

# Компьютерная лингвистика в школе как интеграция математических и филологических знаний

*Международный Институт Менеджмента ЛИНК*

ФАКУЛЬТЕТ лингвистики

К.п.н., доцент Матвеева Н.В.

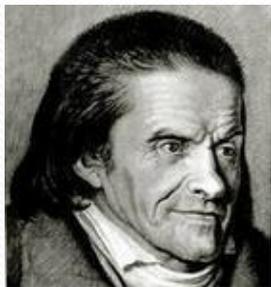


Москва-Жуковский, ноябрь 2017

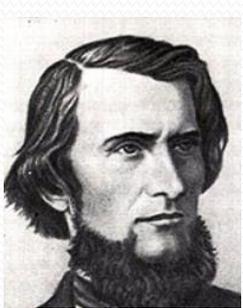
# Об интеграции знаний



О межпредметной интеграции, как фундаментальном принципе образования, писали в своих трудах еще Я.А. Коменский, И.Г. Песталоцци, К.Д. Ушинский и другие.



Однако, структура школьных знаний и существующие формы обучения, до сих пор формирует «лоскутное» мировоззрение.



Вопросы интеграции своей актуальности не теряют .

# Роль математики в науках



«Использование науками  
отвлеченных понятий и методов  
**математики** расширяет их  
возможности, способствует  
открытию новых, более глубоких  
закономерностей».

*А.Т.Хроленко*

# Лингвистика – пионер использования математики в гуманитарных науках

«Науки естественного цикла – и информатика в их числе – давно заговорили на языке математики.

Гуманитарные науки обратились к математике только в XX веке.

И первой из них была лингвистика

# Межпредметная интеграция – основа целостной картины мира

Интеграция знаний, как правило, происходит лишь индивидуально в сознании школьника.

Как показано в ряде исследований (Д.В.Гайворонский, Л.Г.Кузнецова и др.), интеграция школьных дисциплин возможна при соблюдении ряда условий:

- 1) содержательная совместимость наук, соответствующих интегрируемому учебным предметам;
- 2) наличие общих формируемых понятий;
- 3) наличие общих методов и подходов.

# Математика и информатика

**В математике** мы идем от формирования понятия числа, знака и переменной к количественным отношениям и пространственным формам действительного мира в виде математических моделей

(Мордкович А.Г. и др.).

# Математика и информатика

**В информатике** – от формирования представлений об информационном процессе как феномене реальности к информационной модели как инструменту познания этого феномена и его применения в различных областях человеческой деятельности (Бешенков С.А., Ракитина Е.А. Семакин И.Г. Хеннер Е.К. и др.).

# Математика и информатика

Для математики и информатики условия интеграции соблюдаются.

Это позволило включить **информатику** в образовательную область «Математика».

# О филологии в школе

Филология в школе включает в себя изучение языков и литературы.

Литература представлена в основном как искусство слова.

Родной русский язык, как правило, изучается лишь с точки зрения его структуры, грамматики и синтаксиса.

# Об иностранном в школе

Иностранный язык можно рассматривать как предмет, который вносит вклад в общую системную картину мира.

Тексты, используемые в учебниках иностранного языка интересны и познавательны, взяты из разных областей – истории, биологии, географии, литературы и т.д., что придает этому предмету развивающий межпредметный характер.

# Возможна ли интеграция математики, информатики и филологии?

**ДА, безусловно ВОЗМОЖНА!**

Во-первых:

Понятия «язык», «слово», «знак» и «модель» являются общими понятиями для информатики, математики и филологии – то есть содержание этих предметов взаимно пересекается и дополняет друг друга.

# Возможна ли интеграция математики, информатики и филологии?

Во-вторых:

Связь математики, информатики и филологии находит свое отражение в способе мыслительной деятельности – и математика, и информатика, и филология учат моделированию явлений окружающей действительности.

То есть в методологической основе всех трех дисциплин лежат одни и те же методы: **формализация и моделирование явлений окружающей действительности.**

# Язык как универсальное средство моделирования

Язык, есть универсальное средство моделирования, а слово естественного языка есть его смысловая единица и форма представления информации в сознании человека.

# Как можно осуществить интеграцию филологии с математикой и информатикой?

Возможны различные варианты, но самый простой, на наш взгляд: ввести в образовательную область филология новый предмет «Компьютерную лингвистику», которой, как информатике уже более 30 лет!

**Компьютерная лингвистика похожа на информатику – их история происхождения и содержание существенно пересекается!**

# Чем похожи «Компьютерная лингвистика» и «Информатика»

Первое – моментом рождения.

Компьютерная лингвистика начала активно развиваться, когда фирма IBM выпустила в **1981** году свой первый персональный компьютер.

Информатика начала активно развиваться, когда в **1985** году началось экспериментальное обучение информатике в школе (тогда появилось в русском языке слово «информатика»).

# Все надо записывать в цифрах



Уже в X веке ученый и философ эпохи возрождения Николай Кузанский в своем трактате «Об ученом познании» утверждал, что все познания о природе необходимо записывать в цифрах, а все опыты над нею производить с весами в руках.

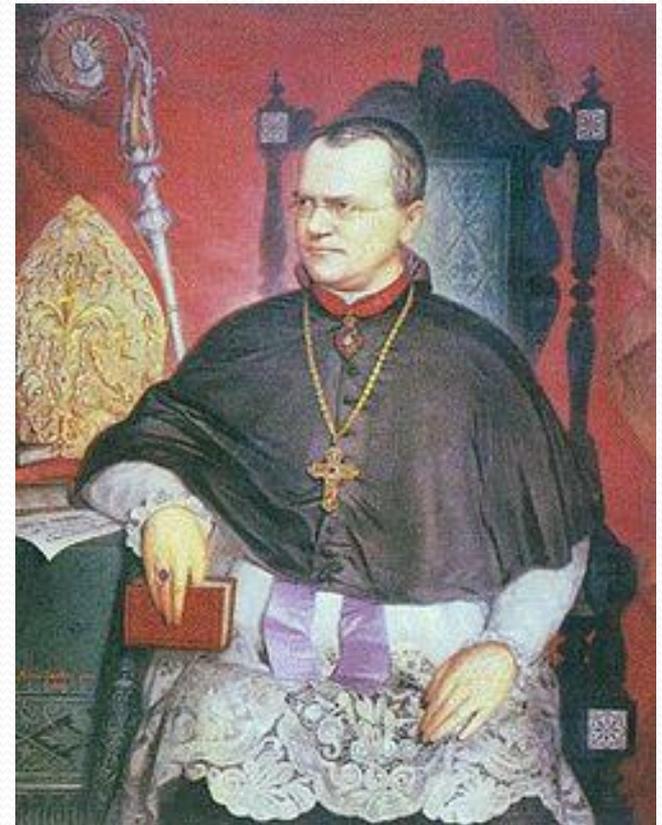
# Математику нужно применять в лингвистических исследованиях



В начале XIX века, известный русский математик Виктор Яковлевич Буняковский говорил о необходимости применения математики в лингвистических исследованиях.

# Статистические методы в генетике

- Тогда же основоположник генетики австриец Грегор Иоганн Мендель пробовал применить статистические методы исследования не только в биологии, и метеорологии, но и лингвистике.



# Тесная связь языкознания и математики



В начале XX века российский и польский языковед И.А.Бодуэн де Куртенэ указывал на тесную связь языкознания и математики.

Он говорил о том, что применение количественных методов в языкознании приблизит его к точным наукам.

# Марков АА – марковские цепи и лингвистика



Тогда же, в начале XX века, известный математик Андрей Андреевич Марков, ученик П.Л.Чебышева, применяя методы теории вероятностей и математической статистики пришел к открытию «марковских цепей»ю

Он исследовал доли гласных и согласных в тексте А.С.Пушкина «Евгений Онегин».

# Норберт Винер – сын лингвиста Лео Винера



- Математик и основоположник кибернетики Норберт Винер – сын известного американского лингвиста Лео Винера, который перевел с русского на английский 24-томное собрание сочинений Л. Н. Толстого – не без оснований считал, что в науке лингвистике есть все условия, необходимые для математического исследования.

# С чего все началось?

В ходе Нюрнбергского процесса в связи с необходимостью синхронного перевода был задуман машинный перевод. Это было в 1944-45 годах.

Личный переводчик Эйзенхауэра и его личный знакомый – руководитель одного из отделов фирмы IBM – затеяли так называемый «Джорджтаунский эксперимент»: перевод с русского на английский с помощью ЭВМ.



Была создана программа. Словарь содержал 250 слов. При переводе использовалось 6 правил. Данные хранились на перфокартах и вводились латиницей.

# Начало машинного перевода в СССР и России

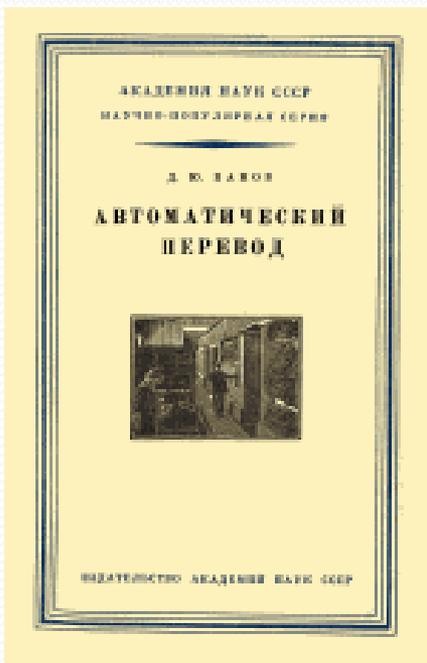
В конце **1955** г. были проведены первые опыты перевода научно-технического текста с английского языка на русский при помощи электронной счетной машины (ЭВМ) БЭСМ Академии наук СССР ...

# Начало машинного перевода в СССР и России

В **1958** году вышла работа Панова Дмитрия Юрьевича «Автоматический перевод».

В книге рассказывается о проблеме автоматического перевода с одного языка на другой.

Занимался этой проблемой Институт точной механики и вычислительной техники и Институт научной информации Академии наук СССР.



# А.П.Ершов и компьютерная лингвистика



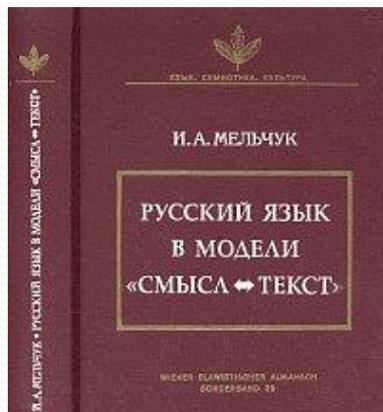
Во второй половине XX века, основоположник школьной информатики А.П.Ершов, один из пионеров теоретического и системного программирования, создатель Сибирской школы информатики, становится одним из основателей российской корпусной и компьютерной лингвистики.

# Лингвист Мельчук И.А.



В 60-тые годы А.П. Ершов уже работает над проблемой **общения человека с ЭВМ на естественном языке.**

К решению этой проблемы он привлекает лингвиста Мельчука Игоря Александровича (р. 1932) – в настоящее время канадского лингвиста российского происхождения, создателем лингвистической теории «Смысл — Текст».



# Ершов Андрей Петрович - лингвист, математик и информатик в одном лице



Ершов А.П. — один из пионеров  
русской **корпусной**  
**лингвистики**.

По его инициативе начал  
создаваться Машинный фонд  
русского языка при Институте  
русского языка АН СССР – начало  
цифrofизации в России!

## 50-тые и 60-тые годы

Перед лингвистами **ЖИЗНЬЮ** была поставлена задача **автоматического анализа текста**, такая как **машинный перевод и распознавание речи**. И они решили эту задачу!

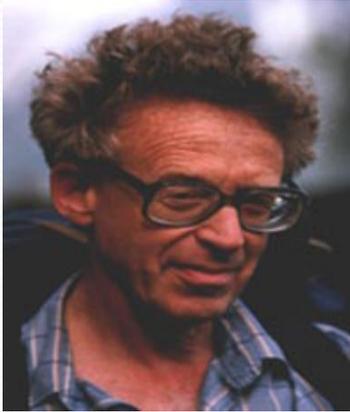
А в каком состоянии были в это время средства автоматического анализа , то есть ЭВМ – электронные вычислительные машины?

## 50-тые и 60-тые годы

Первые громоздкие ЭВМ умеют только работать с числовыми данными: хранить их в громоздкой ламповой «памяти» и обрабатывать в соответствии с заданным алгоритмом с помощью программ, написанных на «понятном» ЭВМ «языке» - искусственном языке программирования.

Еще не появились мощные персональные компьютеры – они появятся только в **1981** году!

# Лингвистика и информатика давно в тесном содружестве



Мельчук окончил испанское отделение филологического факультета МГУ.

В 1956 работал в Институте языкознания АН СССР, где занимался проблемой машинного перевода и к началу 70-х годов он является лидером в области **структурной прикладной лингвистики**.

В 1974 начинает работу над интегральной моделью языка «Смысл — Текст», которая в то время значительно опережала аналогичную теорию западного лингвиста **Н. Хомского**

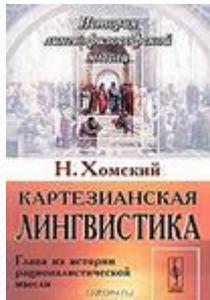
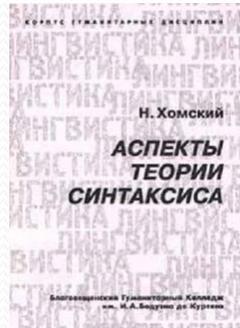
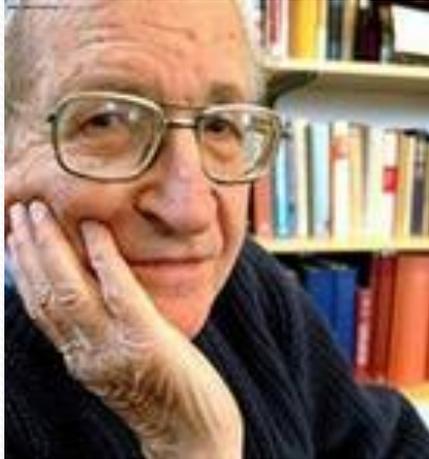
В настоящее время Мельчук профессор университета в Монреале.

А. И. КОРДАДСКАЯ, И. А. МЕЛЬЧУК

СМЫСЛ  
И СОЧЕТАЕМОСТЬ  
В СЛОВАРЕ

И. А. МЕЛЬЧУК  
КУРС  
ОБЩЕЙ  
МОРФОЛОГИИ  
Том IV

# Ноам Хомский



Ноам Хомский – лингвист, основатель когнитивной психологией и уже в 60-ые годы принимал участие в конференциях по искусственному интеллекту, где обсуждались вопросы **моделирования** процессов порождения и восприятия речи.

Сейчас проблема искусственного интеллекта – это один из важнейших разделов науки информатики.

# Ершов и его последователи

В **1985** году Ершов А.П. пишет первый учебник по информатике для общеобразовательной школы.

В качестве эксперимента преподает этот предмет в школе и внедряет ее в учебный план 10-11 классов всей страны.



Сергей Александрович Бешенков – тогда молодой специалист в области математики и информатики, становится соавтором этих учебников.

Сейчас он руководитель отдела по дидактике обучения информатике (РАО).



# Лингвистика, информатика, компьютер и информационные технологии

Есть мнение (А.Т.Хроленко, А.В.Денисов), что  
лингвистика и информатика **ПОШЛИ**  
навстречу друг другу, когда:

# Лингвистика и информатика пошли навстречу друг другу, когда

- компьютеры стали персональными (1981),
- начали интенсивно развиваться текстовые процессоры;
- появились средства проверки орфографии и грамматики;
- появились программы оптического распознавания текстов и речи;
- и так далее.

# Лингвистика и информатика всегда были рядом и вместе,

а не когда информатика достигла чего-то!

Лингвистика и информатика изначально, уже  
больше 50 лет идут рука об руку, взаимно  
поддерживая и «толкая» вперед друг друга!



# Интеграция в жизни

Появился интернет – стала актуальной задача информационного поиска в сети, содержащей огромный массив текстовых и иного рода данных.

Интернет и современные мобильные коммуникационные устройства сформировали всеобщую **потребность в языковых информационных технологиях.**

# Язык человека и язык компьютера

Формальный язык, которым «пользуется» компьютер для обработки данных, принципиально отличается от естественного разговорного языка и между ними нет ничего общего.

Компьютер в принципе не способен «понять» ни текст, ни человеческую речь. Он может лишь имитировать понимание, как бы адекватно реагируя на вопрос.

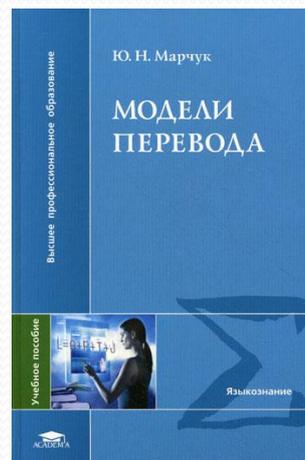
# Возможности компьютера как инструмента для лингвиста

Пользователь, например, может произвольно задавать размер листа, размеры полей и отступов, выбирать самый разнообразный шрифт, выделять, переставлять и убирать части текста, править и вносить изменения, автоматически проверять орфографию и пунктуацию, обращаться к различным словарям (которые находятся в памяти компьютера), вставлять иллюстрации и т.д. и т.п. Он может вызвать сразу несколько документов и работать с ними одновременно, перенося данные из одного в другой. Наряду с редактором есть множество прикладных программ, ориентированных на узких специалистов.

# Марчук Юрий Николаевич



Профессор МГУ, доктор филологических наук, специалист по прикладной и **компьютерной лингвистике**, машинному переводу, автоматическому анализу и синтезу текстов, терминологии и терминоведению, лексикологии и лексикографии, общему языкознанию.



Где она – компьютерная  
лингвистика (КЛ)?



# Что такое компьютерная лингвистика (КЛ)?

Компьютерная лингвистика – одна из наиболее актуальных современных лингвистических дисциплин.

**«По мере того, как расширяется информатизация современного общества, возрастает значение компьютерной лингвистики, науки, находящейся на стыке глубоко человеческой, гуманитарной науки лингвистики (языкознания), изучающей законы развития и пользования могучим средством мышления и коммуникации – языком, – и компьютерного знания, с помощью которого машине передастся все большая часть интеллектуального труда человека»**

**(Марчук Ю.Н.)**

# Задачи компьютерной лингвистики

Компьютерная лингвистика решает задачи **автоматического анализа текста**, такие как:

- ✓ машинный перевод,
  - ✓ распознавание речи,
  - ✓ информационный поиск стали
- и др.

# Задачи компьютерной лингвистики

Эти задачи стали актуальными с момента появления персонального компьютера, но только технологическая революция, связавшая нашу жизнь с интернетом и многочисленными мобильными коммуникационными устройствами, сформировала **всеобщую потребность в языковых технологиях.**

## «Компьютерная лингвистика»

О новом месте компьютерной лингвистики в жизни общества говорит то, состоялся в 2012-2013 году уже третий цикл лекций «Компьютерная лингвистика».

Лекции проходили в политехническом музее и были организованы Институтом лингвистики РГГУ совместно с Политехническим музеем при информационной поддержке сайта [Полит.ру](http://polit.ru).

Публичные лекции: сезон 2012/2013.

## «Компьютерная лингвистика»

О задачах компьютерной лингвистики рассказали те, кто уже достиг серьёзных успехов в их решении: представители исследовательских и коммерческих проектов в области автоматической обработки естественного языка.

# Владимир Селегей

Владимир Селегей – заведующий кафедрой компьютерной лингвистики Института лингвистики РГГУ и НИУ МФТИ, директор по лингвистическим исследованиям компании АВВУУ. Председатель Оргкомитета крупнейшей российской конференции по компьютерной лингвистике «Диалог».



**Компьютерная лингвистика  
сегодня: от автоматической  
обработки текста до  
машинного понимания.**

Владимир Селегей

Компьютерная лингвистика сегодня

# С чего все началось



За полвека существования компьютерная лингвистика переживала периоды больших надежд и таких же больших разочарований.

Как будет развиваться компьютерная лингвистика, на что она способна сегодня, и что сможет завтра – об этом была первая лекция цикла.

# Владимир Селегей

## Компьютерная лингвистика сегодня: от автоматической обработки текста до машинного понимания

Одним из усвоенных уроков стало понимание того, что **компьютерная лингвистика и лингвистика теоретическая** – это пусть очень тесно связанные, но всё же разные науки.

Нечто похожее произошло в области моделирования шахматной игры, когда выяснилось, что не всегда решать задачу на компьютере нужно так, как предположительно решает ее человек.



# Владимир Плунгян:

## зачем лингвисты изучают языки?



Плунгян Владимир

2 ноября 2012,  
пятница



# Язык человека и компьютера

В компьютерной лингвистике **язык** рассматривается как одна из ментальных или интеллектуальных способностей человека наряду с памятью, вниманием, интуицией и пр.

Компьютер - это электронное устройство с программным обеспечением. Он не понимает человеческого языка. Между собой компьютеры «разговаривают» на компьютерном языке – языке двоичных кодов.

# Внутренний язык компьютера

Лет 20-25 назад, общение с вычислительными машинами было уделом немногих специалистов – программистов.

Сейчас компьютеры доступны даже детям.

Компьютеры становятся все более совершенными, но «собственный язык» компьютеров за все эти годы не изменился: кроме сложения нулей и единиц компьютер ничего делать не умеет.

# Внутренний язык компьютера

Внутренний язык компьютера по своей  
природе не совместим с языком  
человеческим,

и поэтому –

пользоваться людям языком компьютера в  
коммуникационном процессе общения  
**абсолютно невозможно!**

# Компьютерная лингвистика

Таким образом, компьютерная лингвистика – это актуальная область современного языкознания и информатики.

Компьютерная лингвистика занимается различными проблемами **компьютерной обработки** текстов на естественных языках, в том числе проблемами компьютерного перевода.

# Компьютерная лингвистика

Компьютерная лингвистика нужна  
специалистам, перед которыми  
стоят задачи построения **систем  
искусственного интеллекта**

И

всем интересующимся лингвистикой  
и ее приложениями.

# Дисциплина «Компьютерная лингвистика» по Марчуку Ю.Н. (МГУ)

Как учебная дисциплина, «компьютерная лингвистика находится **в процессе становления** и нет единой точки зрения на ее содержание, задачи и методы обучения»

(Марчук Ю.Н.)

# Дисциплина «Компьютерная лингвистика» по Марчуку Ю.Н. (МГУ)

Проектирование структуры и содержания данного курса основывалось на относительно сложившихся концепциях ведущих отечественных (МГУ, МГЛУ) и некоторых зарубежных школ США (университеты Денвера, Луисвилла, Флориды), а также общего подхода к обучению компьютерной грамотности и привития лингвистам профессиональной информационной культуры.

# Теоретические темы курса «Компьютерная лингвистика» в МГУ

В данном курсе рассматриваются темы связанные с освоением :

- информационных и коммуникационных технологий (ИКТ);
- основных положений лингвистики и перевода в точках соприкосновения естественных и формальных языков, ИКТ и технологий обучения, языка и коммуникации.

# Практическая часть курса «Компьютерная лингвистика»

Практическая часть курса в МГУ основана на освоении:

- лингвистических механизмов организации и функционирования сети Интернет (гипертекст, поиск, навигация, базы данных),
- включает задания по созданию учебных продуктов (перевод, совместный перевод, список ресурсов с описаниями).

# Учебные проекты при изучении компьютерной лингвистики

Курс сопровождается созданием учебных проектов с организацией виртуальных классов с группами студентов зарубежных университетов (Western Kentucky, Indiana-Purdue, USA, Valenciennes, France)

# Учебные проекты при изучении компьютерной лингвистики

В соответствии с целью реализуемого проекта студентами создается совместный учебный продукт:

**эссе, тематический рассказ, анализ языковых ошибок, особенностей электронной коммуникации и пр.**

# Достижение целей обучения

Достижение целей профессиональной информационной компетентности реализуется путем решения следующих задач:

обобщение теоретических знаний об основных понятиях, методах и направлениях компьютерной лингвистики как междисциплинарной научной области

сравнение моделей представления, хранения, обработки и передачи знаний (информации) в языкознании и информатике на основе естественных и формальных языков, средств ИКТ

ознакомление с общими принципами построения систем автоматической обработки текста, перевода, информационного поиска и компьютерной коммуникации и формирование умений и навыков работы с ними

# Лингвистическая информатика

**Лингвистическая информатика** интересуется естественным языком постольку, поскольку, во-первых, он является важнейшим средством получения, хранения, накопления и трансляции информации

# Лингвистическая информатика

В круг интересов «чистой» информатики входит создание и передача информации в виде схем, графиков, статичных картинок типа фотографий, анимационных блоков вплоть до появившихся в самое последнее время так называемых Computer Generated Images (CGIs)

CGIs – это электронные копии реальных людей (известных актеров и актрис), которые «снимаются» в электронном кинематографе (часто не ведая об этом)

# Лингвистическая информатика

Для человека наиболее приемлемый и простой способ общения — **языковой**.

Именно поэтому специалисту информатику приходится немало потрудиться с тем, чтобы **симулировать** появление у своего детища — персональной ЭВМ — некоего подобия естественно-языковой способности.

# Симуляция

**Симуляция** (от лат. *simulatio* — видимость, притворство), ложное изображение болезни или отдельных её симптомов человеком, не страдающим данным заболеванием.

В медицине различают умышленную (с корыстными целями) и патологическую симуляцию (болезненную).

•

# Задача лингвистической информатики

**Задачей лингвистической информатики является обеспечение связи между компьютером и пользователем, каковым выступает человек**

Для человека наиболее приемлемый и простой способ общения — языковой, а программистам-информатикам приходится потрудиться и **именно симулировать у компьютера подобие естественно-языковой способности**

# Растет производительность компьютеров ...

В начале 60-х годов в связи с бурным развитием информационных технологий и ростом производительности вычислительных систем формируется новое научное направление — коммутационная лингвистика.

# Растет производительность компьютеров ...

Бóльшая часть информации, накопленной обществом к настоящему времени - представлена в языковом виде.

Это наводит на мысль о близости сфер интересов лингвистики и информатики.

# Растет производительность компьютеров и тогда ...

... возникает заманчивая идея использовать компьютер для решения лингвистических задач: индексирование, реферирование, аннотирование, перевод, обработка текста, установление авторства, формирование репрезентативных корпусов и пр.

# Растет производительность компьютеров и тогда ...

Производительность компьютера – это хорошо.

Однако, без должного **лингвистического обеспечения** современные информационные системы и технологии зайдут в тупик и не смогут помогать человеку справляться с колоссальным потоком информации.

# Коммутационная лингвистика



Коммутационная лингвистика  
смотрит на компьютер как на  
эффективное и удобное  
**средство** максимально полного  
и всестороннего изучения языка  
во всех его ипостасях и формах.

# Коммутационная лингвистика

Несмотря разнонаправленность целевых векторов **коммутационной лингвистики** и **лингвистической информатики**, эти направления не являются антагонистичными

2+dGCDKROA75BITDxtgWbN93X  
Hq1FOGSU1YychqT69c1Iome  
YiuDkToC3CC+8h17kdvkUoch  
7fT/GRJqIh9ZX+jL4qfuNC  
icZgFlag320QSFmk191DM600  
x06bH7RMAh014xGS1Fnq  
hFaOvrFhHETU1160

gEUEZ+Y21D  
6iBwAe1Q1Vodc  
p47H+nSPVCDx  
c6z1A72IH

c3CC+8h1  
GRJqIh9ZX+jL  
Flag320QSFmk19  
7RMAh014xGS  
FhHETU1



Они дополняют друг друга, обеспечивая **синергетический** (взаимно усиливающий) эффект при решении информационных и лингвистических задач пользователей.

# Лингвистическая информатика

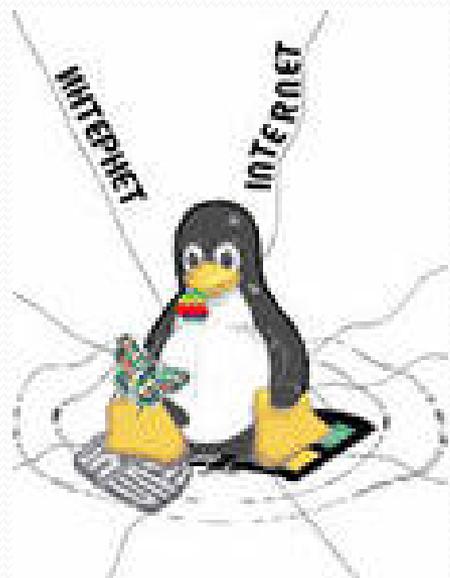
**Лингвистическая информатика** смотрит на язык как на средство оптимизации диалога между человеком и компьютером, между человеком и человеком посредством компьютера, а также как на своего рода модель, по образу и подобию которой можно создавать существенную часть программного обеспечения.



# Математическая лингвистика

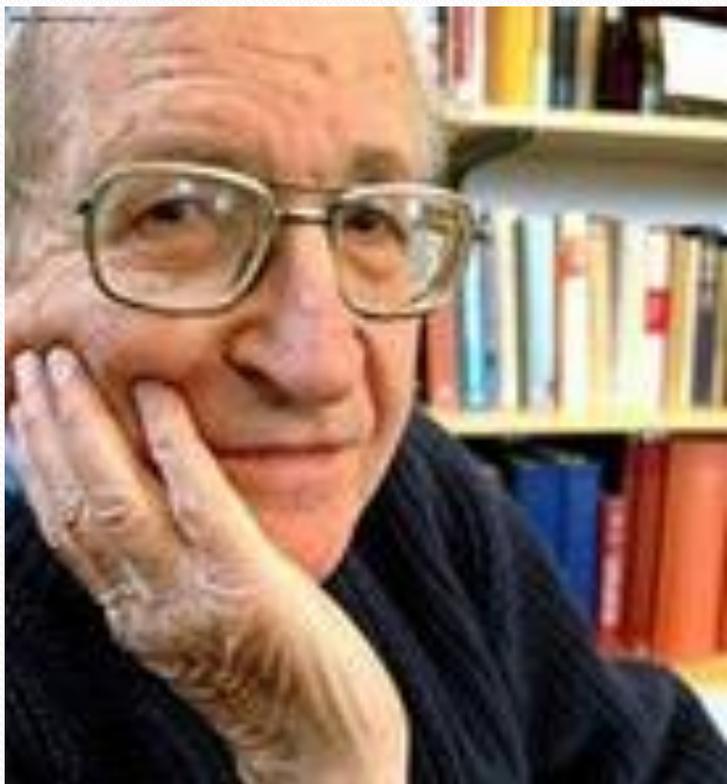
В 50-х годах XX века (в связи с назревшей в языкознании потребностью уточнения его основных понятий) возникает **математическая лингвистика**.

Математическая лингвистика разрабатывает **формальный аппарат** для описания строения естественных и некоторых искусственных языков.



Центральное место в современной математической лингвистике принадлежит теории формальных грамматик, основанной на работах Н. Хомского.

# Ноам Хомский



**Аврам Ноам Хомский** — американский лингвист, политический публицист и теоретик.

Институтский профессор лингвистики Массачусетского технологического института, автор классификации формальных языков, называемой *иерархией Хомского*. Его работы о порождающих грамматиках внесли значительный вклад в развитие когнитивных наук.

Хомский предположил, что люди обладают врожденной способностью к языку

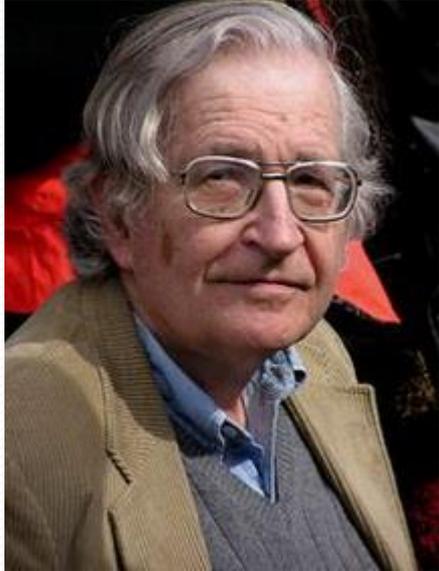
# Глубинная и поверхностная структуры языка



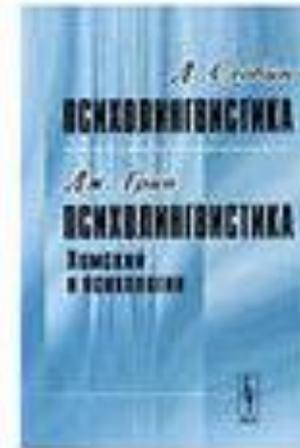
Хомский полагает, что дети быстро развивают языковые навыки, позволяющие им извлекать одинаковый смысл (глубинную структуру) из различного порядка слов (поверхностной структуры)

Например, в предложениях: «собака погналась за мальчиком» и «мальчика преследовала собака» поверхностная структура различается, но глубинная структура одинакова

# Универсальные правила языка



Воспринимаемая речь на конкретном языке, естественный механизм овладения языком извлекает из нее универсальные правила, подходящие для ЭТОГО языка.



# Язык как системный механизм



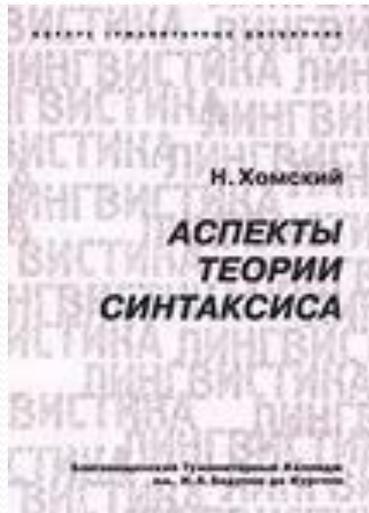
Фердинант де Соссюр:  
Язык – это системный механизм, функционирование которого проявляется в речевой деятельности его носителей



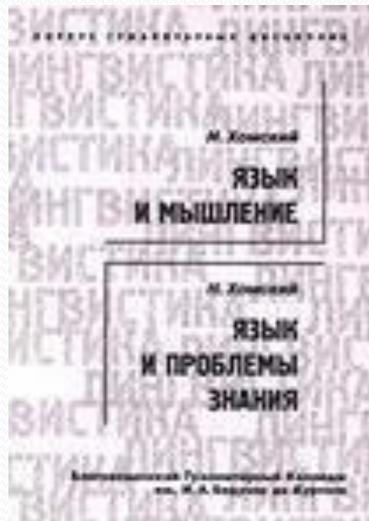
Язык подчиняется определенным закономерностям, что ввело лингвистику в своего рода «семиотический цикл» дисциплин, к которому принадлежат математика и классическая аристотелевская логика



# Языкознание и математика



Взаимодействие языкознания и математики имеет богатую и во многом противоречивую историческую традицию.



Задача построения формальной математической модели языка и речи на протяжении десятилетий занимала и занимает умы многих исследователей.

# Компьютерная лингвистика

Компьютерная лингвистика уже 50 лет с переменным успехом пытается разработать алгоритмы автоматического понимания текста

Фундаментальная проблема заключается в алгоритмоцентрическом характере систем обработки лингвистической информации: их центральным понятием является **алгоритм** — формально определенная заранее заданная последовательность действий, которые преобразуют исходные данные в конечный результат решения поставленной задачи

# Компутационная лингвистика

Коммутационная лингвистика  
отводит человеку лишь роль  
стороннего наблюдателя,  
ожидающего получить на выходе  
алгоритма **готовый результат**  
**вычислений**

# Коммутационная лингвистика

Известно, что никакой алгоритм не может работать с неформализованными данными.

Язык и речь можно формализовать лишь на **поверхностном уровне.**

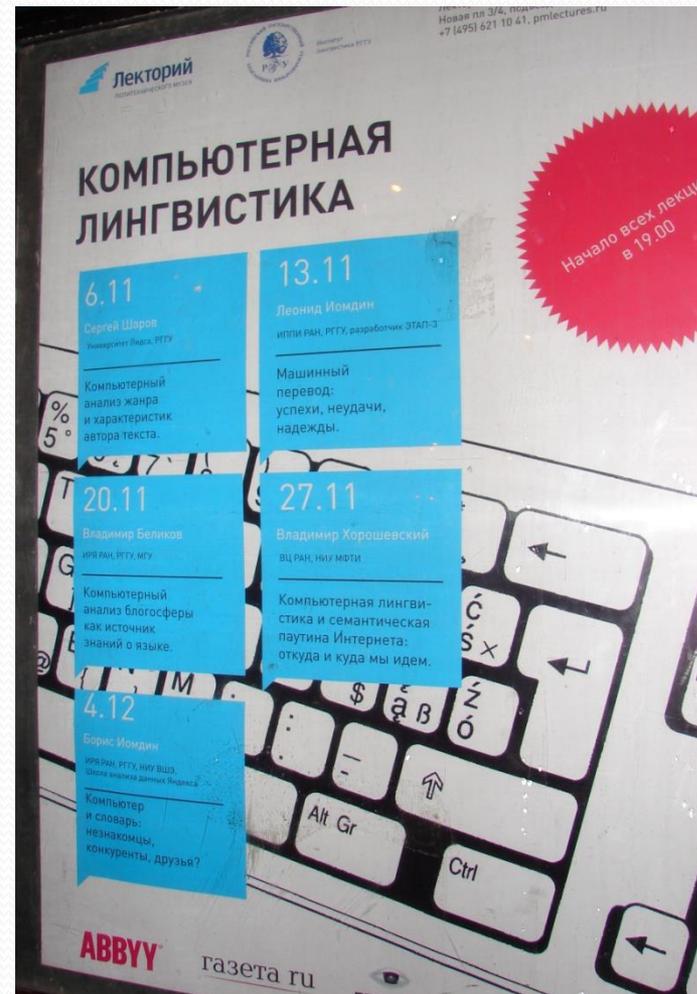
Задача анализа речевых высказываний достаточно глубокого для понимания их смысла, **до сих пор формально не решена**

# О компьютерной лингвистике

## сегодня

К чему пришли

современные ученые в связи с проблемой обработки текста с помощью компьютера, было представлено в цикле лекций на тему «Компьютерная лингвистика»



# Иомдин Леонид



Кандидат филологических наук, ведущий научный сотрудник, и.о. зав. Лабораторией компьютерной лингвистики Института проблем передачи им. А.А. Харкевича Российской академии наук, доцент Института лингвистики РГГУ.



Лингвист, специалист по современному синтаксису и семантике, компьютерной лингвистике и машинному переводу.

# Иомдин Леонид



Ведущий разработчик известной системы автоматического перевода ЭТАП-3. Работал в крупных европейских проектах по машинному переводу. Преподавал теоретическую и компьютерную лингвистику в ряде западноевропейских университетов (Мюнхен, Прага, Барселона).

Один из авторов Нового большого англо-русского словаря под редакцией акад. Ю. Д. Апресяна.

Переводчик-синхронист.

## Машинный перевод: успехи, неудачи, надежды



История машинного перевода текстов с одного языка на другой с помощью компьютера – насчитывает без малого **шестьдесят** лет.

# Машинный перевод: успехи, неудачи, надежды

За это время сменилось несколько поколений систем машинного перевода: от почти игрушечных моделей, переводивших текст слово за словом без учета контекста, ученые перешли к сложным системам, создавая правила, учитывающие тонкие смысловые оттенки переводимого текста.

# Машинный перевод: успехи, неудачи, надежды

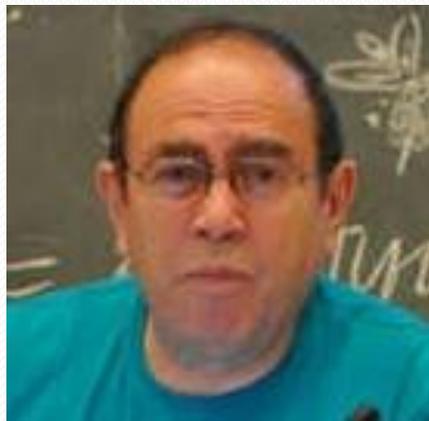
Наряду с системами перевода, основанными на правилах, стали создаваться «статистические» системы, обращающиеся к сверхбольшим корпусам параллельных текстов и находящие в них наилучшие эквиваленты для как можно более крупных фрагментов переводимого текста.

# Машинный перевод: успехи, неудачи, надежды



В дополнение к системам перевода письменных текстов приобретают все более широкое распространение системы устного перевода, которые распознают живую речь на входном языке и синтезируют звучащий текст на выходном языке.

# Машинный перевод: успехи, неудачи, надежды



В истории машинного перевода были свои взлеты и падения: энтузиазм первопроходцев сменялся глубоким пессимизмом, когда видные специалисты приходили к убеждению, что задача машинного перевода не может быть решена в обозримом будущем.

Сейчас машинный перевод **переживает второе рождение**: благодаря сочетанию различных методов и подходов качество перевода заметно улучшается и в эту область вовлекаются все новые языки.

# Беликов Владимир Иванович

Российский лингвист, этнограф, социолог, доктор филологических наук. Известен публикациями в разных гуманитарных науках

Последние годы занимается изучением социальных различий в русском языке.

А настоящее время ведущий научный сотрудник Отдела культуры русской речи Института русского языка РАН.



# Компьютерный анализ блогосферы как источник знаний о языке

В написанном в конце прошлого века учебнике социолингвистики говорилось, что дневниковые записи «среднестатистических» носителей языка представляют для исследователей языка исключительно интересный, но труднодоступный материал

# Компьютерный анализ блогосферы как источник знаний о языке

Широкое распространение новых информационных технологий и блогосферы

дало неограниченные возможности для компьютерного анализа повседневных текстов

# Компьютерный анализ блогосферы как источник знаний о языке

Теперь лингвисты могут работать с повседневными записями, авторы которых легко классифицируются по важнейшим социолингвистическим параметрам.

Теперь несложно получить объективные данные о языке мужчин и женщин, подростков и лиц старших возрастов в любом регионе, стала доступной статистика языковых изменений за последнее десятилетие.

# Хорошевский Владимир Федорович



Доктор технических наук,  
профессор, заведующий сектором  
«Интеллектуальные прикладные  
системы» отдела  
«Интеллектуальные системы»  
Вычислительного центра им. А.А.  
Дородницына РАН.

Профессор базовой кафедры  
«Интеллектуальные системы»  
НИУ МФТИ.

# В. Хорошевский

## Компьютерная лингвистика и семантическая паутина Интернета: откуда и куда мы идем



Один из основателей  
ассоциации  
искусственного  
интеллекта СССР (САИИ),  
а затем России (РАИИ).

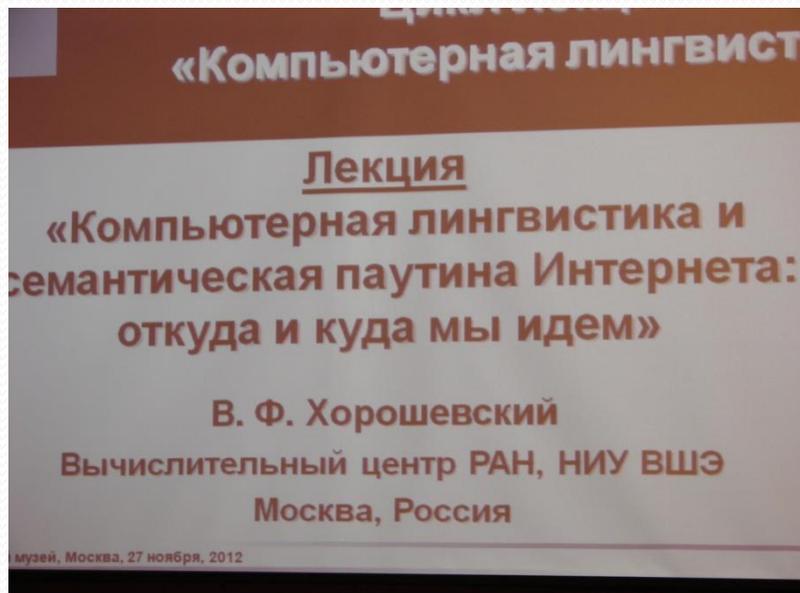
# Компьютерная лингвистика и семантическая паутина Интернета: откуда и куда мы идем

Лекция посвящена обсуждению 3-х тем:

- 1) ретроспективному обзору методов и средств компьютерной лингвистики, которые в настоящее время активно используются при построении систем извлечения знаний из текстов

# Компьютерная лингвистика и семантическая паутина Интернета: откуда и куда мы идем

- 2) обсуждению концепции семантической паутины, которая постепенно трансформирует современный Интернет за счет формирования пространств знаний



# Компьютерная лингвистика и семантическая паутина Интернета: откуда и куда мы идем

3) сравнительному анализу существующих в данной области решений и результатов:

1, 2, 3, 4, 5, 6, 7, 8, ...



**Несколько замечаний вместо заключения**

- ❖ В мире активно ведутся не только исследования и разработки в области извлечения информации из текстов, но и существуют реальные прикладные системы.
- ❖ Для российских специалистов в области извлечения информации из текстов время полномасштабного и скоординированного развертывания исследований и разработок в этой важнейшей области компьютерной лингвистики и интеллектуальных технологий еще впереди.

**Родуктов, связанных с обработкой естественного языка и семантическими технологиями, RCO занимает положение в течение последних 3-5 лет.**

**Объекты:**

- Гарний разработка Morphology SDK
- Morphology Professional SDK
- Fact Extractor SDK
- Гарний разработка RCO Fact Extractor
- RCO КАОТ

# Компьютерная лингвистика и семантическая паутина Интернета: откуда и куда мы идем

Обсуждение сопровождалось презентацией семантических порталов, функционирующих в Интернет, где уже используются системы обработки естественного языка.

# Иомдин Борис



Кандидат филологических наук,  
старший научный сотрудник  
Института русского языка им.  
В. В. Виноградова Российской  
академии наук, доцент  
Института лингвистики РГГУ  
и факультета филологии  
Высшей школы экономики,  
преподаватель Школы анализа  
данных Яндекса.



# Иомдин Борис



Один из авторов словарей и монографий по семантике и лексикографии под редакцией акад. Ю. Д. Апресяна («Новый большой англо-русский словарь», «Новый объяснительный словарь синонимов», «Перспектив активнаго словаря русскаго языка», «Русская языковая картина мира и системная лексикография»), автор научных статей.

Организатор словарного проекта и научного семинара «Словарь бытовой терминологии». Автор множества лингвистических задач, член методических комиссий Всероссийского конкурса-игры «Русский медвежонок — языкознание для всех», Московской традиционной олимпиады по лингвистике, международной лингвистической олимпиады и других олимпиад и конкурсов по лингвистике и русскому языку.

# Иомдин Борис



Член совета сайта  
«Лингвистика для  
школьников» и  
оргкомитета Летней  
лингвистической школы



# Компьютер и словарь: незнакомцы, конкуренты, друзья?

Когда-то толкования непонятных слов писали прямо на полях рукописей, потом стали собирать их в отдельные книги – словари, позже возникла целая наука лексикография.

Лексикографы выписывали примеры употребления слов на карточки и составляли из них огромные картотеки – еще не так давно шкафы со словарными карточками стояли во всех коридорах Института русского языка.

# Компьютер и словарь: незнакомцы, конкуренты, друзья?

С появлением компьютеров карточки стали стремительно уходить в прошлое, а в последнее время они потянули за собой и словари.

Все чаще люди просто смотрят значение слова в Википедии или вводят его в окошко онлайн-машинного переводчика.

Как повлиял компьютер на развитие лексикографии, ее материала, методов и практических результатов?

Какие новые задачи ставит перед лексикографами развитие компьютеров и Интернета? И есть ли у словаря будущее в новую эпоху?

# Компьютер и словарь: незнакомцы, конкуренты, друзья?

Как повлиял компьютер на развитие лексикографии, ее материала, методов и практических результатов?

Какие новые задачи ставит перед лексикографами развитие компьютеров и Интернета? И есть ли у словаря будущее в новую эпоху?

На эти и другие вопросы компьютерной лингвистики лектор дал самые подробные ответы: [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), ...

# Сергей Шаров



Кандидат физико-математических наук, профессор кафедры переводоведения Университета Лидса и доцент кафедры компьютерной лингвистики РГГУ

# Компьютерный анализ жанра и характеристик автора текста



В лекции рассматриваются подходы к автоматическому определению этих характеристик

# Компьютерный анализ жанра и характеристик автора текста

Определение тематики текста является достаточно простой задачей.

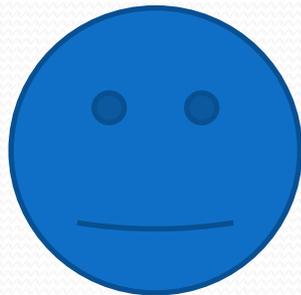
Так, например, Яндекс/Новости определяет сюжеты новостей по совпадению ключевых слов)

А вот жанр текста, характеристики его автора (например, пол, возраст) не зависят прямо от ключевых слов, поскольку текст в одном и том же жанре или одного и того же автора может быть написан на разные темы

# Какое впечатление лекции произвели на студентов

(видеосюжеты с семинара по обсуждению результатов посещения лекций)

1



2



3





# СПАСИБО ЗА ВНИМАНИЕ

ФАКУЛЬТЕТ лингвистики МИМ ЛИНК

Михайлова Т.О.

Абрамова Е.И.

Матвеева Н.В.