить убеждение Н. С. Трубецкого в том, что «язык лежит вне меры и числа».

Приведем ряд высказываний лингвистов и математиков.

Г. Хердан: «Стабильность относительных частот символов является общей характеристикой лингвистических форм <...>. Это выражение

того факта, что даже здесь, где человеческая воля и выбор имеют широчайшую сферу, где сознательно избираемый замысел (*intention*) и бесконечная игривость быстро сменяют друг друга, – даже здесь выявляется в большем целом регулярность главного значения» [Herdan 1966, p. 36].

Ю. К. Орлов: «Ни один биолог никогда не исследовал фарша из множества лягушек, чтобы установить “среднелягушечные” характеристики. Ни один филолог не отнесется всерьез к фразе: “Все счастливые семьи похожи ... на перекладных из Тифлиса”. А ведь это – модель репрезентативной выборки... Нужно отказаться от привычной трактовки частоты как оценки вероятности, так как в условиях недостаточных выборок эта трактовка приводит к слишком большим ошибкам» [Орлов 1980, с. 39].

В. В. Налимов: «Усредненные характеристики здесь, в отличие от физики, не имеют значения. Важны отдельные явления в своем индивидуальном проявлении вне зависимости от того, какова вероятность их появления <...>. К нашей концепции, может быть, ближе стоит невероятностный подход к построению количественной семантики расплывчатого смысла слов» [Налимов 2003, с. 96].

Ю. В. Чайковский: «Вероятностный язык здесь заведомо неприемлем, но если все-таки пользоваться языком частот, то придется признать модели, в которых дисперсии неограниченны, наиболее удобными <...>. Дальнейшее продвижение в статистической лингвистике будет, как мне видится, достигнуто после решения более общего вопроса – почему столь общ феномен квазигипербол» [Чайковский 2001, с. 207].

Исследования в квантитативной лингвистике – статистике речи и текста, словаря и ассоциативных полей, стилеметрии и лингвогеографии – позволяют по-новому оценить природу важнейших системных закономерностей, носящих существенно «недискретный», «нежесткий» характер и плохо поддающихся описанию на неадекватном языке «алгебраической» лингвистики. В то же время методы теории вероятностей и математической статистики, применяемые в лингвистике, имеют свои границы. Распределения частот в текстах и в ассоциативных полях, полисемия, стилеметрия, языковые различия и многие другие объекты изучения как в лингвистике, так и в биологии, экономике, социологии, технетике, – слишком сложны и нестабильны для построения их моделей, исходя из удаленных от эксперимента общих принципов, характерных для фундаментальной физики. Следует настороженно относиться к привлечению сложной математики, отвлеченным построениям, легковесному теоретизированию. Здесь,

по-видимому, более уместны феноменологические модели, опирающиеся на реально измеряемые характеристики, чем опора на принцип максимума статистической энтропии.

Так, в настоящее время эксперимент позволяет ручаться за относительную точность выполнения закона сохранения энергии в ядерных реакциях, равную 10–6, тогда как законы Менделя выполняются в эксперименте гораздо менее точно.

В литературе встречаются настоятельные рекомендации измерять функции распределения на том основании, что стабилизация функций $F_m(x)$ с ростом m всегда более заметна, нежели стабилизация гистограмм – эмпирических плотностей распределения. Между тем это достоинство, по мнению Ю. И. Алимова, «<подобно> преимуществу нечувствительного прибора, дающего малое рассеяние результатов измерений, перед прибором чувствительным, у которого рассеяние в тех же условиях оказывается значительно большим, как раз из-за чувствительности» [Алимов 1980, с. 24].

Следует отметить, что многие ранговые распределения не являются *вероятностными* в принятом смысле слова [Налимов 2003; Чайковский 2001]. Для ассоциативных полей слов, как и для квазитекста неограниченной длины, неопределенным остается понятие «генеральная совокупность», и дело здесь не только в бесконечной дисперсии. Однородность культурноязыкового сообщества и однородность ассоциативного поля находятся в тесной взаимосвязи и взаимозависимости. К неформальной должна быть отнесена и проблема исходных посылок при определении представительности выборок. С ростом выборки доля вновь появляющихся слов не обнаруживает заметного падения. Вспомним афоризм: статистике часто принадлежит первое слово, но последнее – никогда.

Вот что писал об этом Джордж Кингсли Ципф:

Демонстрируя наши формулы, мы не должны считать вопрос решенным статистически; наши формулы – не «статистические артефакты», в обычном понимании этого термина. Кроме того, имея в виду число объектов, к которым формулы применялись эмпирически, нельзя сказать, что соответствия – всего лишь дело случая, как обыкновенно интерпретируется слово «случай»; на самом деле, то, что мы обнаружили такое близкое соответствие нашим формулам, проявившееся даже в двух произвольных языковых выборках, *чрезвычайно противоположно вероятностям* [Zipf 1938, с. 360]. (Курсив наш. – В. Д.).

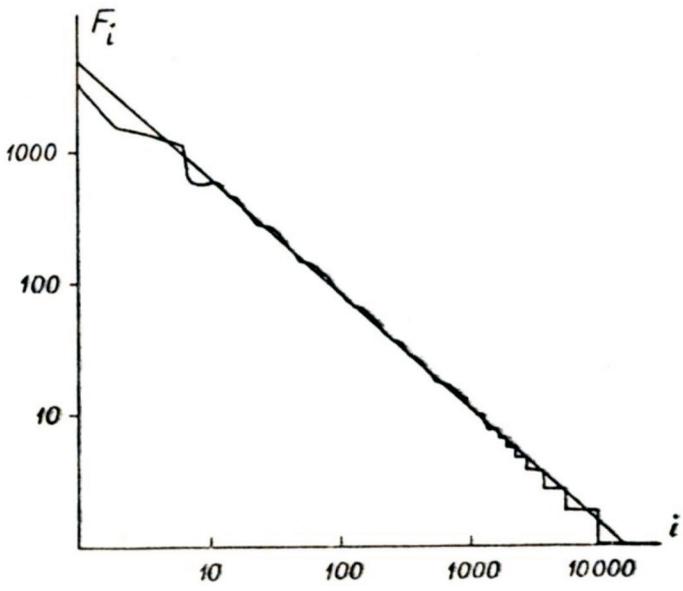
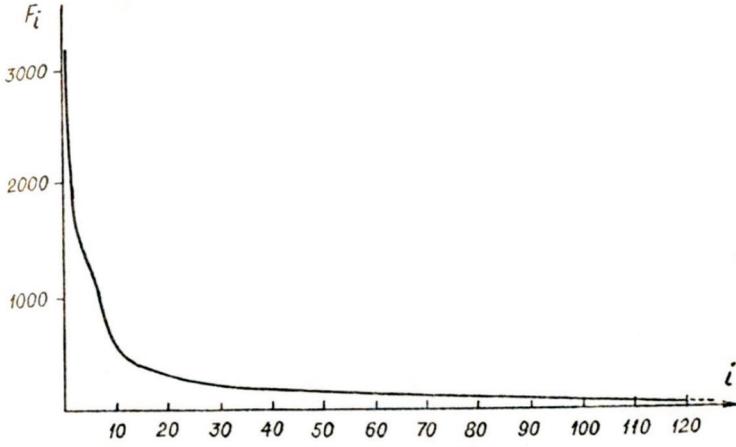
Основатель кибернетики Норберт Винер подчеркивал:

Как ни труден отбор надежных данных в физике, гораздо сложнее собрать обширную информацию социологического характера, состоящую из многочисленных серий однородных данных. <...> В этих обстоятельствах безнадежно добиваться слишком точных определений величин, вступающих в игру. Приписывать таким неопределенным по самой своей сути величинам какую-то особую точность бесполезно, и, каков бы ни был предлог, применение точных формул к этим слишком вольно определяемым величинам есть не что иное, как обман и пустая трата времени [Винер 2003, с. 220–221].

Лингвистические модели являются наиболее «узким местом» в информационных технологиях, что связано с противоречием, существующим между смысловым характером информации, выражаемой средствами естественного языка, и необходимостью формальной ее обработки с помощью компьютерных программ. Преодоление этого противоречия, или, скорее, его «смягчение», видится, в частности, в разработке языковых моделей, развиваемых в русле лингвистического ассоцианизма, синергетики и, в целом, количественной лингвистики.

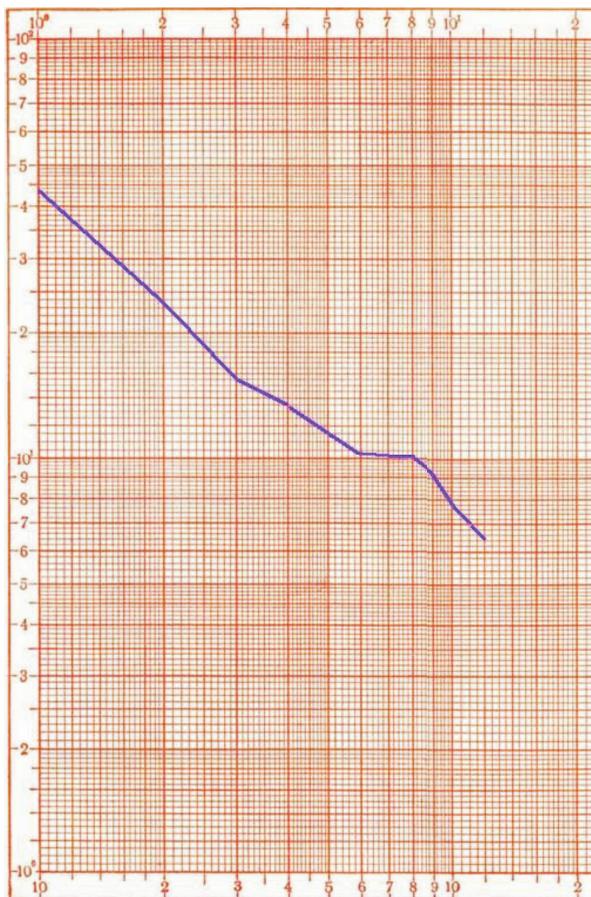
Илл. 1

Связь между частотой F_i и рангом слова i
(а – обычный масштаб; б – билогарифмическая система координат)
[Тулдава 1987]



Частотный словарь Нового Завета (Четвероевангелие)
Фрагмент частотного списка [Алексеев 2001]

Ранг	Слово	Частота	Ранг	Слово	Частота
1	<i>и</i>	4400	7	<i>я (Я)</i>	1046
2	<i>он (Он)</i>	2352	8	<i>сказать</i>	1023
3	<i>в (во)</i>	1591	9	<i>вы</i>	918
4	<i>не</i>	1383	10	<i>говорить</i>	776
5	<i>они</i>	1151	11	<i>ты (Ты)</i>	706
6	<i>быть</i>	1049	12	<i>Иисус</i>	649



Частотный словарь русского языка
Фрагмент частотного списка [Частотный 1977]

Ранг	Слово	Частота	Ранг	Слово	Частота
1	<i>в (во)</i>	42854	7	<i>что</i>	13185
2	<i>и)</i>	36266	8	<i>он</i>	13143
3	<i>не</i>	19228	9	<i>с (со)</i>	12975
4	<i>на</i>	17262	10	<i>а</i>	10719
5	<i>я</i>	13839	11	<i>как</i>	7425
6	<i>быть</i>	13307	12	<i>это</i>	6940



Илл. 4

**Ранг слова
по данным частотного словаря русского языка
в четырех подмассивах и в корпусе в целом [Частотный 1977]**

РАНГ \ ЛЕКСЕМА	в	и	не	на	я	быть
Словарь в целом	1	2	3	4	5	6
По подвыборкам:						
Художественные произведения	2	1	4	3	5	6
Драма	5	2	1	4	6	3
Научно-публицистические	2	1	4	3	6	5
Газетно-журнальные	1	2	4	3	5	6

Илл. 5

**Полисемичность слова
по данным толковых словарей русского языка:
в среднем и в четырех словарях**

ПОЛИСЕМИЯ \ ЛЕКСЕМА	в	и	не	на	я	быть
Словарь в целом	18,25	9	6,5	22,75	3	5,75
По словарям:						
ССРЛЯ	27	16	9	35	6	6
МАС	23	13	8	34	2	7
ТСРЯ	12	3	7	12	2	6
СО	11	4	2	10	2	4

- ССРЛЯ – Словарь современного русского литературного языка: в 17 т. (1950–1965); 120 тыс. слов
 МАС – Словарь русского языка: в 4-х т. (1981–1984); 90 тыс. слов
 ТСРЯ – Толковый словарь русского языка под ред. Н. Ю. Шведовой (2007); 82 тыс. слов
 СО – Словарь русского языка С. И. Ожегова (14 изд., 1983); 57 тыс. слов

**Распределение лексем по числу значений
по данным толковых и авторских словарей русского языка
[Поликарпов 1987]**

Полисемия	Число лексем			
	ССРЛЯ	МАС	СО	СЯП
Всего	120481	82017	56998	20196
1	76382	59920	44166	16494
2	26102	13236	8390	3164
3	9316	4680	2546	911
4	3989	2060	960	336
5	1928	996	426	154
6	1057	491	231	55
7	587	282	116	41
8	342	191	42	16
9	222	97	51	8
10	151	79	30	5
11	118	40	14	5
12	83	25	13	3
13	37	17	1	–
14	38	11	6	–
15	35	9	3	2
16	22	11	2	–
17	7	3	–	2
18	12	2	3	–
19	14	1	–	–
20	14	2	1	1
21	9	–	–	–
22	4	–	–	–
23	1	1	–	–
24	2	1	–	–
25	1	3	–	2
26	3	–	2	–
27	1	–	–	1
28	1	–	–	–
29	–	1	–	1
30	1	–	–	1
31	1	–	–	–
32	–	–	–	–
33	1	–	–	–

- ССРЛЯ – Словарь современного русского литерат. языка: в 17 т. (1950–1965)
 МАС – Словарь русского языка: в 4-х т. (1957–1961)
 СО – Словарь русского языка С. И. Ожегова (9 изд., 1972)
 СЯП – Словарь языка А. С. Пушкина (1956–1961)

Илл. 7

**Распределение персонажей романа М. А. Булгакова
«Мастер и Маргарита»
по частоте их упоминаний [Чайковский 2001]**

Количество персонажей	201	90	34	19	9	4	3	1	3	4	4	3	1	2	3	1	1	2	2
Количество упоминаний	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	более 19

Илл. 8

**Распределение лексем с семантикой цветообозначения
в текстах Н. В. Гоголя
(цветовой спектр, в %) [Белый 1934]**

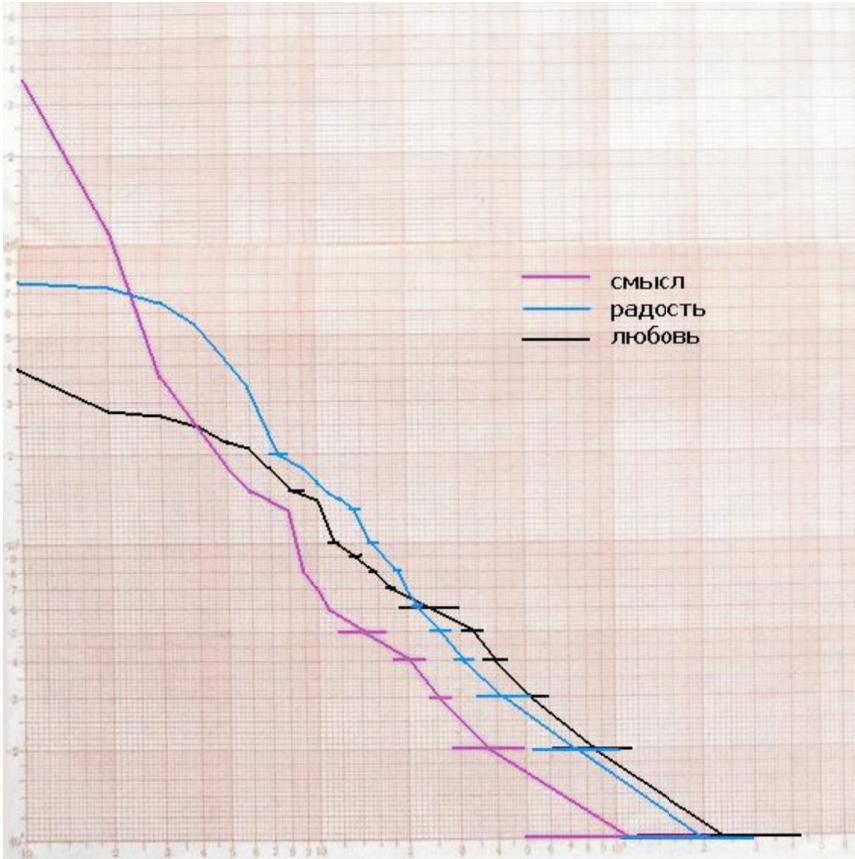
Цвет	Средняя	Группа произведений			
		1	2	3	4
Красное	17,4	26,8	12,5	10,3	6,4
Белое	14,0	9,5	9,0	22,0	17,0
Чёрное	12,0	11,0	14,1	11,8	4,8
Зелёное	9,4	8,6	7,7	9,6	21,6
Золотое	9,2	11,6	8,9	2,8	12,8
Синее	8,7	10,7	6,1	4,9	6,4
Жёлтое	7,0	3,5	8,5	10,3	12,8
Серое	5,8	2,6	8,9	10,5	6,4
Голубое	4,8	4,4	5,7	7,0	1,6
Серебряное	4,8	7,1	3,2	2,8	4,8
Коричневое	4,0	0,9	6,5	8,4	1,6
Розовое	2,3	3,8	0,8	2,1	3,2
Оранжевое	1,2	0,3	1,6	2,8	0,0
Лиловое	0,9	1,7	1,2	0,0	1,6

Обозначения групп произведений:

1. «Вечера на хуторе близ Диканьки», «Вий», «Тарас Бульба»
2. Повести и комедии
3. «Мёртвые души», первый том
4. «Мёртвые души», второй том

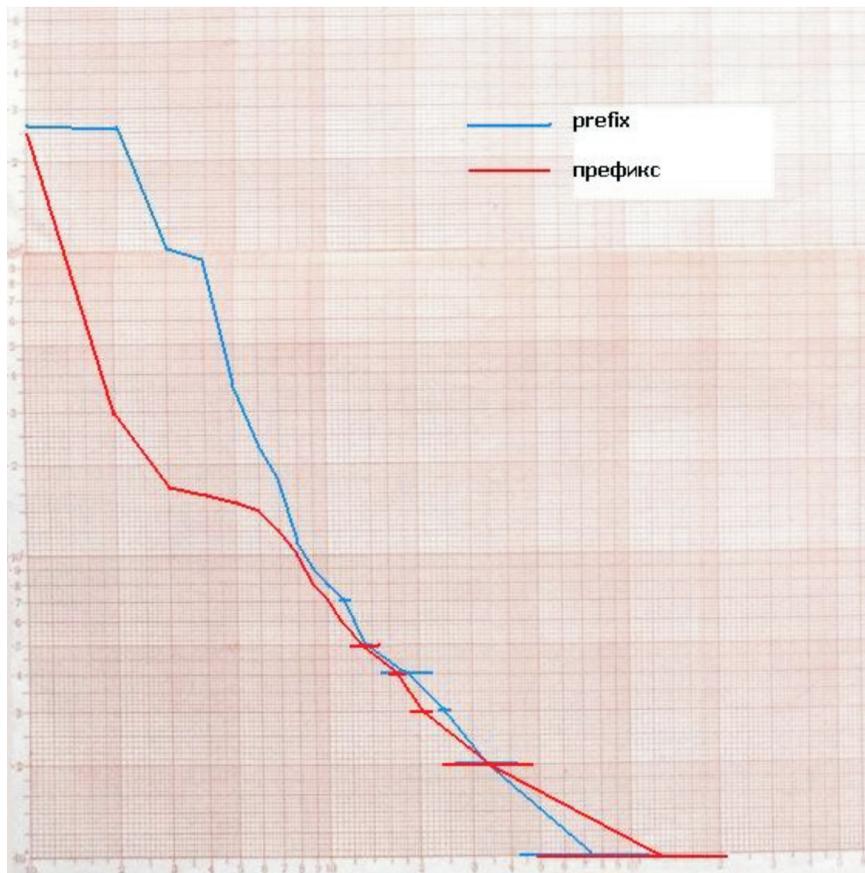
Илл. 9

Частотно-ранговые распределения ассоциативных полей
существительных русского языка: СМЫСЛ, РАДОСТЬ, ЛЮБОВЬ
(билогарифмическая система координат)



Илл. 10

**Частотно-ранговые распределения ассоциативных полей
слов-стимулов PREFIX и ПРЕФИКС**
(Языки: американский английский [Norms 1970]; русский [Долинский 2012];
билогарифмическая система координат)



PREFIX (N = 1000; F1 = 264; L = 114; m1 = 72)

suffix 264	adjective 4	after 2
before 260	end 4	ante 2
word(s) 105	index 4	ex 2
beginning 96	pre 4	in 2
first 34	preceding 4	meaning 2
English 23	re 4	name 2
book(s) 18	start 4	noun 2
grammar 11	telephone 4	part 2
begin 9	number 3	precede 2
letter (s) 8	sub 3	reflex 2
ending 7	verb 3	sentence 2
front 7	ad 2	stem 2
spelling 5	addition 2	Subject A 2
syllable (s) 5	affix 2	un 2

43–114 (F=1)

add, additive, adverb, ahead, antifix, are, beginning of, break, came before, change, comes before, coming before, complex, con, dash, de, definition, dictionary, dis, dis/appear, dog, dry, en, first, word, fixe, for, forehead, French, German, glass, head, im, in front of, ing, interesting, into, library, locate, longer, "me"tcalf, modify, mutilation, part of word, preamble, prearrange, precedence, prediction, preface, prefix, preplex, previous, pro, pronominal prefix, rain, ramrod, reading* renew, short, sink, stop, subject, surname, syllable before, telephone number, the, three, verbal, vowel, wood, working, wrench)

ПРЕФИКС (N = 1010; F1 = 246; L = 212; m1 = 164)

суффикс – 246	Аникина – 1	квадрат – 1
приставка – 29	Суздальцева – 1	кекс – 1
аффикс – 17	Фикс – 1	кино – 1
слово – 16	Фукс – 1	книга – 1
дефис – 15	анекдот – 1	код – 1
русский язык – 14	апостериори – 1	колесо – 1
грамматика – 12	апостроф – 1	континиус – 1
постфикс – 10	без ассоциаций – 1	коронка – 1
язык – 8	болезнь – 1	круто – 1
школа – 7	бугор – 1	крыша – 1
часть слова – 6	буквы – 1	ксерокс – 1
английский – 5	взлёт – 1	кс-кс-кс – 1
английский язык – 5	власть – 1	лагерь – 1
префект – 5	во – 1	ластик – 1
преферанс – 5	выделение, зелёной ручкой – 1	латынь – 1
глагол – 4	выключатель – 1	лексикология – 1
окончание – 4	выравнивание – 1	линия – 1
часть – 4	выступающий – 1	литература – 1
икс – 3	грамматический разбор – 1	логика – 1
инфикс – 3	гримаса – 1	маленький – 1
корень – 3	да – 1	маска – 1
начало – 3	деталь – 1	мафия – 1
флексия – 3	детство – 1	мистер – 1
знак – 2	дефект – 1	мм! – 1
идея-фикс – 2	дефикс (?) – 1	морфема – 1
иностраный язык – 2	длинный – 1	морф-анализ – 1
кавычки – 2	длинный нос – 1	мура – 1
компьютер – 2	добавление – 1	мэр – 1
лингвистика – 2	дом – 1	написан – 1
морфология – 2	дурацкий – 1	на- – 1
немецкий – 2	дурость – 1	недоделанный – 1
офис – 2	дыра – 1	неизвестно – 1
полиция – 2	ерунда – 1	немецкий язык – 1
сегмент – 2	загадка – 1	неповоротливый – 1
словообразование – 2	загадочный – 1	нечто – 1
сложно – 2	запятая – 1	не въехал – 1
существительное – 2	знание – 1	не знаю – 1
сфинкс – 2	значение – 1	ничего – 1
тире – 2	значок – 1	нудистика (!) – 1
урок – 2	золото – 1	облако – 1
учебник русского языка – 2	игра – 1	округ – 1
учитель – 2	издание – 1	орфография – 1
фикция – 2	инспектор – 1	остриё – 1
филология – 2	интересно – 1	отделяемый – 1
часть речи – 2	интересно узнать – 1	падеж – 1
чёрточка – 2	какаду – 1	папка – 1
экзамены – 2	карты – 1	пенис – 1
языкознание – 2		перед – 1

перепись – 1	ручей – 1	умно – 1
перо – 1	серьёзно – 1	унитаз – 1
перфект – 1	сила! – 1	уравнение – 1
печать – 1	синтагма – 1	учебник – 1
правильный – 1	синтаксис – 1	ученик – 1
предлог – 1	сифилис – 1	учить – 1
предложение – 1	скиферп (!) – 1	у- – 1
префектура – 1	скобка – 1	фактура – 1
префект и суффикс – 1	слово, напр., «при-шёл» – 1	фиксатор – 1
префикс – 1	собака – 1	фикса (!) – 1
пре- – 1	странный – 1	фикс – 1
принтер – 1	строгость – 1	фикус – 1
при- – 1	супрефикс – 1	фи-фи – 1
программирование – 1	суффикс, флексия – 1	циркумфлекс – 1
профессия – 1	телефак (!) – 1	частица – 1
раз- – 1	теория – 1	что – 1
рамка – 1	тоска – 1	чухня (!) – 1
расстрел – 1	точки – 1	широкий – 1
ребёнок – 1	тумак – 1	экзамен – 1
речь – 1	тюфикс (!) – 1	языкознание (МГУ) – 1
рога – 1	угол – 1	«2» по русскому языку – 1
розовый – 1	уголок – 1	Deutsch – 1
русский – 1	ужасный – 1	I don't know – 1

**Распределение языков
по числу стран, в которых на них говорят,
в сопоставлении с числом носителей, числом изучающих,
числом интернет-пользователей и долей веб-сайтов**

[Ethnologue-Languages of the World, CM-US, Unesco, United Nations,
University of Düsseldorf, Internetworldstats, Foundation for Endangered Languages]

Ранг	Язык	Число стран	Кол-во носителей (млн)	Число изучающих (млн)	Число интернет-пользователей по языкам (млн)	Доля веб-сайтов (%)
1	Английский	110	341 (4)	1500 (1)	948,6 (1)	54,8 (1)
2	Арабский	60	412 (2)	2,5	168,4	1,2
3	Французский	51	78 (12)	82 (2)	102,1	4,4
4	Китайский	33	1212 (1)	30 (3)	751,9 (2)	4,5
5	Испанский	31	322 (5)	14,5 (5)	277,1 (3)	4,8
6	Персидский	29	31			0,1
7	Немецкий	18	100 (10)	15 (4)	83,8	5,8 (3)
8	Русский	16	167 (8)	1	103,1	6,0 (2)
9	Малайский	13	36		109,4	0,3
10	Португальский	12	176 (7)	2	154,5	2,3
11	Итальянский	11	62 (18)	8 (6)		1,5
12	Украинский	9	47			0,1
13	Польский	9	44			1,1
14	Турецкий	8	61 (19)			1,4
15	Нидерландский	6	22			1,8
16	Тамильский	6	66 (17)			0,1
17	Урду	6	60 (20)			0,1
18	Корейский	5	78 (11)			0,3
19	Хинди	4	366 (3)			0,1
20	Бенгали	4	207 (6)			0,1
21	Яванский	3	75 (13)			
22	Вьетнамский	3	68 (16)			
23	Японский	2	125 (9)	3 (7)	115,1	5,7
24	Телугу	2	69 (14)			
25	Маратхи	1	68 (15)			
26	Панджаби	1	57			
27	Гуджарати	1	46			
28	Тайский	1	46			
...

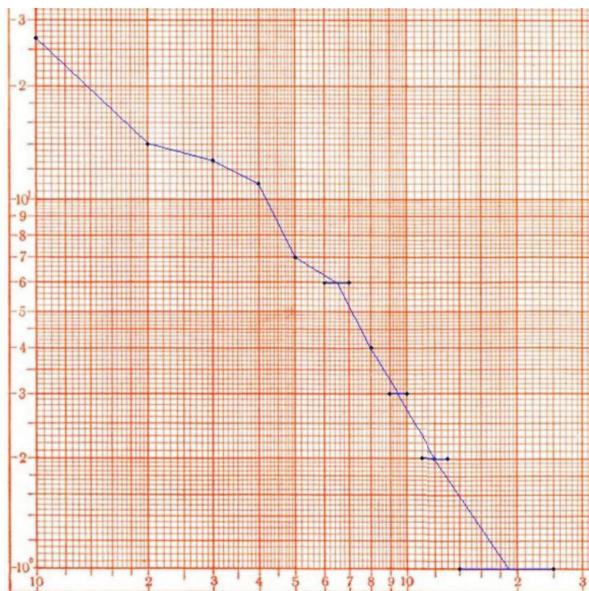
В скобках указаны ранги для распределений по каждому из столбцов таблицы

Илл. 12

**Распределение литературных языков
по числу лауреатов Нобелевской премии по литературе
[Лауреаты 2016]**

Лауреаты Нобелевской премии по литературе
(113 чел. с 1901 по 2016 гг.)
писали на следующих 25 языках:

Ранг (X)	Язык	Число лауреатов (Y)	Ранг (X)	Язык	Число лауреатов (Y)
1	Английский язык	28	14	Арабский	1
2	Французский	14	15	Бенгальский	1
3	Немецкий	13	16	Венгерский	1
4	Испанский	11	17	Иврит	1
5	Шведский	7	18	Идиш	1
6	Итальянский	6	19	Исландский	1
7	Русский	6	20	Окситанский	1
8	Польский	4	21	Португальский	1
9	Норвежский	3	22	Сербско-хорватский	1
10	Датский	3	23	Турецкий	1
11	Греческий	2	24	Финский	1
12	Китайский	2	25	Чешский	1
13	Японский	2			1



СПИСОК ЛИТЕРАТУРЫ

- Алексеев П. М.* Частотные словари. СПб. : Изд-во С.-Петерб. ун-та, 2001. 156 с.
- Алимов Ю. И.* Альтернатива методу математической статистики. М. : Знание, 1980. 64 с.
- Аллахвердов В. М.* Методологическое путешествие по океану бессознательного к таинственному острову сознания. СПб. : Речь, 2003. 368 с.
- Арапов М. В.* Квантитативная лингвистика. М. : Наука, 1988. 184 с.
- Белый Андрей.* Мастерство Гоголя. М.–Л., 1934. 320 с.
- Винер Н.* Кибернетика и общество. Творец и робот / пер. с англ. М. : Тайдекс Ко., 2003. 248 с.
- Долинский В. А.* Распределение реакций в экспериментах по вербальным ассоциациям // Уч. записки Тартуского университета. Вып. 827. Тарту, 1988. С. 89–101.
- Долинский В. А.* Теория ассоциативных полей в квантитативной лингвистике. М. : Тезаурус, 2012. 512 с.
- Кибрик А. Е.* База естественного человеческого языка и ее основные параметры: доклад на конф. «Диалог-2011». URL: www.dialog-21.ru/digests/dialog2011/
- Лауреаты Нобелевской премии по литературе. URL: www.nobelprize.org/nobel_prizes/literature/laureates/index.html/
- Ляшевская О. Н., Шаров С. А.* Частотный словарь современного русского языка. М., 2009. 1152 с. Ljashevskaja O. N., Sharov S. A. Chastotnyj slovar' sovremennogo russkogo jazyka. M., 2009. 1152 p.
- Мартыненко Г. Я.* Введение в теорию числовой гармонии текста. СПб. : Изд-во СПб. ун-та, 2009. 252 с.
- Налимов В. В.* Вероятностная модель языка. О соотносительности естественных и искусственных языков. 3-е изд. Томск–М. : Водолей Publishers, 2003. 368 с.
- Орлов Ю. К.* Невидимая гармония // Число и мысль. Вып. 3. М. : Знание, 1980. С. 70–106.
- Пиотровский Р. Г.* Лингвистическая синергетика: исходные положения, первые результаты, перспективы. СПб. : Филол. фак-т СПбГУ, 2006. 160 с.
- Поликарпов А. А.* Полисемия: системно-квантитативные аспекты // Уч. зап. Тартуского университета, вып. 774. Тарту, 1987. С. 135–154.
- Потапов В. В.* К современному состоянию проблемы вымирающих языков в некоторых регионах мира // Вопросы языкознания. 1997. № 5. С. 3–15.
- Тулдава Ю. А.* Проблемы и методы квантитативно-системного исследования лексики. Таллин : Валгус, 1987. 204 с.
- Чайковский Ю. В.* О природе случайности. М. : Центр системных исследований. Институт истории естествознания и техники РАН, 2001. 272 с.

- Частотный словарь русского языка / под ред. Л. Н. Засориной. М.: Русский язык, 1977. 936 с.
- Altmann G.* Two models for word association data // В. Rieger (Ed.), *Glottometrika* 13. Bochum: Brockmeyer, 1992. P. 105–120.
- Altmann G.* The art of Quantitative Linguistics // *Journal of Quantitative Linguistics*, 1997. Vol. 4. No 1–3. P. 13–22.
- Herdan G.* The Advanced Theory of Language as Choice and Chance. Berlin–Heidelberg–N. Y. : Springer Verlag, 1966. 459 p.
- Köhler R., Altmann G., Piotrowski R. G.* (Hrsg.): *Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch.* Berlin–New York : de Gruyter, 2005. 1041 p.
- Norms of Word Association* / Eds. L. Postman and G. Keppel. N. Y. & L., 1970. 943 p.
- Zipf G. K.* Homogeneity and heterogeneity in language // *The psychological records*. 2. 1938. P. 347–367.
- Zipf G. K.* *Human behavior and the principle of least effort.* Cambridge (Mass.): Addison-Wesley, 1949. 573 p.

УДК 81'36

В. В. Потапов

доктор филологических наук, главный научный сотрудник НИР кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики английского факультета МГЛУ; ст. научный сотрудник филологического факультета МГУ им. М. В. Ломоносова;
e-mail: volikpotapov@gmail.com

**К ВОПРОСУ О СОВРЕМЕННЫХ ПОДХОДАХ К КОНЦЕПЦИИ
«НЕИКОНИЧЕСКАЯ ХРОНОЛОГИЯ»
(по в. С. Евсееву)**

В рамках концепции В. С. Евсеева иконичность в языке рассматривается с учетом фактора соотнесенности событий в реальном времени и их передачи с помощью языковых средств. Рассматривая синтаксические и семантические факторы, необходимо и целесообразно определить специфику основных временных отношений и противопоставления иконической хронологии неиконической хронологии. Таким образом, иконичность обозначает последовательность событий, расположенных в одном темпоральном ряду с их описанием при помощи языка. В данном случае могут возникнуть следующие вопросы: при каких условиях будут появляться какие-либо отклонения от принципа временной иконичности, от каких факторов зависит это отклонение, как связаны эти факторы друг с другом и т. д. Проанализированный В. С. Евсеевым языковой материал позволил совершенно по-новому охарактеризовать синтаксические конструкции, передающие отклонение от естественной темпоральной последовательности событий в нарративном тексте. Таким образом, в статье представлена концепция В. С. Евсеева, которая связана со сравнительным описанием неиконической хронологии в немецком, английском и русском языках на материале текстов различных жанров.

Ключевые слова: неиконическая хронология; временная семантика; временная референция; таксис; синтаксические конструкции.

V. V. Potapov

D. Sc. Chief Researcher, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the Humanities and Applied Sciences of Moscow State Linguistic University (MSLU); Senior Researcher, Faculty of Philology, Lomonosov Moscow State University;
e-mail: volikpotapov@gmail.com

**TO THE QUESTION OF MODERN APPROACHES TO THE CONCEPT
OF "NON-ICONIC CHRONOLOGY" (by V. S. Yevseyev)**

As part of the concept of V. S. Yevseyev iconic chronology in the language is considered taking into account the factor of the correlation of events in real time and their transmission by language means. Considering syntactic and semantic factors, it

is necessary and expedient to determine the specifics of the main temporary relations and the opposition of the iconic chronology of the non-iconic chronology. Thus, iconic chronology denotes the sequence of events located in one temporal series with their description by language. In this case, the following questions may arise: under what conditions will there be any deviations from the principle of the iconic chronology, on what factors this deviation depends, how these factors are related to each other, etc. Analyzed by V.S. Yevseyev linguistic material allowed a completely new way to characterize syntactic constructions that convey a deviation from the natural temporal sequence of events in a narrative text. Thus, in the article the author presents the concept of V.S. Yevseyev, which is associated with a comparative description of non-iconic chronology in German, English and Russian languages on the basis of texts of a different genre.

Key words: non-iconic chronology; temporary semantics; temporary reference; taxis; syntactic constructions.

Введение

В рамках концепции В. С. Евсеева [Yevseyev 2003; Yevseyev 2008; Yevseyev 2011; Yevseyev 2012; Yevseyev 2013; Yevseyev 2017] *иконичность* в языке рассматривается с учетом фактора соотносительности событий в реальном времени и их передачи с помощью языковых средств. Таким образом, *иконичность* обозначает последовательность событий, расположенных в одном темпоральном ряду с их описанием при помощи языка. В данном случае могут возникнуть следующие вопросы: при каких условиях будут появляться какие-либо отклонения от принципа врененной *иконичности*, от каких факторов зависит это отклонение, как связаны эти факторы друг с другом и т. д.

Рассматривая синтаксические и семантические факторы, необходимо и целесообразно определить специфику основных временных отношений и противопоставления иконической хронологии неиконической хронологии. Анализ исследуемого в монографии материала осуществлялся с учетом обращения к знаковой (семиотической) модели Ч. С. Пирса [Peirce 1978; Peirce 1991] и переформулирования его подхода к данному вопросу со стороны Р. Якобсона [Jakobson 1971a; Jakobson 1971b]. Согласно Ч. С. Пирсу, *иконические знаки* определяются как знаки, согласованные с воспринимаемыми признаками детоната. Благодаря этому иконические знаки отличаются от символических и указательных знаков. В данной теории рассматриваются предложения с неиконической хронологией, когда события передаются в обратном порядке, и отклонение от естественного порядка выражено грамматическими средствами.

Специфика текстов в рамках литературно-художественной коммуникации

По мнению В. С. Евсеева, *неиконическая хронология* – это отклонение от естественной последовательности событий на каком-либо отрезке (повествовательного) текста [Евсеев 2009б, с. 115–116]. Термин *(не)иконический* используется автором в традиции Ч. С. Пирса, определявшего иконичность как внешнее сходство знака с передаваемым им предметом. При определении объекта исследования автор придерживался принципов, предложенных Петербургской школой функциональной грамматики для *таксисных явлений* [Yevseyev 2003]. «Таксис – линейная последовательность единиц, часто – временные отношения между действиями» [Баранов, Добровольский 1996, с. 583]. *Таксис* как термин впервые появился в научном обиходе в работе Р. О. Якобсона в 1957 г. “Shifters, verbal categories and the Russian verb” [Jakobson 1957]. Лингвисты, как правило, обращают внимание на две особенности таксисных форм: во-первых, таксисные формы указывают на хронологическое соотношение, и, во-вторых, таксисные формы прототипически являются синтаксически зависимыми и обозначают возрастное «действие» [Маслов 1983, с. 42].

Согласно детальному исследованию Петербургской типологической школы, посвященному описанию глагольных конструкций, таксисные последовательности в разноструктурных языках получают в этих языках различную формальную реализацию, что дает возможность дифференциации языков по *двум типам* [Типология таксисных конструкций 2009]. К первому типу относят языки, в которых преобладающим средством выражения таксисных отношений служат сложные придаточные предложения (далее – СПП), несмотря на то, что могут использоваться и другие средства. Ко второму типу относят языки, в которых преобладающим средством выражения таксисных отношений служат осложненные предложения (далее – ОП) с *деепричастными и причастными оборотами*. В языках подобного типа могут функционировать и другие средства, однако практически отсутствуют СПП «...хотя разные компоненты конструкции вносят часто неодинаковый вклад в формирование конкретного таксисного значения, в общем случае в формировании конкретного таксисного значения конструкции принимают участие все ее значимые компоненты» [Храковский 2009, с. 112].

Неиконическая хронология рассматривается В. С. Евсеевым как одна из категориальных ситуаций *таксисных отношений разновременности*. Изучаются как СПП с придаточными времени, так и союзные сложные предложения (далее – ССП), бессоюзные сложные предложения (далее – БСП), а также СФЕ, в которых между клаузами прослеживаются временные отношения, иногда сопровождающиеся условными, причинными и следственными отношениями, ср. [Yevseyev 2008].

В рамках исследования корпуса выборок из 83 русских нарративных текстов XIX–XX вв. идентифицировано 164 случая неиконической хронологии, относящихся к прошедшему времени [Евсеев 2009б, с. 115–116]. Почти половина всех случаев¹ приходится на структуру ‘ПСВ², ДСВ’ (48,8 %). Второе место по распространенности занимает структура ‘ПСВ, когда ПСВ’ (18,3 %). Третья по частоте конструкция – ‘ПНСВ, ДСВ’ (5,5 %) – структурно примыкает к первому типу, а четвертая – ‘ПНСВ, когда ПСВ’ (2,4 %) – ко второму типу. Таким образом, по мнению В. С. Евсеева, на структуры с ДСВ или союзом приходится три четверти всех выявленных случаев. Для передачи неиконической хронологии сравнительно редко используются союзы предшествования / следования, которые обычно считаются средствами, специализированными для выражения отношений разновременности (союз *когда* может выражать и одновременность). По данным корпуса, структуры типа ‘ПСВ *после того, как* ПСВ’, ‘ПСВ, *как только* ПСВ’, ‘*прежде чем* ИСВ, ПСВ’ и ‘*перед тем как* ИСВ, ПСВ’ встречаются в общей сложности лишь в 5,5 % случаев. Как пишет в своей работе В. С. Евсеев, представляют интерес также те примеры, в которых отклонение от естественной хронологии событий эксплицитно не выражено языковыми средствами и следует лишь из характера логических, причинно-следственных или условных отношений между предикациями (называемые в классической риторике *гистерон протерон*). К типу *гистерон протерон* в русском языке относятся 10,4 % случаев неиконической хронологии, формально представленные СФЕ, БСП с разными знаками препинания и ССП с союзами *а, и, но* [Евсеев 2009б, с. 116].

¹ Все примеры взяты из работы [Евсеев 2009б].

² Условные обозначения: ПСВ – форма прошедшего времени совершенного вида; ПНСВ – форма прошедшего времени несовершенного вида; ДСВ – деепричастие совершенного вида; ИСВ – инфинитив совершенного вида).

Неиконическая хронология: временная семантика и временная референция

В данном случае речь идет о физических и языковых категориях временной семантики и временной референции, которые никоим образом не «покрывают» друг друга. Данное утверждение способствует большей ясности в исследовании категоризации временных отношений, и способа передачи действий применительно к немецкому, английскому и русскому языкам. На этом принципе построено исследование, отраженное в монографии [Yevseyev 2017].

Данная монография написана на материале предыдущих работ автора [Евсеев 2009а; Yevseyev 2011; Yevseyev 2012; Yevseyev 2013 и др.], а также является итогом продолжительного научно-исследовательского проекта.

Постановка проблемы заканчивается утверждением, что на современном уровне состояния исследуемой концепции «вопрос о функциональных мотивациях темпоральной неиконичности может считаться решенным». И это обосновывается в последующих главах монографии.

Три языка, на материале которых проводилось исследование, различаются по признаку основных моделей порядка слов, что позволяет сравнить вариативность вышеуказанного признака и построить типологическую теорию неиконичности.

В работе также представлена методика подбора экспериментального материала – корпусов литературных текстов применительно к трем анализируемым языкам. Последовательно описаны этапы создания баз данных: от ручного маркирования и аннотирования материала до этапов оцифровки данных. Автор подробно останавливается на методических трудностях создания электронных корпусов данных, методике отбора иконических структур, распределения данных по принципу, например, «главное предложение перед придаточным предложением» versus «более раннее событие в придаточном предложении» и т. д.

В главе «Обобщение результатов» изложены общие наблюдения, полученные автором в процессе исследования корпусов литературных текстов, включающих около 12.750 фрагментов (из 445 текстов) и содержащих 750 примеров неиконических структур. Согласно выводам автора, некоторые писатели исследуемых текстов широко используют принцип иконичности (например, М. Булгаков, Дж. Стейнбек,

Т. Фонтане). У других авторов подобной тенденции не наблюдается (например, у Ф. М. Достоевского, Ч. Диккенса, И. В. Гёте). Многие авторские находки отражены детально и соотносятся с критериями особенностей анализируемых языков, употребления наиболее значимых синтаксических структур применительно к письменной и устной формам текстов. Заслуживает внимания тот факт, что, как установлено автором монографии, для немецкого языка в его устной форме иконичность кодируется немного экспликативнее, чем в письменной речи. В то же время, напротив, для русского и английского языков подобной тенденции не наблюдается [Yevseyev 2017, с. 261–273].

Следующая глава «Структуры, служащие для передачи неиконической хронологии» посвящена разработке проблемы неиконической хронологии и связанных с этим языковых структур. В частности, речь идет о временных формах глагола, а также о порядке их следования, комбинаторике с участием временных союзов, наречий и т. д.

В главе «Обоснования для изучения неиконической хронологии» представлены теоретическая и методологическая база исследования, интерпретация полученных данных, историко-теоретическая соотнесенность, например с концепцией «Функциональной перспективы предложения» В. Матезиуса, «Принципа актуальности» О. Есперсена, «Закона растущих составляющих (членов)» О. Бехагеля, согласно которому категория «краткости» предшествует категории «протяженности».

Применительно к неиконическим структурам автор устанавливает соотношение, равное приблизительно 2:1. При этом «краткие структуры занимают преимущественно начальную позицию».

Согласно следующему принципу *main clause first* («главное в начале») применительно к немецкому языку главное предложение в составе сложноподчиненного предложения должно предшествовать придаточному. Это утверждение, по мнению автора, правомерно для трех анализируемых языков, включающих неиконические структуры. Вместе с тем это утверждение может быть опровергнуто наличием конструкций, где «главное расположено в конце» (*anterior in sub*), согласно которому предшествующее событие тендирует к его вербальному выражению в рамках придаточного предложения и причастных конструкций. При этом несущественна позиция и взаиморасположение вышеуказанных составляющих.

Выводы

В заключение следует подчеркнуть, что проанализированный В. С. Евсеевым языковой материал позволил совершенно по-новому охарактеризовать функционирование синтаксических конструкций, передающих принцип неиконичности в темпоральной картине текстов. Таким образом, ученым предложено глубокое исследование сравнительного описания неиконической хронологии на материале немецкого, английского и русского языков (см. также об этом [Потапов 2018]), а также представлена специфика функционирования темпоральной неиконичности в текстах литературно-художественной коммуникации.

СПИСОК ЛИТЕРАТУРЫ

- Баранов А. Н., Добровольский Д. О.* (ред.): Англо-русский словарь по лингвистике и семиотике. Т. I. / А. Н. Баранов [и др.]. М.: Помовский и партнеры, 1996. 656 с.
- Евсеев В. С.* Неиконическая хронология в немецких повествовательных текстах Нового времени // Русская германистика. Ежегодник Российского союза германистов. М.: Языки славянской культуры, 2009а. Т. 5. С. 287–296.
- Евсеев В. С.* Синтаксические конструкции для выражения неиконической хронологии в русском языке: материал. Межд. науч. конф., посв. 100-летию со дня рождения В. Г. Адмони, 9–13 ноября 2009 г. СПб.: Нестор-История, 2009б. С. 115–116.
- Маслов Ю. С.* Результатив, перфект и глагольный вид // Типология результативных конструкций (результатив, статив, пассив, перфект) / отв. ред. В. П. Недалков. Ленинград: Наука, 1983. С. 41–54.
- Потапов В. В.* Евсеев В. Неиконическая хронология: временная семантика и временная референция в немецком, английском и русском языках // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Реферативный журнал. Кёльн–Веймар–Вена: Бёлау, 2017. 322 с. (Серия 6. Языкознание.)
- Храковский В. С.* (отв. ред.): Типология таксисных конструкций. М.: Знак, 2009. 912 с.
- Jakobson R.* Shifters, verbal categories and the Russian verb. Cambridge (Mass.): Harvard Univ. Press, 1957. (Русск. пер. Якобсон Р. Штифтеры, глагольные категории и русский глагол // Принципы типологического анализа языков различного строя / отв. ред. О. Г. Ревзина. М.: Наука, 1972. С. 95–113.)

-
- Jakobson R.* Zur Struktur des russischen Verbums // R. Jakobson: Selected writings. Vol. 2: Word and language. The Hague–Paris : Mouton 1971a [1931]. S. 3–15.
- Jakobson R.* Quest for the essence of language // R. Jakobson: Selected writings. Vol. 2: Word and language. The Hague–Paris : Mouton 1971b [1965]. S. 345–359.
- Peirce Ch. S.* Logic as semiotic. The theory of signs // J. Buchler (Hg.): The philosophy of Peirce. Selected writings. New York : Ams Press 1978 [1940]. S. 98–119.
- Peirce Ch. S.* Prolegomena to an apology for pragmatism // J. Hoopes (Hs.): Peirce on signs. Writings on semiotic by Ch. S. Peirce. Chapel Hill (North Carolina). L. : University of North Carolina Press, 1991 [1906]. S. 249–251.
- Yevseyev V.* Ikonizität und Taxis. Ein Beitrag zur Natürlichkeitstheorie am Beispiel des Deutschen und Russischen. Frankfurt am Main, etc.: Lang 2003. 307 S.
- Yevseyev V.* Kodierung von Taxisrelationen und narrative Eigenschaften russischer Texte des Realismus und der Moderne // R. Hodel, V. Lehmann (Hg.): Textkohärenz und Narration. Untersuchungen russischer Texte des Realismus und der Moderne. Berlin, 2008. 368 S.
- Yevseyev V.* Non-iconic chronology in English narrative texts // P. Michelucci, O. Fischer, Ch. Ljungberg (Hg.): Semblance and signification. Amsterdam–Philadelphia : Benjamins 2011. S. 191–210.
- Yevseyev V.* Non-iconic chronology in Russian narrative texts: Its types and motivations // Zeitschrift für Slawistik 2012. Band 57. № 1. S. 46–70.
- Yevseyev V.* Motivations for non-iconic chronology in English narrative texts // Linguistics 2013. Band 51. № 3. S. 555–584.
- Yevseev V.* Nicht-ikonische Chronologie: Zeitlichkeit und Zeitreferenz im Deutschen, Englischen und Russischen. Köln–Weimar–Wien : Böhlau Verlag, 2017. 322 S.

УДК 004.4, 004.9

А. А. Харламов, Б. И. Васин

А. А. Харламов, доктор технических наук, профессор кафедры прикладной и экспериментальной лингвистики факультета английского языка МГЛУ;
e-mail: kharlamov@analyst.ru

Б. И. Васин, Генеральный директор, ООО «ДИАВЕР»; e-mail: bivasin@diaver.com

ИНСТРУМЕНТАРИЙ ДЛЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ ЭКСПЕРТНОЙ ОЦЕНКИ НАУЧНОЙ ПРОДУКЦИИ С ЦЕЛЬЮ ВЫЯВЛЕНИЯ ПРИОРИТЕТНЫХ НАПРАВЛЕНИЙ И КОЛЛЕКТИВОВ

Автоматизированная информационно-аналитическая экспертная оценка научной продукции с целью выявления приоритетных научных направлений и коллективов является одним из актуальных направлений анализа текстовых данных больших объемов, и в настоящее время далеко от практических решений. Экспертная оценка научных публикаций и научной продукции сталкивается с двумя основными проблемами: (1) субъективность экспертной оценки и трудность объединения оценок разных экспертов; (2) необходимость анализа очень больших объемов неструктурированной текстовой информации. Обе эти проблемы преодолеваются применением технологии автоматического смыслового анализа текстовой информации TextAnalyst. Предлагается методика использования технологии TextAnalyst, которая позволяет решить несколько задач из области экспертной оценки научной продукции: смоделировать предметную область, представленную в текстах автора, выявить особенности представления предметной области автором, сформировать общую для всех экспертов модель предметной области и отслеживать динамику ее изменения. При этом появляется возможность обрабатывать большие объемы текстовой информации, а возможность анализа различных корпусов текстов на единой основе привносит в анализ элемент объективности.

Ключевые слова: информационно-аналитическая экспертная оценка; анализ текстовых данных больших объемов; технология автоматического смыслового анализа текстовой информации TextAnalyst.

A. A. Kharlamov

Dr. Tekhn. Sciences, Department of applied and experimental linguistics of the FAI,
Professor, fsbi MSLU, kharlamov@analyst.ru

B. I. Vasin

General Director, Diaver, Ltd., bivasin@diaver.com

TOOLS FOR INFORMATION-ANALYTICAL EXPERT EVALUATION OF SCIENTIFIC PRODUCTION FOR SIGNIFICATION OF PRIORITY SCIENTIFIC STREAMLINES AND COMMUNITIES

Automatical information-analytical expert evaluation of scientific production for signification of priority scientific streamlines and communities is one of great current interest streamlines for analysis of text big data. And now there are not practical decisions in the direction. Expert evaluation of scientific papers and scientific production has two main difficulties: (1) subjectivity of expert evaluation; and (2) necessity of large volume unstructured text information analysis. The both the difficulties overcome by using the TextAnalyst technology of automatical semantic analysis of text information. In the paper the methods of using the TextAnalyst technology are shown. They allow decide several tasks of expert evaluation of scientific production: to model a subject domain by using author texts, to signify special aspects of the author conceptualization of the subject domain, to analyze dynamics of subject domain model changes. This is allow to analyze large volumes of text information. And the single base of different text corpora analysis introduce the objectiveness element into the analysis.

Key words: Information-Analytical Expert Evaluation, Analysis of Large Volumes of Text Information, TextAnalyst Technology of Automatical Semantic Analysis of Text Information.

Введение

Автоматизированная информационно-аналитическая экспертная оценка научной продукции с целью выявления приоритетных научных направлений и коллективов является одним из актуальных направлений анализа текстовых данных больших объемов, и в настоящее время далеко от практических решений [Kharlamov et al. 2014].

Экспертная оценка научных публикаций и научной продукции сталкивается с двумя основными проблемами: (1) субъективность экспертной оценки и трудность объединения оценок разных экспертов; (2) необходимость анализа очень больших объемов неструктурированной текстовой информации.

Обе эти проблемы преодолеваются посредством применения современных технологий для автоматической смысловой обработки текстов. При этом появляется возможность обрабатывать большие объемы текстовой информации, а возможность анализа различных корпусов текстов на единой основе привносит в анализ элемент объективности.

В статье для анализа текстов рассматривается методика использования технологии автоматического смыслового анализа текстовой

информации TextAnalyst, которая позволяет решить несколько задач из области экспертной оценки научной продукции:

- 1) смоделировать предметную область, представленную в текстах автора;
- 2) выявить особенности представления предметной области автором;
- 3) сформировать общую для всех экспертов модель предметной области;
- 4) анализировать динамику изменения модели предметной области, отражающую как динамику изменения предметной области, так и динамику изменения ее модели, соответствующую представлениям эксперта.

Модель предметной области формируется в рамках упомянутой технологии, на основе корпуса описывающих ее текстов, как однородная семантическая (ассоциативная) сеть, вершины которой соответствуют ключевым понятиям предметной области. Понятия предметной области так же, как и связи между ними в этой сети, взвешиваются их рангом, характеризующим степень связности понятия с другими понятиями (чем выше степень связности, тем выше ранг) [Жоршунов & Гомзин 2012].

Модель предметной области в виде семантической сети порождает другие представления текста: тематическую структуру текста, реферат текста, тематический реферат. Тематическая структура текста (корпуса текстов) – минимальный древовидный подграф семантической сети – иерархия понятий, характеризующая тематику текста: основную тему, подтемы, подподтемы, и т. д. Аналогично тематической структуре текста можно построить тематическое дерево на любую заданную тему текста. Реферат текста – это подмножество предложений текста, суммарный вес понятий которых превышает некоторый заранее заданный порог, в порядке их появления в тексте. Тематический реферат текста – это множество предложений текста, содержащих понятия тематической структуры на заданную тему.

Сравнение состояний моделей предметной области на временной оси позволяет выявить динамику изменения как индивидуальных представлений о предметной области отдельных авторов, коллективов и экспертов, так и объективные изменения модели, характеризующие изменения предметной области.

Сравнивая тематические структуры корпусов текстов разных авторов, разных научных коллективов, разных экспертов между

собой, а также с тематической структурой усредненной модели предметной области, полученной интегрированием информации от разных авторов, научных коллективов, экспертов, можно выявить отличия в представлениях об этой предметной области разных авторов, научных коллективов, экспертов. Числовые характеристики понятий и связей модели предметной области, полученные на основе строгих алгоритмов, позволяют объективно характеризовать эти различные модели одной и той же предметной области.

Такое представление предметной области может быть основанием для формирования единой экспертной оценки, полученной как объективными методами, так и на основе субъективных мнений отдельных экспертов, суммированием этих различных оценок с их взвешиванием степенью валидности каждого конкретного экспертного мнения.

Вопрос выявления метаданных текстов (наукометрических показателей) оставлен в стороне, так как реализуется принципиально другими методами [Налимов & Мульченко 1970].

Основное назначение инструментария – снижение нагрузки на экспертов при выполнении рутинных работ, и формализация показателей экспертной деятельности с целью ранжирования экспертов по качеству экспертизы – предполагает использование механизмов автоматического и автоматизированного анализа больших массивов плохо структурированной текстовой информации.

1. Обработка текстов автоматическая, автоматизированная

В основе механизмов обработки информации – подход, базирующийся на использовании сетевого представления смысла текста [Коршунов & Гомзин 2012].

1.1. Формирование смыслового портрета предметной области

Под смысловым портретом предметной области понимается однородная (ассоциативная) семантическая сеть с ранжированными вершинами и дугами, вершинами которой являются понятия текста, а дуги соответствуют ассоциативным связям между этими понятиями в рамках смысловых единиц текста – предложений или абзацев, или, возможно, небольших текстов (новостной ленты, например). Неоднородная семантическая сеть, в противоположность однородной (ассоциативной), имеет дуги, размеченные типами отношений [Осипов 1997].

Смысловой портрет предметной области позволяет реализовать всю нижеперечисленную функциональность: формирование тезауруса текста – гипертекстовой структуры текста, формирование лексической маски текста, тематическое моделирование текста, реферирование общее, тематическое, сравнение текстов по смыслу, классификацию текстов, кластеризацию текстов.

Формирование однородной семантической сети осуществляется путем выявления попарной сочетаемости слов упомянутых фрагментов текста (выявления частоты встречаемости слов текста) после проведения первичной обработки текста – удаления слов, не несущих в тексте смысла, с целью уменьшения информационного шума.

Слова текста подвергаются процедуре лемматизации (также для увеличения устойчивости процедуры анализа): вместо слов вершины сети становятся корневыми основы слов.

Для еще большей устойчивости анализа необходимо перечисление синонимов понятий заданной предметной области. Учет синонимии – процедура, которую пользователь должен провести в заданной предметной области.

Для ранжирования вершин и связей полученная частотная сеть текста подвергается переранжированию – итеративной процедуре, в процессе которой вес вершин увеличивается за счет весов связанных с ними вершин. При этом учитывается и вес соответствующих связей. Таким образом, вершины с наибольшей связностью оказываются имеющими наивысший ранг. В качестве начальных числовых характеристик итеративной процедуры переранжирования используются частота встречаемости слов в тексте (как весовых характеристик понятий) и частота попарной встречаемости слов в тексте (как весовых характеристик связей).

1.2. Формирование тезауруса текста – гипертекстовой структуры

Тезаурус текста представляет собой гипертекстовую структуру, в которой в качестве средства навигации по тексту используется упомянутая ассоциативная сеть, а объяснительными статьями понятий тезауруса (текста) является перечень предложений текста, содержащих эти понятия.

Имеется дополнительная возможность адресоваться в исходный текст (при необходимости расширения контекста за рамки одного

предложения) из этого конкретного предложения. По этой причине гипертекстовая структура текста становится удобным цитатником с ассоциативной навигацией на основе семантической сети.

1.3. Тематическое моделирование текста

Из ассоциативной сети текста выделяется минимальный древо-видный подграф с корневой вершиной максимального веса, который представляет собой тематическую структуру текста: корневая вершина – главная тема текста, последующие по уровню вершины – подтемы, подподтемы и т. д.

Тематический анализ, реализованный традиционными методами LSA, pLSA, LDA можно использовать с помощью программных продуктов с открытым кодом.

1.4. Формирование лексической маски

Перечень слов, используемых в словаре слов-предпочтений, позволяет визуализировать из сети по п. 1.1 подсеть, которую можно считать лексической маской. Термины лексической маски как подсети семантической сети ранжированы весовыми показателями, которые характеризуют степень важности термина в описании предметной области.

Формирование лексической маски может осуществляться экспертом в полуавтоматическом режиме из перечня слов ассоциативной сети. Эксперт также в полуавтоматическом режиме может назначать ранги терминов лексической маски, отличные от усредненных представлений, полученных автоматически при анализе текстов предметной области.

Возможно оперирование этой маской с целью удаления понятий, в ней содержащихся, из основной сети, сравнения лексических масок, визуализации динамики изменения лексических масок во времени.

1.5. Реферирование общее, тематическое

Выявление предложений текста, суммарный вес понятий, содержащихся в которых, превышает некоторый заранее заданный порог, в порядке их следования в исходном тексте, представляет собой некоторый реферат текста, полученного методом извлечения выдержек.

Подобно формированию тематического дерева текста возможно формирование дерева ответа на заданный вопрос (корневой вершиной которого является заданное понятие, или группа понятий). Выявление

в тексте предложений, суммарный вес понятий, содержащихся в которых, превышает некоторый заранее заданный порог, в порядке их следования в исходном тексте за вычетом предложений, не относящихся к этому дереву ответа на запрос, представляет собой реферат текста на тему запроса (тематический реферат).

1.6. Сравнение текстов по смыслу

Вычисление степени пересечения ассоциативных сетей (вершин и дуг с их весовыми характеристиками) двух текстов позволяет выявить степень смыслового подобия текстов [Харламов 2014].

1.7. Классификация текстов

Сравнение текста (по п. 1.6) с корпусами текстов, описывающих конкретные предметные области, позволяет выявлять степень смыслового пересечения текста с упомянутыми предметными областями, то есть классифицировать тексты.

1.8. Кластеризация текстов

Разбиение ассоциативной сети некоторого корпуса текстов на подсети путем удаления слабых связей, и выявления главных тем этих подсетей, позволяет выявить степень пересечения сетей отдельных текстов с этими подсетями, что равноценно кластеризации исходного корпуса текстов на разделы, соответствующие выделенным главным темам.

1.9. Визуализация динамики изменения смыслового портрета предметной области во времени

Последовательное представление смысловых портретов предметных областей по их временным срезам (позавчера–вчера–сегодня) позволяет визуализировать динамику изменения предметной области во времени [Харламов 2017].

2. Автоматизированные механизмы формирования экспертных оценок

Так же как текст является результатом интеллектуальной деятельности автора (коллектива авторов), текст является результатом интеллектуальной деятельности эксперта как автора. Поэтому анализ текста позволяет оценить как качество работы автора, так и качество работы эксперта, оценивающего работу автора. При этом текст автора

рассматривается как текст, а текст эксперта одновременно и как текст, и как метатекст.

Для оценки качества текстов и в том, и в другом случае используются одни и те же механизмы. Но при этом, оценки, которые формируются при анализе метатекстов во втором случае, влияют на формирование оценок текстов в первом случае – оцениваемых текстов.

В качестве основного инструмента оценки используется смысловой портрет текста (предметной области). Дополнительным инструментом анализа является лексическая маска (набор лексических меток), которая контрастирует представление о предметной области, характерное для разных авторов и для разных экспертов.

2.1. Лексическая маска эксперта

Перечень лексических маркеров (в составе лексической маски), т. е. перечень ключевых понятий текста (полученный автоматически) и скорректированный вручную экспертом, соответствует представлениям эксперта о предметной области.

2.2. Точка зрения эксперта на семантику предметной области

Семантическая сеть текста, сформированная автоматически, более или менее объективно оценивает (любой) текст. Модель предметной области, представленная семантической сетью, по мнению конкретного эксперта, меняет свою структуру. Конкретный эксперт акцентирует свое внимание на конкретных понятиях предметной области, исключая другие из круга своих интересов. Поэтому семантическая сеть предметной области (или конкретного текста, относящегося к предметной области), отфильтрованная с помощью лексической маски, становится семантическим портретом текста с точки зрения данного эксперта.

2.3. Усреднение представлений о предметной области как пересечение точек зрения разных экспертов

Сравнение семантических моделей предметных областей разных экспертов дает возможность выявить объективные представления о предметной области, но и анализировать частные мнения экспертов о предметной области. Пересечение ассоциативных сетей дает усредненное представление группы экспертов о предметной области. Дополнения к пересечению – особенности представлений отдельных экспертов о предметной области.

3. Автоматизация экспертизы

Интерпретация различных точек зрения экспертов на предметную область является основой для оценки качества экспертизы.

3.1. Оценка текста автора

Сопоставление результатов сравнения смысловых моделей, принадлежащих конкретным экспертам, как с оценками, сделанными этими экспертами, так и с оценками самих экспертов, позволяет с точностью до мнения экспертов, а также сообщества экспертов оценить качество автоматического анализа с целью его последующего использования вместо экспертов.

Для примера рассмотрим оценку конкретного текста. Авторский текст подвергается автоматическому выявлению его смыслового портрета (семантической сети). Сеть подвергается фильтрации с помощью лексической маски. Оставшаяся часть сети оценивается суммированием со взвешиванием рангов терминов, которые получены при оценке авторского текста (т. е. соответствует оценке автора). Такая же оценка с помощью лексической маски (усредненной по множеству текстов предметной области) дает максимальную оценку. Тогда процент оценки авторского текста от оценки лексической маски дает оценку текста по балльной шкале: 20 %, 40 %, и т. д. – 2 балла, 3 балла и т. д.

3.2. Оценка текста научного коллектива

Формирование модели предметной области на основе анализа текстов научного коллектива позволяет сравнивать модель предметной области с моделями других коллективов. Этот анализ не отличим от анализа текста конкретного автора в сравнении с текстами других авторов, а также в сравнении с усредненной оценкой.

3.3. Оценка степени новизны научной продукции

Выявление в научной продукции информации, новой относительно усредненной за некоторый промежуток времени модели предметной области, позволяет оценивать степень новизны научной продукции.

3.4. Оценка вклада научного коллектива

Выявление разницы усредненной за некоторый промежуток времени модели предметной области и такой же модели, полученной на

основе анализа корпуса текстов научных публикаций конкретного коллектива, позволяет оценить вклад (продуктивность) научного коллектива в состояние предметной области.

3.5. Тезаурус в оценке степени новизны научной продукции

Возможно проведение качественного (в противовес количественному) сравнения текстов. Для этого можно сравнивать не смысловые портреты текстов предметных областей, а их тезаурусы (по п. 1.2). При этом можно сравнивать тезаурусы, построенные на текстах разных экспертов, а также на текстах разных научных коллективов.

4. Анализ жизненного цикла научного коллектива

Всё вышеперечисленное касалось анализа статистики предметной области (ее состояния в конкретный момент времени). Анализ динамики модели предметной области, сформированной на основе анализа научной продукции научного коллектива, позволяет оценивать динамику развития коллектива.

4.1. Качество научной деятельности научного коллектива

Наличие изменений в структуре (составе и связях) модели предметной области позволяет формировать количественную оценку предметной области. Динамика этих численных характеристик, сопоставленная с динамикой экспертных оценок результатов научной деятельности, дает возможность автоматизировать анализ продуктивности научного коллектива, а в конечном итоге – анализировать его жизненный цикл. Так, длительное неувеличение продуктивности (качества) научного коллектива может говорить о завершении его жизненного цикла.

5. Отбор экспертов

Отбор экспертов предполагает использование как механизмов автоматической оценки качества экспертизы, так и оценок, проведенных другими экспертами (как внутренними, так и внешними).

Численная оценка точки зрения эксперта, построенная на основе анализа его текстов (по п. 2.2) с использованием лексической маски эксперта (как в п. 2.1) может быть соотнесена с результатами экспертной деятельности данного эксперта экспертным сообществом.

Проведение процедуры такой оценки по текстам конкретного автора позволит принять предварительное решение при ранжировании претендентов в эксперты в процессе отбора.

Заключение

В статье изложена методика автоматизации экспертной деятельности, проводимой на основе публикаций авторов и творческих коллективов. Представлен инструментарий для проведения такой экспертной оценки, использующий технологию для автоматического смыслового анализа текстов TextAnalyst, реализующий отдельные этапы автоматической обработки больших текстовых корпусов с формированием численных характеристик отдельных составляющих этих корпусов. Соотнесение таких усредненных численных оценок с оценками, выработанными экспертами, позволяет автоматизировать (и таким образом объективизировать) процесс проведения экспертизы, что облегчает и делает более доступным проведение более качественной экспертизы, и делает возможным сравнение экспертных оценок, сделанных отдельными экспертами.

СПИСОК ЛИТЕРАТУРЫ

- Коришунов А., Гомзин А.* Тематическое моделирование текстов на естественном языке: тр. Института системного программирования РАН. Т.23. 2012. С. 215–244.
- Налимов В. В., Мутьченко З. М.* Об использовании статистических методов при управлении развитием науки // Управление, планирование и организация научных и технических исследований. Т. 3. М. : ВИНТИ, 1970. С. 327–342.
- Осинов Г. С.* Приобретение знаний интеллектуальными системами: Основы теории и технологии. М. : Наука, Физматлит, 1997. 112 с.
- Харламов А. А.* Технология автоматического смыслового анализа текстов TextAnalyst // Вестник МГЛУ. 2014. Вып. 13 (699). С. 234–244.
- Харламов А. А.* Ассоциативная память – среда для формирования пространства знаний. От биологии к приложениям. Дюссельдорф : Palmarium Academic Publishing, 2017. 109 с.
- Kharlamov A. A., Yermolenko T. V., Zhonin A. A.* Modeling of Process Dynamics by Sequence of Homogenous Semantic Networks on the Base of Text Corpus Sequence Analysis: доклад на Международной конференции SPECOM 2014, Нови Сад, Сербия, 2014. LNAI 8773 Springer P. 300–307.

УДК 81'42

М. В. Хитина

доктор филологических наук; доцент; профессор кафедры прикладной и экспериментальной лингвистики; Институт прикладной и математической лингвистики факультета английского языка МГЛУ;
e-mail: khitina49mail.ru;

ИССЛЕДОВАНИЕ ИДЕНТИФИКАЦИОННЫХ ПРИЗНАКОВ ЛЕКСИЧЕСКОГО УРОВНЯ В МОНОЛОГАХ, ДИАЛОГАХ, ПОЛИЛОГАХ

Пилотное экспериментальное исследование посвящено анализу представленности идентификационных признаков лексической группы в монологе, диалоге, полилоге (трилоге). В качестве материала использовались фонограммы из аннотированной устно-речевой базы данных (УРБД), разработанной на кафедре прикладной и экспериментальной лингвистики (научный руководитель – д-р филол. наук, проф. Р. К. Потапова). На основании литературных источников, в том числе и экспертной практики, был подготовлен список лексических идентификационных признаков, в дальнейшем используемых для анализа. Рассматриваются как сами идентификационные признаки, так и конкретные лексические единицы, относящиеся к соответствующему идентификационному признаку. Осуществлялось сравнение наборов идентификационных признаков для каждого диктора в монологе, диалоге, полилоге, а также использованных лексических единиц. Анализировались особенности речи дикторов. При исследовании монологов, диалогов, полилогов учитывался фактор темы. Данный анализ представляет интерес для лингвистов, изучающих лексические особенности говорящих, а также для проведения криминалистических экспертиз (фоноскопия).

Ключевые слова: идентификационный признак; лексика; монолог; диалог; полилог (трилог); фоноскопическая экспертиза.

M. V. Khitina

D. Sc., Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the English Language of Moscow State Linguistic University (MSLU); e-mail: khitina49mail.ru

RESEARCH OF IDENTIFICATION FEATURES ON LEXICAL LEVEL OF A MONOLOGUE, DIALOGUE AND POLYLOGUE

Pilot experimental research is devoted to the analysis of the presented identification attributes of a lexical group in a monologue, dialogue or polylogue (trilogue). Phonograms from the annotated spoken discourse database were used as the material. The database was developed at the Department of Applied and Experimental Linguistics (Academic advisor: Doctor of Philology, Professor R. K. Potapova). According to the literature sources and taking into account expert practice, a list of lexical identification features was prepared, which were used for

the analysis later. Both the identification features and specific lexical units relating to the corresponding identification feature are considered. The sets of identification attributes for each speaker and used lexical units were compared in a monologue, dialogue, polylogue. The features of the speech of the announcers were analyzed. While studying monologues, dialogues, polylogues, the topic factor was taken into account. The analysis is of interest for linguists studying lexical features of speakers, as well as for forensic examinations (phonoscopy).

Key words: identification feature; lexis; monologue; dialogue; polylogue (trilogue), phonoscopic examination.

Введение

В настоящее время в мире отмечается активное развитие информационных технологий, передача данных по различным каналам связи, требующих защиты от несанкционированного доступа. Повышается роль автоматической обработки данных, использование речевых технологий, а также практических приложений, предполагающих использование методов идентификации и верификации личности по голосу и речи, в том числе и в криминалистических исследованиях для решения специальных задач – идентификации личности по голосу и речи.

Задачей идентификации личности по голосу и речи является определение «на основе характеристик речевого сигнала и речевого потока в целом, принадлежит ли данное высказывание конкретному говорящему при условии выбора из *n*-стимулов, принадлежащих *n*-лицам» [Потапова, Потапов 2006, с. 12]. Некоторые характеристики речи, обусловленные физиолого-биологическими, психологическими и социальными факторами, являются индивидуальными и в совокупности образуют «индивидуальную систему признаков, тождественную только самой себе» [Попов 1995, с. 18]. Эти индивидуальные признаки называются идентификационными. Под идентификационным признаком понимается такое свойство речи, которое является характерным для какого-либо человека, обладающее относительной устойчивостью и отличающее его от других людей. Предметом данного пилотного исследования являются лексические признаки, характеризующие речь человека на уровне фраз и слов.

Лексический уровень языка обладает большой вариативностью. Лексика каждого человека определяется множеством факторов: происхождением, социальным статусом, возрастом, уровнем образования, профессией и т. п. Благодаря этому лексические признаки,

обнаруживаемые в речи человека, могут расцениваться как идентификационные.

По виду речевого общения выделяют монологическую, диалогическую и полилогическую речь, которые отличаются друг от друга не только количеством участников общения и коммуникативной установкой, но и языковой и структурно-композиционной организацией.

Для проведения пилотного эксперимента были использованы фонограммы, содержащие записи монологов, диалогов и полилогов (трилогов) говорящего¹.

Цель работы – исследовать, как идентификационные лексические признаки представлены в таких видах речи, как монолог, диалог и полилог, а также сравнить распределение лексических единиц в пределах каждого выделенного идентификационного признака.

Основная гипотеза: в устно-речевом дискурсе человека на лексическом уровне могут присутствовать идентификационные признаки, сохраняющиеся независимо от вида речи (монолога, диалога или полилога).

Речевую культуру человека отражают уровень освоения норм разговорной речи и умение использовать языковые средства соответственно речевой ситуации. Языковой единицей лексического уровня является лексема (слово). Лексика как словарный состав языка в русском языке рассматривается с нескольких точек зрения. Анализ ряда источников показал, что в большинстве классификаций аспекты, по которым происходит дифференциация лексики, совпадают [Фомина 2001; Розенталь и др. 2010; Крысин 2007]. Обычно выделяют пять основных категорий: происхождение слова, принадлежность к активному или пассивному запасу лексики, к определенной сфере употребления, функционально-стилевая и экспрессивная окраска слова.

Индивидуальное словоупотребление проявляется в предрасположенности к использованию определенных слов, устойчивых словосочетаний, фразеологизмов, специальной лексики и т. д. [Попов и др. 1995, с. 91–92]. Употребление устаревших слов, словосочетаний и фразеологизмов может указывать, например, на возраст человека.

¹ Материалы взяты из устно-речевой аннотированной базы данных, подготовленной на кафедре Прикладной и экспериментальной лингвистики в рамках НИР «Исследование синтаксической и структурно-смысловой организации спонтанной речи» (шифр «Описание СРФ»), науч. рук. д-р филол. наук, проф. Р. К. Потапова).

Иностранные слова чаще встречаются в речи людей образованных, изучающих языки. Об образованности свидетельствует также употребление слов, относящихся к книжному стилю, и наоборот, использование просторечных слов и выражений характерно для необразованного человека. Наличие в речи специальной лексики, жаргонизмов и диалектизмов может дать информацию о месте рождения и проживания, его социальном статусе, принадлежности к определенным группам и роде деятельности. Использование эмоционально-оценочной и экспрессивной лексики характеризует человеческий характер. При этом речь человека характеризует и то, правильно ли эта лексика употребляется, имеются ли искажения. Довольно показательной является и незнаменательная лексика.

В настоящее время в фоноскопии не существует определенного метода выявления и описания лексических признаков устной речи. Это связано с неопределенностью в описании самого лексического уровня, с постоянным развитием русского языка и др. Так, например, Е. А. Брызгунова приводит в своей работе следующий список лексических признаков: богатство активного словаря; правильность словоупотребления; грамматическая правильность речи; наличие лексики ограниченной сферы употребления; наличие разговорной и просторечной лексики; наличие и доля афункциональной лексики (слов «паразитов»); наличие и типы сбоев речи (оговорки, неверное начало слова с последующей коррекцией или без нее, повторы первого слога, заикание); наличие стилистически окрашенных лексических средств и средств ограниченной сферы употребления (просторечных, книжных, официально-деловых, диалектных, профессиональных, жаргонных); употребление слов с ослабленным стандартным лексическим значением; наличие индивидуально синонимированных слов; употребление устаревших слов и выражений; выбор конкретного синонима из синонимического ряда значений; употребление одних и тех же слов для обозначения различных явлений действительности; употребление аффективно-символической лексики: а) устойчивые словосочетания-клише; б) заместительная лексика; в) идеоматические выражения-штампы; г) экспрессивные единицы; д) нерелевантные повторы одного и того же слова; употребление лексических средств речевого контакта; предпочтительное употребление конкретных лексем и значений многозначных слов; наличие лексических актуализаторов,

используемых для усиления значения знаменательных слов, выражающих отношение говорящего к предмету речи и имеющих добавочные оттенки значений; наличие часто употребляемых оборотов, конструкций, устойчивых выражений, фразеологизмов; наличие повторов слов и высказываний, смысловых повторов [Брызгунова 2003].

В свою очередь, Н. Б. Кураченкова в пособии для экспертов считает показательными следующие идентификационные лексические признаки [Попов 1995, с. 91–92]: соответствие (несоответствие) употребления слова его значению по толковому словарю; предпочтение конкретного синонима из синонимического ряда и др.

В данном исследовании решено было проанализировать ряд лексических признаков, используемых при идентификации личности по устной речи. Среди них: употребление заимствованных слов; устаревших слов; наличие слов ограниченной сферы употребления (специальной лексики; жаргонизмов; диалектизмов); стилистически окрашенной лексики (книжного стиля; разговорного стиля; просторечной лексики); использование эмоционально-оценочной и экспрессивной лексики; фразеологизмов; незначительной лексики. Считается, что совокупность лексических признаков, свойственных говорящему, составляет индивидуальный набор, который может служить для него идентификационной характеристикой.

Экспериментальная часть

В качестве материала для проведения исследования были использованы фонограммы спонтанной речи трех дикторов – монологов, диалогов и полилогов (трилогов) на темы «Спорт», «Учеба», «Отдых». Каждый из трех участников продуцировал монолог на каждую из трех тем (всего 9 фонограмм), попарно продуцировали диалоги (также 9 фонограмм) и по три участника создавали трилоги (3 фонограммы). Длительность записей составляла: монологов (в среднем) две минуты, диалогов три минуты, полилогов пять минут. Качество записей было оценено как хорошее. Экспериментальную группу дикторов составляли трое мужчин-преподавателей (Д1, Д2 и Д3 соответственно). Перед записью фонограмм каждый диктор прошел предварительное анкетирование.

В качестве дикторов выступали преподаватели университета, среднего и старшего возраста, проживающие в Москве более 5 лет,

имеющие высшее образование и ученые степени, родной язык – русский, владеют иностранными языками, имеющие средний социальный и экономический статус.

Исследования проводились в соответствии с методами, описанными в учебно-методическом пособии «Идентификация лиц по фонограммам русской речи на автоматизированной системе «Диалект» [Попов 1995].

На предварительном этапе исследования были прослушаны все фонограммы монологов, диалогов и полилогов и установлено их словесное содержание.

Далее дикторы были оценены с помощью метода общего аудитивного анализа. Учитывались такие признаки, как общий характер аудитивного восприятия, высота и сила голоса, его тембральная окраска, особенности голосовой атаки и просодические характеристики речи. Так, речь диктора Д1 отличается от речи других участников общения по таким признакам, как общий характер аудитивного восприятия, высота голоса (низкий), тембральная окраска голоса (неоднородный, глухой, сиплый, мягкий), сила голоса (слабый), речевое дыхание (нормальное), просодические характеристики речи. Речь диктора Д2 определялась по высоте голоса (высокий по сравнению с другими двумя дикторами), тембральной окраске голоса (однородный, звонкий, мягкий), силе голоса (сильный), речевому дыханию (нормальное) и просодическим характеристикам. Характеристики речи диктора Д3 следующие: средняя высота голоса, по тембральной окраске голос неоднородный, звонкий, скрипучий, голос сильный, речевое дыхание нормальное.

Затем на основе изученного теоретического материала была составлена таблица лексических признаков, в которой фиксировались признаки, выявленные для дикторов в монологах, диалогах и полилогах (трилогах). При наличии у диктора лексического идентификационного признака в соответствующую строку вносились примеры. При многократном появлении лексем, относящихся к одному и тому же признаку, указывалось их количество.

Затем были составлены сводные таблицы по всем формам речи для каждого диктора и проведено сравнение лексических признаков в указанных видах речевого общения. На следующем этапе анализировалось употребление конкретных лексических единиц в рамках каждого идентификационного признака.

Результаты исследования

Для диктора Д1 на лексическом уровне для всех видов речи присутствуют следующие признаки: разговорная лексика (преобладает); книжная лексика; незначительная лексика; употребление заимствованных слов; использование эмоционально-оценочной и экспрессивной лексики (в большинстве случаев положительно окрашенной). Из этого можно сделать вывод, что они являются характерными для диктора Д1 и могут считаться относительно стабильными идентификационными признаками. В меньшей степени в речи диктора присутствуют фразеологизмы (5 фонограмм из 9), специальная лексика из разных областей и местоимения, просторечная лексика и жаргонизмы. При анализе речи данного диктора не было обнаружено использования устаревших слов и диалектной лексики.

В речи диктора Д2 были выявлены следующие признаки: книжная лексика (преобладает); разговорная лексика; употребление специальной лексики; наличие эмоционально-оценочной лексики; заимствования из других языков; использование незначительной лексики. Стоит отметить, что хотя признак «наличие эмоционально-оценочной лексики» выделен как идентификационный для данного диктора (он был обнаружен во всех видах речи), количество лексем, имеющих эмоционально-экспрессивную окраску, меньше чем у двух других дикторов. Специальная лексика относится к различным областям, таким как математика, экономика, лингвистика, образование, медицина, биология, физика и химия. Помимо этого были обнаружены следующие признаки: употребление фразеологизмов (3 фонограммы) и жаргонная лексика (молодежный сленг). В речи диктора Д2 отсутствуют диалектизмы, устаревшие слова и просторечная лексика.

В речи диктора Д3 стабильно обнаруживаются такие лексические признаки, как употребление заимствованных слов; специальная лексика; книжная лексика; разговорная лексика; эмоционально-оценочная и экспрессивная лексика; использование фразеологизмов; незначительная лексика. Интересным является тот факт, что диктор употребляет большое количество эмоционально-экспрессивных единиц, а также часто употребляет фразеологизмы, которые принадлежат к разговорному стилю речи. К специальной лексике, использованной диктором Д3, относятся слова из области спорта, образования, математики, биологии, лингвистики и медицины. Кроме того, в речи диктора Д3 было

обнаружено присутствие просторечной лексики (5 фонограмм) и жаргонизмов (молодежный сленг) (4 фонограммы). Такие признаки, как устаревшие слова и диалектная лексика, не обнаружены.

Из лексических признаков, выявленных в речи каждого из трех дикторов, можно сделать вывод о том, что наборы этих признаков оказались для каждого из них индивидуальными. При этом, хотя некоторые лексические признаки присутствуют в речи всех трех дикторов, их соотношение может быть различным. Например, объем использования книжной и разговорной лексики у дикторов в нашем эксперименте отличается: диктор Д1 чаще использует разговорную лексику, для диктора Д2 характерно употребление лексики книжного стиля, а у диктора Д3 объемы книжной и разговорной лексики в речи примерно одинаковы.

Полученные данные также показывают, что особенности монологической, диалогической и полилогической видов речи в некоторой степени определяют наличие лексических признаков в речи человека. Исследование подтверждает, что во время монолога люди больше используют книжную лексику, в их речи реже появляются слова-«паразиты» и фразеологизмы. Диалогическая и полилогическая виды речи сходны: в обоих случаях лексический уровень характеризуется использованием преимущественно разговорной лексики, употреблением просторечий, фразеологизмов и жаргонной лексики, большей эмоциональностью, по сравнению с монологом, увеличением количества незначительной лексики. Следует отметить, что слова-«паразиты» чаще употребляются во время диалога, что, возможно, связано с нехваткой времени на обдумывание. Из этого можно сделать вывод, что вид речи влияет на количественное проявление лексических признаков, но не определяют однозначное наличие или отсутствие лексических идентификационных признаков.

При анализе конкретных лексических единиц, представленных в отмеченном признаке, был сделан вывод о том, что в большинстве случаев лексемы либо не повторяются, либо повторение слов происходит при реализации одной темы в монологе, диалоге и полилоге. Тем не менее количественные характеристики употребления лексем могут быть использованы для определения словарного запаса личности. Например, в речи диктора Д1 значительная часть лексем идентификационного лексического признака «Эмоционально-оценочная и экспрессивная лексика» представлена такими словами, как

«хороший» (и его вариациями «хорошие», «хорошая», «хорошо») – 16 случаев употребления, «замечательный» (и его вариациями «замечательные», «замечательно») – 7 случаев и словом «прекрасный» (с вариантами «прекрасно», «прекрасная») – 9 случаев употребления. Количество повторяющихся слов данного признака и однообразность самих лексических единиц могут свидетельствовать о том, что словарный запас Д1 при обсуждении заявленных тем ограничен. Единственный признак, представители которого стабильно повторяются в речи всех трех дикторов во всех видах речи – это незначительная лексика. Так, в речи диктора Д1 повторяются вводные слова и словосочетания «так сказать» (6 раз в монологах, 18 раз в диалогах, 3 раза в полилогах) и «конечно» (3 раза в монологе, 10 раз в диалоге, 3 раза в полилоге), частицы «ну» (10 раз в монологе, 42 раза в диалоге и 25 раз в полилоге) и «вот» (3 раза в монологе, 47 раз в диалоге и 12 раз в полилоге). Диктор Д2 чаще других использует в своей речи частицы «ну» (6 раз в монологе, 22 раза в диалоге и 25 раз в полилоге) и «вот» (6 раз в монологе, 30 раз в диалоге и 22 раза в полилоге). У диктора Д3 в речи также часто появляются частицы «ну» (9 раз в монологе, 39 раз в диалоге и 27 раз в полилоге) и «вот» (2 раза в монологе, 32 раза в диалоге и 28 раз в полилоге) и вводное словосочетание «так сказать» (3 раза в монологе, 8 раз в диалоге и 2 раза в полилоге). Данные цифры могут свидетельствовать о чистоте или «засоренности» речи человека.

Исследование показало, что некоторые лексические признаки сохраняются во всех трех видах речи (в пределах одной темы), что дает нам право считать их стабильно появляющимися и, следовательно, более значимыми для идентификации личности по речи.

Для ряда идентификационных лексических признаков достаточно только их наличия (употребление фразеологизмов, использование специальной лексики, заимствований и др.). При анализе других признаков важна и количественная оценка употребления конкретных лексических единиц (она может дать информацию о частоте употребления определенной лексики, соотношении разных типов лексики, чистоте речи, наличии у человека «любимых слов» и т. д.).

Таким образом, пилотный эксперимент показал, что в устной речи человека на лексическом уровне могут присутствовать идентификационные признаки, сохраняющиеся независимо от вида речи (монолога, диалога, полилога на определенную тему). Однако

представляется, что для окончательного решения по этому вопросу требуется проведение большего числа экспериментов с анализом речи дикторов различного возраста, социального статуса, профессии, происхождения и др., при их участии в монологе, диалоге и полилоге на разные темы, и сравнение представленности лексических признаков во всех видах речи. Только после этого можно будет говорить об относительном постоянстве конкретных лексических признаков в устной речи человека.

СПИСОК ЛИТЕРАТУРЫ

- Брызгунова Е. А.* Лингвистический анализ в криминалистической экспертизе звукозаписей // Современные методы, технические и программные средства, используемые в криминалистической экспертизе звукозаписей. Методич. пособие для экспертов / под ред. А. Ш. Каганова. М. : РФЦ СЭ при Минюсте России, 2003. С. 59–83.
- Крысин Л. П.* Современный русский язык. Лексическая семантика. Лексикология. Фразеология. Лексикография: учеб. пособие для студ. филол. факультетов высш. учеб. заведений. М. : Академия, 2007. 240 с.
- Попов Н. Ф. [и др.]. Идентификация лиц по фонограммам русской речи на автоматизированной системе «Диалект»: пособие для экспертов / под ред. А. В. Фесенко. М. : Войсковая часть 34435, 1996. 102 с.
- Потапова Р. К., Потапов В. В.* Язык. Речь. Личность. М. : Языки славянской культуры, 2006. 496 с.
- Розенталь Д. Э., Голуб И. Б., Теленкова М. А.* Современный русский язык. 11-е изд. М. : Айрис-пресс, 2010. 448 с.
- Фомина М. И.* Современный русский язык. Лексикология: учебник. 4-е изд., испр. М. : Высшая школа, 2001. 415 с.

ИНФОРМАЦИОННО-КОММУНИКАЦИОННАЯ ЛИНГВОДИДАКТИКА

УДК 81-139

И. К. Ахэрн, А. И. Горожанов

И. К. Ахэрн, доцент кафедры методики обучения иностранным языкам и литературе (английский, французский и испанский) Педагогического колледжа Центра подготовки преподавателей Университета Комплутенсе г. Мадрид;
e-mail: akahern@ucm.es

А. И. Горожанов, кандидат филологических наук, доцент, доцент кафедры грамматики и истории немецкого языка факультета немецкого языка, директор Центра дистанционного обучения;
e-mail: gorozhanov@linguanet.ru

ИНФОРМАЦИОННЫЕ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ В ПЕДАГОГИЧЕСКОЙ ПРАКТИКЕ (на опыте педагогического факультета университета Комплутенсе и центра дистанционного обучения МГЛУ)

С 2003 г. Университет Комплутенсе (Мадрид) располагает виртуальным образовательным пространством (Виртуальным кампусом), одним из компонентов которого является LMS Moodle. Преподаватели высоко оценивают возможности Виртуального кампуса, однако нельзя утверждать, что большинство пользователей LMS Moodle единодушны в исключительно положительной оценке системы. В Московском государственном лингвистическом университете дистанционное обучение иностранным языкам началось в 2016 г. на базе Центра дистанционного обучения. Деятельность этого подразделения направлена на развитие краткосрочных дистанционных курсов иностранных языков различного типа. В качестве технологической платформы используется LMS Moodle 3.0+. Опыт Центра дистанционного обучения МГЛУ и опыт Университета Комплутенсе по использованию LMS Moodle для реализации дистанционных и очных образовательных программ показывает, что пристального внимания заслуживают следующие вопросы: дальнейшее формирование ИКТ-компетенции преподавателей и студентов; разработка специальных методов использования систем управления обучением в высшем образовании; углубление обмена опытом между МГЛУ и Университетом Комплутенсе в области использования систем управления обучением для реализации широкого спектра образовательных программ.

Ключевые слова: LMS Moodle; Московский государственный лингвистический университет; Университет Комплутенсе (Мадрид); дистанционное обучение; иностранные языки.

A. K. Ahern

Ph.D., Associate Professor in the Department of Language and Literature Didactics (English, French and Spanish), School of Education of the Centre for Teacher Training, Complutense University of Madrid; e-mail: akahern@ucm.es

A. I. Gorozhanov

Ph.D., Associate Professor in the Department of German Language Grammar and History at the Faculty for German Language, Director of the E-Learning Centre; e-mail: gorozhanov@linguanet.ru

**INFORMATION AND COMUNICATION TECHNOLOGIES
IN PEDAGOGICAL PRACTICE
(based on the experience of the school of education at the UCM
and the e-learning centre at the MSLU)**

Since 2003, the UCM avails of an online educational space, called the Virtual Campus, through which, among other potential uses, teaching staff develop a set of resources for their students using the LMS Moodle. Instructors highly appreciate the usefulness and value of the Virtual Campus for their teaching. They are in clear agreement that it helps them to be efficient in their course delivery. In examining the specifics of the kinds of advantages using Moodle offers, there is no clear, unanimous opinion, although opinions on potential disadvantages are also divided. Teaching foreign languages online began at the MSLU 2016 with the establishment of the E-Learning Centre. The activity of this department was aimed at rapid development of online, short-term foreign language courses of different kinds. As a learning platform, the LMS Moodle 3.0+ is being used. The experience of the E-Learning Center at the Moscow State Linguistic University and the Faculty for Education at the UCM in using the LMS Moodle for the implementation of distance and full-time educational programs shows that the following questions deserve close attention, answers to which will improve the quality of the learning process: further development of ICT-competence of teachers and students; development of special methods for using learning management systems in higher education; deepening experience exchange between the MSLU and the UCM in the field of using LMSs for the realization of a wide range of educational programs.

Key words: LMS Moodle; Moscow State Linguistic University; Complutense University of Madrid; e-learning; foreign languages.

ICT at the Faculty of Education, Complutense University of Madrid

The Complutense University School of Education offers a range of courses, of which the most popular, in terms of the number of students

enrolled, is the Bachelor's Degree in Primary School Education¹ (in Spanish, *Grado en Magisterio en Educación Primaria*). The teachers of the different under- and post-graduate degree courses at this School of Education are specialists in diverse scientific disciplines, who provide instruction that combines its focus on the disciplines themselves and on relevant pedagogical approaches, while other instructors offer courses on general pedagogy, developmental psychology, history of education, and related fields, to students who are interested in becoming professionals in the field of education. Thus, both the teaching staff and the students can be expected to hold an interest in the use of new technologies in the teaching-learning process, given their growing importance and the obvious fact that they will have an increasing impact on education in the future. Such interests can be considered universal to today's teacher education and development programmes, as reflected in work based on the framework known as Technological Pedagogical Content Knowledge, and its widespread discussion and application [Tpack ...].

Since 2003, the UCM avails of an online educational space, called the Virtual Campus (in Spanish, *Campus Virtual*), through which, among other potential uses, teaching staff develop a set of resources for their students using the LMS Moodle. Currently (early 2018) the version of the LMS in use is 2.9, although a move to version 3.2 of Moodle is to take place very soon, and will probably be an easy transition thanks to the lack of modifications to the programming core version in use. In Moodle 3.2, users at the UCM will be able to use a plug-in connecting cloud services provided thanks to a contract with Google, and thereby have a simple means for including content in Moodle spaces directly from Google Drive. On the other hand, a single, central web server is used at the UCM, which also provides advantages in terms of efficiency.

All of the UCM professors and students can gain access to the platform using the institutional e-mail and password that they are provided with from the beginning of their studies or work at the University. Teachers then have the capacity to open a Moodle space for each course that they teach, or alternatively, a "Research Seminar" space for developing other projects related to their teaching and / or research.

¹ Information about this course can be consulted at the following link: educacion.ucm.es/data/cont/docs/titulaciones/57_en.pdf

Each course instructor at the UCM, thanks to this arrangement, can use the Moodle as part of their teaching activity in the ways that they prefer or consider necessary. The use of the online teaching platform is optional, but due to the interest that is shared among teaching staff and students in providing up-to-date teaching methods, it is taken advantage of for a substantial proportion of the University courses that are currently on offer. However, as the choice of whether to use the online teaching space and the capabilities that it provides is up to each teaching staff member, the particular options of exactly how it is most frequently used are a factor that can be seen as a reflection of the pedagogical traditions and tendencies of the Complutense teachers. Thus, in the next section, some information on what kind of preferences characterise the UCM School of Education's course instructors will be offered, as a window onto this group's teaching practices, and of their perceptions on the value of this kind of new technology.

Instructors' practices and perceptions related to Moodle at the UCM School of Education

In order to obtain information about how the School of Education professors and other teaching staff (we will refer to these staff members in general as "instructors"¹ henceforth) use the resources that are available to them within the LMS Moodle, and their perceptions on these resources and on their students' usage of them, a brief questionnaire was completed in November 2017, by 21 participants belonging to the collective of instructors of courses at this School. In this section of the article, we offer data obtained from the questionnaire, beginning with information on the uses that instructors and students make of the LMS Moodle resources available to them, followed by information on the instructors' perceptions of the advantages and challenges these practices bring about.

Firstly, it is of interest to identify the age group of the participants. The majority (61.9 %) of the instructors that participated were between 40 and 50 years of age; 23.8 % were between 50 and 60 years old, and the remaining 14.3 % were between 30 and 40 years old.

¹These include lecturers and professors of various professional levels, details of which are not provided since they are not considered relevant to the purpose of the present article.

Instructor practices

Regarding the use that instructors make of the range of Moodle capabilities, the most popular way to use it is to provide students with the required readings for the course, by means of having the texts are compulsory available or linked into the course space. All of the participants confirmed that they use Moodle in this way. It is also very common practice to provide access through Moodle to complementary readings (confirmed by 86% of participants), to PowerPoint presentations shown during face-to-face class sessions (81 % of participants do so), to online videos related to the course content (76 % of participants) or other online learning resources (71 %). Finally, 5 instructors (23.8 %) confirmed that they use Moodle tests or quizzes.

There are, however, a range of potential uses of Moodle applications that were specifically asked about in the questionnaire but shown to be less prevalent in instructors' practices: only one instructor confirmed using wikis or glossaries. On the other hand, so as to complete the option of participants adding any resources that they use, in addition to those included in the questionnaire, the following resources were also mentioned, although only once for each: uploading of tasks (assignments) done by students, forums, submission of – and sharing ideas about – assignments.

As to the instructors' view on the frequency with which their students connect to their course Moodle space, the majority (71.4 %) affirm that their students enter it once a week; but 19 % of the instructors' students are said to open it almost every day. Finally, only 2 of the participating instructors claim to have students that use their course Moodle spaces less than once each week.

On the other hand, regarding the possible impact of students' usage of Moodle resources on their grades in each course, the majority of the instructors surveyed (57 %) do not acknowledge that there is such a direct impact. However, almost half (47 %) confirm that their students' participation or activity in a course Moodle space does impact the grades that they obtain. In particular, instructors mention that participation in forums is taken into account, or that performance in the final exam is strongly affected by students having taken advantage of the learning resources provided through Moodle.

Thus, based on these data, it can be highlighted that in this context, instructors' main uses of the Virtual Campus consist in posting fundamental texts, complementary readings, slides from presentations made in class and

sometimes other resources like videos, in order to complement the class sessions and provide easy access to – mainly independent – study materials. Not much use is made of the tools for collaborative work that Moodle can provide, such as forums, wikis, glossaries, nor of other kinds of activities that probably require more preparation and administration work, such as quizzes. The fact that the Virtual Campus is complementary to face-to-face class sessions most likely influences these choices, since it may be understood that these sessions offer instructors the simplest and most straightforward option for certain task types, particularly those involving group collaboration.

Instructor perceptions and views

In general terms, instructors highly appreciate the usefulness and value of the Virtual Campus for their teaching: 90 % of the survey participants agreed (28 %) or completely agreed (62 %) that it constitutes “a very useful teaching tool”. There is a very similar level of agreement with the statement that the Virtual Campus helps instructors to be more efficient and makes their teaching work easier.

Regarding the details of exactly what aspects instructors appreciate most, or the concerns that they have related to the Virtual Campus and its uses, there is, of course, a diversity of points of view, which are reflected in the following observations and assessments.

A majority of the participants are in agreement that the Virtual Campus helps the instructor to act more as a learning facilitator than as an information transmitter (11 % agree, and 52 % totally agree); an idea based on the assumption that the use of online resources favours students’ active participation in learning [Laurillard 2002; Richards 2005]. However, a few disagree slightly with this statement.

Regarding the idea that the use of the Virtual Campus contributes to the development of interpersonal competencies and teamwork in the students, opinions among the instructors are divided, although in total, 57 % disagree with this idea (34 % slightly disagree and 24 % disagree). Yet quite a few do agree with it (28 % agree or completely agree; 14 % slightly agree). The exact same proportion of the participating instructors disagree that the use of the Virtual Campus promotes peer collaboration among the students.

As to some issues which might be expected to present challenges to instructors in relation to the use of the Virtual Campus, three were brought up in the questionnaire: that of lacking or needing to acquire technical

knowledge or skills in order to use these resources adequately; a need for a large amount of time to be able to use them; that of a possible neglect of valuable learning resources, such as library or textbooks, due to the increasing use of the Virtual Campus; and the possibility that this technology might increase students' stress and the pressure that they experience.

Instructors were of diverse opinions related to the statement that they might "need training in order to be able to use the Virtual Campus efficiently"; none totally agreed with this statement, whilst about 45 % slightly agreed or agreed, and the remaining 55 % disagreed, (slightly disagreement (20 %), disagreed (20 %), or totally disagreed 15 %). In other words, a slim majority of the participants felt that they did not require (further) training to use these technological resources adequately.

Regarding the potential view that using the Virtual Campus implies spending large amounts of time on technical tasks, the participants mainly (67 %) disagreed, although only 2 participants (9.5 %) totally disagreed. This suggests that there is a general perception that time is required, but not necessarily perceived as excessive amounts of it, to deal with the technical side of teaching with these resources. And about a third of the participants do agree that it requires a lot of time to spend on technical issues.

The possible perceptions of how the Virtual Campus might be considered to have certain negative effects on students included the question of it being a potential source of additional pressure and stress felt by the students, an idea which was mainly (81 %) disagreed with. One third of the participants totally disagreed, and half of the participants disagreed or slightly disagreed. Only 4 participants agreed, (3 slightly agreed; 1 agreed). On the other hand, the participants agreed, to a greater measure, with the potentially negative effect of the use of the Virtual Campus as a reason for students to neglect other useful study and learning resources like library materials and books. Two thirds of the participants agreed with this idea, although only 20 % totally agreed. Some disagreed, in fact 14 % completely disagreed with it, however.

All in all, it is clear that the UCM instructors who participated value the convenience and capabilities that Moodle offers. Overall, they are in clear agreement that it helps them to be efficient in their course delivery. In examining the specifics of the kinds of advantages using Moodle offers, there is no clear, unanimous opinion, although opinions on potential disadvantages are also divided.

Having described the online teaching resources used in education studies at the UCM, and some data related to instructor practices and opinions of these, in the next sections we provide similar kinds of information related to the MSLU context.

Teaching online at the E-Learning Centre of the MSLU

Teaching foreign languages online began at the MSLU 2016 with the establishment of the E-Learning Centre (russ. Центр дистанционного обучения). The activity of this department was aimed at rapid development of online, short-term foreign language courses of different kinds. On 20 September of the same year, a German language course for beginners was launched, and now, after about a year and a half, over 15 courses of different type – for English, German, French and Spanish are being provided. Among them there are general courses (foreign language from beginning to intermediate levels), “aspect courses” (communicative grammar) and foreign languages for specific purposes (written translation from a foreign language into the Russian language).

All courses are being provided on a commercial basis, and the statistics indicate that the most popular among the students are: by foreign language – English; by course type – Written translation. In detail, the percentage by foreign languages is the following: English – 63 %, French – 21 %, German – 15 %, Spanish – 1 %. The distribution by courses is the following: Written translation – 49 %, General courses – 47 %, Aspect courses – 4 %.

As a learning platform, the learning management system Moodle 3.0+ (LMS Moodle 3.0+) is being used. By adding “+” to the version number we mean that the standard core of the LMS Moodle 3.0 installation has been modified to be a more effective tool for teaching foreign languages.

After having used the LMS for about a year and a half, some conclusions about the convenience and disadvantages of the system can be drawn.

In conversations, the students and tutors spoke positively about the LMS, despite some minor challenges. Basically, the difficulties arose with the students with low level of general IT-literacy. So, several times cases of using obsolete software (web browsers mostly) were recorded, which prevented some GUI (graphical user interface) widgets from being displayed. Despite the fact that the requirements for software and connection speed are prescribed in the contract and in the course’s curriculum and manual, there were some incidents of software incompatibility, which had to be corrected during the learning process.

As to the students positive appraisals, they noted the convenience of step-by-step algorithm (linear program by B. F. Skinner [Skinner 1968, p. 49]) and the fact that all the learning material is situated within the course, without the need to address to any other sources of information. Also, the availability of the tablet and mobile interface was assessed positively.

Some criticism toward the system was mistakenly expressed, because some students could not navigate through the course freely. As it turned out later, such students understood the study of the language as a simple sum of one-time execution of tasks, and with the possibility to skip as many of them as possible. Of course these cases can in no way be related to the quality of the LMS or the selected pedagogical approach.

Tutors' comments about the system were extremely positive. Highly appreciated were the monitoring subsystem and the user-friendly interface of the assignment module, which allows commenting on the student's response, also by highlighting the background color of the text.

It is interesting to note, anecdotic ally, that some students refused to take part in online lessons with the personal tutor, neglecting the most valuable type of learning activity provided in the distance courses. In our opinion, such a refusal can be explained either by the student's unwillingness to communicate, or by a banal laziness. Only in one case, working on a aspect course, a student did not want to use online lessons, but nonetheless, passed the examination successfully, which can be regarded more as an exception than a rule.

From the point of view of the system administrator, the LMS can also be evaluated positively. The administrator is responsible, in general, for the following actions (in chronological order, for one student):

- create a user account;
- enroll student and tutor to the course;
- control the first access of the student to the system;
- monitor the student's work in the first lesson;
- protocol the work of students and tutors for the preparation of financial documents (monthly);
- upload the examination assignment to the course just before the exam;
- prepare the final protocol and delete the student from the course after the examination.

Though there are plenty of foreign language courses for Russian language speakers on the educational market, our courses tend to be a

unique offer for several reasons. First, they are individual; that means that every student is being taught not in a group but by a personal tutor, who can focus on her or his needs and be as flexible as possible in finding the best way to optimize the learning process. Second, students can start learning immediately after the conclusion of the contract, in independence from the current date. Third, students need not buy any additional learning materials (including dictionaries), because they have all they need inside the course.

Conclusion

In general, we can conclude that the LMS Moodle 3 is a modern and effective learning management system that is able to perform various learning tasks, both in blended and distance learning, including teaching foreign languages.

The fact that the installation pack of the system is distributed free facilitates optimizing the costs of the learning process, and the possibility of modifying the programming core opens the prospect of implementing various methods.

At the same time, practice says that the success of learning through the LMS Moodle is directly related to the level of development of ICT-Competence of students and teachers, which, however, is now a must for any student and professional.

The experience of the E-Learning Center at the Moscow State Linguistic University and the Faculty for Education at the Complutense University of Madrid in using the LMS Moodle for the implementation of distance and full-time educational programs shows that the following questions deserve close attention, answers to which will improve the quality of the learning process:

Further development of ICT-competence of teachers and students.

Development of special methods for using learning management systems in higher education.

Deepening experience exchange between the MSLU and the Complutense University of Madrid in the field of using learning management systems for the realization of a wide range of educational programs.

REFERENCES

- Laurillard, D.* Rethinking University Teaching: a conversational framework for the effective use of learning technologies (2nd Edition). L. : Routledge, 2002. 292 p.
- Richards, C.* The design of effective ICT-supported learning activities: exemplary models, changing requirements, and new possibilities // Language Learning & Technology. 2005. Vol. 9–1. P. 6079.
- Skinner, B. F.* The Technology of Teaching. Des Moines : Meredith Corp., 1968. 271 p.
- TPACK Framework Official Website. URL: tpack.org (06.01.2018).

УДК 81.2

Н. А. Свердлова

кандидат филологических наук, доцент, начальник отдела научно-образовательной деятельности и экспертной оценки ИНЦ СО РАН, заведующий кафедрой педагогики и экспертной оценки ИНЦ СО РАН (Иркутск);
e-mail: nsverdlova@yandex.ru

ЛИНГВОПСИХОЛОГИЧЕСКИЕ ОСОБЕННОСТИ ФОРМИРОВАНИЯ ЯЗЫКОВОЙ ЛИЧНОСТИ БИЛИНГВА

В статье рассматриваются психологические и философские особенности процесса формирования языковой личности билингва. Описанные процессы основаны на постмодернистских представлениях о плюрализме человеческого сознания. Выбор билингом формы для выражения идеи обусловлен требованиями коммуникации многоязычного и многокультурного общества. Новая социальная реальность или намерение человека участвовать в межкультурном общении приводят к необходимости использования подходящих для эффективного общения языковых знаков.

Формирование двуязычного индивида происходит в ситуации взаимодействия родной и неродной культуры и языка. Психически человек готов к изменениям в его языковой картине мира. Его сознание приобретает черты множественного. На скорость процессов формирования билингва влияет внешняя среда, которая может быть преградой для успешного понимания иноязычной культуры и для овладения самим языком. Процесс развития языковой личности билингва связан с формированием особой языковой семиосферы.

Ключевые слова: билингв; сознание; постмодернизм; языковой знак; языковая личность.

N. A. Sverdlova

Ph.D in Philology, Associate Professor, the Head of the Department of Scientific and Education activity of the Federal State Budget-Funded Educational Institution of Science "Irkutsk Scientific Center of the Siberian Branch of the Russian Academy of Sciences", The Head of Department of Teachers-training and Expertise, Irkutsk (Russia)

LINGVOPSYCHOLOGICAL FEATURES OF FORMATION OF LINGUISTIC BILINGUAL IDENTITY

The paper considers the psychological and philosophical features of the process of formation of bilingual linguistic personality. Processes described are based on the Postmodernism's ideas of pluralism of a human consciousness. The bilingual's choice of the form for expressing an idea is due to the communication requirements of a multilingual and multicultural society. The new social reality or the person's intention to participate in intercultural communication leads to the need to use language signs appropriate for effective communication.

The formation of a bilingual person occurs in the situation of the interaction of a native and non-native culture and language. A person is mentally ready for changes in his linguistic picture of the world. The consciousness of a bilingual acquires the features of pluralism. The speed of the processes is influenced by external circumstances, which can be an obstacle to successful perceiving both non-native culture and the acquired language. The process of development of bilingual linguistic personality is associated with the formation of a special linguistic semiosphere.

Key words: bilingual person; consciousness; Postmodernizm; linguistic sign; linguistic personality.

Соприкасаясь с психологией, социологией, психолингвистикой, физиологией, лингвистика находится в поиске ответов на ряд вопросов, актуальных для современного полиэтнического общества, в числе которых билингвизм и билингвальная личность.

Антропоцентрическая направленность научных исследований последних десятилетий выдвигает проблему изучения языковой личности (далее – ЯЛ), национальной языковой личности, что позволяет говорить о языковых способностях говорящего / пишущего индивида, которые влияют на процессы речепорождения и речетворчества. Все большее применение на практике получает изучение и результата указанных процессов для структурирования портрета ЯЛ [Потапова 2002].

Сегодня интерес вызывает языковая личность не в своем инвариантном состоянии (будучи представителем определенного этноса, неотъемлемым свойством которого является способность использовать систему языковых знаков), а как «национально-специфический тип коммуниканта, обладающий культурно обусловленной ментальностью, картиной мира и системой ценностей, придерживающийся определенных когнитивных подходов, языковых, поведенческих и коммуникативных норм и потенциально способный к межкультурной трансформации» [Леонтович 2007].

Необходимо заметить, что наиболее четко проявляются когнитивно-коммуникативные характеристики ЯЛ в процессе межкультурного общения, которое следует признать адекватным при равноправном существовании в сознании индивида родного языка (культуры) и неродного языка (культуры).

Языковая личность, особенности и результаты ее мыслительной деятельности, достаточно объемно представлены в структуре ЯЛ, предложенной Ю. Н. Карауловым, на вербально-семантическом,

когнитивном и прагматическом уровнях [Караулов 1987]. Если взять за основу данную структуру инвариантной языковой личности и учесть аспект межкультурной коммуникации, то следует предположить, что ЯЛ билингва наиболее подвержена трансформации, культурной и языковой, которая базируется на особенностях мыслительной деятельности индивида, а ее первопричину следует искать в социальных условиях реализации личности.

Предметом рассмотрения в данной статье является философское и психологическое осмысление процессов, которые лежат в основе формирования языковой личности билингва, поэтому отправным моментом проведенного исследования является сам человек, использующий практику попеременного пользования более, чем одним языком.

Объектом нашего внимания является не мышление билингва, как более статичная категория, а мыслительная работа индивида, владеющего родным языком (Я1) и языком, приобретенным в естественной среде или искусственным способом (Я2), т. е. «опосредованный и обобщенный способ отражения действительности», «форма познавательной деятельности человека», результатом которого являются идея, мысль, представленные в его суждениях и умозаключениях [Краткий психологический словарь 1998]. Нас интересует мыслительная деятельность как активный процесс, познавательный и эмоционально-волевой – этап, предшествующий речепорождению.

Процесс познания человеком действительности характеризуется аналитическими и синтетическими операциями, т. е. своеобразным членением мира. И ключевой категорией здесь является сознание. Это и есть «высоко аналитически расчлененное познание человеком объективного мира и своей собственной психики ... знание о котором люди сообщают друг другу, что именно необходимо для успешного поведения и деятельности как отдельного человека, так и коллективов в природной и социальной среде» [Чуприкова 2009].

Именно язык, являясь знаковой системой, помогает передать сведения о мире и получить информацию о предметах. Выборочная вербализация проходит благодаря конкретному запросу индивида – участника коммуникации. Вступая в контакт с иной средой, он становится субъектом процесса фрагментации мира и разрушения некой его упорядоченности.

Сознание билингва расширяется за счет появления «новых» понятий из Я2 (см. рис.1). Изучаемый или приобретаемый естественным путем язык на определенном этапе его усвоения характеризуется

ситуацией отсутствия единства означаемого. Можно сказать, что происходит обновление сознания или процесс, который ведет к его плюрализму («плюрализм сознания» – термин З. Баумана).

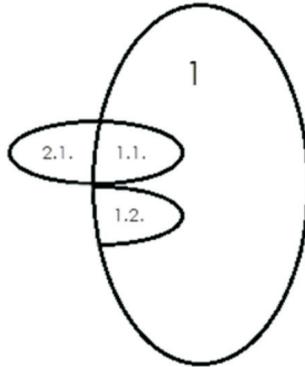


Рис. 1. 1. Сознание билингва. 1.1. Означаемое знака Я1.
1.2. Означаемое знака Я2. 2.1. Означающее (экспонент) знака Я1

Появляясь на определенном этапе развития языковых способностей билингва, означаемое Я2 со временем закрепляется, расширяется. Поиск форм для его выражения в Я2 и дальнейшее их активное использование вытесняет из сознания «неактивные» экспоненты Я1, что может привести к медленному стиранию означающего в системе Я1 (см. рис. 2.).

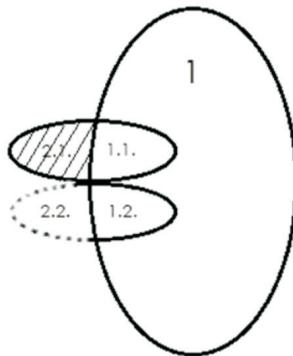


Рис. 2. 1. Сознание билингва. 1.1. Означаемое знака Я1.
1.2. Означаемое знака Я2. 2.1. Означающее (экспонент) знака Я1.
2.2. Означающее (экспонент) знака Я2

Сознание таким образом приспособливается к обновлению средств коммуникации, т. е. в данном случае мы можем говорить о его имманентности. Мотивация в использовании экспоненты для некоего содержания зависит от конкретной социокультурной ситуации.

Разрыв экспоненты знака Я1 с его планом содержания со временем может привести к аттриции Я1. Утрачивается характерное для означаемого развертывание во времени по причине неиспользования Я1. Так нарушается один из главных принципов знака – его линейность. Но это не означает, что Я1 исчезает полностью одновременно с выбором билингвом тех знаков Я2, которые необходимы говорящему/пишущему «здесь и сейчас». Сознание билингва «схватывает» (термин Ж. Деррида) предмет или представление о нем и осуществляет выбор означаемого, пользуясь иной языковой системой.

Очевидно, что определенная прагматическая установка индивида вызывает качественный и количественный рост понятий, что в конечном итоге связано с возможностью выбора вариантов экспоненты, за которой стоит некоторое содержание – «старое» (из Я1) или «новое» (из Я2) понятие.

Использование билингвами языковой системы направлено на интерпретацию окружающего, на поиск наилучших знаков для эффективной коммуникации, или же, как отметил социолог З. Бауман, на то, «какую идентичность выбрать и как суметь вовремя сделать другой выбор, если ранее выбранная идентичность потеряет ценность или лишится ее соблазнительных черт» [Бауман 2002].

Так билингв, обладая идентичностью от рождения, выбирает новую идентичность в зависимости от ряда социальных условий. Этот выбор обусловлен процессом адаптации в сложном современном мире поликультурности и многоязычия. Безусловно, это некое отступление от некогда закреплённых в сознании норм.

Категория выбора рассматривается в рамках теорий личностного выбора. К ним относятся экзистенциально-гуманистические теории выбора (И. Ялом, Дж. Бьюдженталь, Р. Мэй). В них отрицается существование рационального универсального выбора. Основанием в процессе выбора являются субъективные ценностно-смысловые образования. Среди теорий данного направления можно выделить культурно-историческую (Л. С. Выготский), личностно-мотивационные теории (А. Маслоу), ценностно-смысловой подход (С. Мадди).

В соответствии с экзистенциально-гуманистическими теориями, субъект – это человек, который самостоятельно, руководствуясь

своими потребностями, и под влиянием внешней среды выбирает языковые формы, с помощью которых он будет обмениваться своими мыслями. Дальнейшая активность человека будет сложным целым, определенным его доминирующими потребностями, что имеет отношение к психике индивида. Рефлексия, как способность сознания выделять свое «Я», относится к индивидуальному сознанию.

Метапсихологическую категорию (к таким относим и «общение», «чувство», «деятельность», «Я». – Прим. авт.) сознание отличает ответственность, свобода, осознанность. Именно проходя через свое «Я», носитель Я1 может сделать выбор той системы для общения, которая будет отвечать его потребностям. Только будучи включенным в поле сознания, понимая свою идентичность, индивид становится хозяином своего выбора. «Если Я хочу переживать свою жизнь во всей полноте, – пишет Дж. Бьюдженталь в книге «Наука быть живым», – я должен переживать ее в центре – мне необходимо чувствовать свое “Я”. Вот что такое внутреннее осознание. Это переживание своего Я» [Бьюдженталь 1998].

Как и во всем живом, в человеке заложена способность реагировать на наличные факторы среды с опережением – по принципу опережающего отражения [Чуприкова 2009], что обусловлено работой последовательностей во внешней среде. Это утверждение имеет значение для рассмотрения способности к адаптации и ассимиляции человека, оказавшегося в условиях выбора и необходимости вступать в коммуникацию в чуждой для него языковой и культурной среде, т. е. психически такой человек заблаговременно готов к изменениям внешней среды, и этот процесс естественный для него.

Указанный процесс не замыкается на ощущениях отдельного индивида: находясь в поиске благоприятных для себя условий, человек ищет пути «обмена состояниями своей психики». Так мы можем говорить о готовности подсистем речевой и невербальной коммуникации.

Такова идеальная картина формирования психологической готовности индивида и его сознания к выбору. Но внешняя среда по ряду причин может оказаться сильнее, чем сама личность: границы активности и переживания определяются не ей самой, а другими людьми, обстоятельствами, ситуацией и тогда иные (социальные) признаки – половые, статусные и другие – ограничивают «Я» личности, что приводит к уменьшению степени активности и свободы.

Для того чтобы понять психику – образы, мысли, чувства, потребности, состояния – в масштабе взаимодействия с внешним миром, необходима рефлексия. В рамках философской рефлексии современности билингвизм как социальное явление может быть рассмотрен в контексте варианта такой рефлексии – постмодернистских теорий.

Идеи постмодернизма связаны с процессом поликультурации общества, отсутствием инвариантного набора ценностей и паттернов поведения. «Определенное состояние ментальности» (определение З. Баумана) – характеристика социальных изменений, которые происходят постоянно в условиях становления культурно многополярного и многоязычного общества, осложненного вариативностью коммуникаций. Билингв становится объектом процесса реформативного сознания в сторону усложнения языкового сознания, что лежит в основе формирования особой языковой семиосферы [Свердлова 2017].

СПИСОК ЛИТЕРАТУРЫ

- Бауман З.* Индивидуализированное общество. М. : Логос, 2002. 185 с.
- Бьюдженталь Дж.* Наука быть живым: диалоги между терапевтом и пациентами в гуманистической терапии. М. : Класс, 1998. 336 с.
- Караулов Ю. Н.* Русский язык и языковая личность. М. : Наука, 1987. 257 с.
- Краткий психологический словарь / Л. А. Карпенко, А. В. Петровский, М. Г. Ярошевский. Ростов-н/Дону : Феникс, 1998. URL: psychology.academic.ru/1254/мышление
- Леонтович О. А.* Введение в межкультурную коммуникацию. М. : Гнозис, 2007. С. 9–20.
- Потапова Р. К.* Язык, речь, личность. М. : Языки славянской культуры, 2006. 496 с.
- Свердлова Н. А.* Онтологический статус языковой ноосферы билингва: материалы Международной конференции «Жизнь языка в культуре и социуме-6», Москва, 26–27 мая 2017 г. М. : Канцлер, 2017. С. 63–65.
- Чуприкова Н. И.* Функциональная система психики. Филогенетические истоки и сознание человека [Электронный ресурс] // Психологические исследования: электрон. науч. журн. 2009. № 2 (4). URL: psystudy.ru

Сетевое электронное научное издание

ВЕСТНИК Московского государственного лингвистического университета. Гуманитарные науки Выпуск 6 (797)	VESTNIK of Moscow State Linguistic University. Humanitarian Sciences Issue 6 (797)
---	--

Ответственные за выпуск 6 (797):

д-р филол. наук В. В. Потапов (*ответственный редактор*);
д-р филол. наук В. А. Долинский (*зам. ответственного редактора*)
д-р филол. наук, проф. А. В. Зубов (Беларусь);
д-р филологии, проф. Э. Шток (Германия);
А. В. Джунковский (*ответственный секретарь*)

Редактор Е. М. Евдокимова
Компьютерная верстка Г. П. Лопатиной
Дизайн обложки А. Г. Проскуракова

ФГБОУ ВО МГЛУ

Подписано в печать 27.07.2018 г.
Усл. печ. л. 13. Формат 60х90/16
Заказ № 85

Адрес редакции:

119034, Москва, ул. Остоженка, 38, стр. 1
Тел.: (499) 245 33 23
E-mail: ipk-mglu@rambler.ru

В «Вестнике Московского государственного лингвистического университета. Гуманитарные науки» публикуются основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук по следующим отраслям науки и / или группам специальностей научных работников:

10.02.00 – Языкознание
10.01.00 – Литературоведение
24.00.00 – Культурология
09.00.00 – Философские науки

«Вестник Московского государственного лингвистического университета. Гуманитарные науки» является преемником сборников научных трудов с 1940 года и, начиная с 2002 года, следующих выпусков «Вестника Московского государственного лингвистического университета»: «Лингвистика», «Языкознание», «Филологические науки», «Языкознание и литературоведение», «Восточные языки», «Философия», «Философия и культурология».

© ФГБОУ ВО МГЛУ, 2018

Издание зарегистрировано 10 июня 2016 г. Эл № ФС77-66051 Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор)
Доменное имя сайта: libranet.linguanet.ru/prk/vestnik.asp?vest_lang=rus&vest_type=gum
Учредитель: ФГБОУ ВО МГЛУ
Ссылка на издание обязательна