

S T U D I A P H I L O L O G I C A



ИНСТИТУТ ЯЗЫКОЗНАНИЯ РАН

ЯЗЫК и ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Сборник статей по итогам конференции
«Лингвистический форум 2020:
Язык и искусственный интеллект»

Под редакцией А. В. Вдовиченко



Издательский Дом ЯСК

Москва 2023

УДК 81
ББК 32.813
Я 41

Утверждено к печати Ученым советом
Федерального государственного бюджетного учреждения науки
Института языкознания РАН

Рецензенты:

доктор физико-математических наук *В. Д. Соловьев*
кандидат филологических наук *А. А. Котов*

Под редакцией *А. В. Вдовиченко*

Я 41 Язык и искусственный интеллект / Сборник статей
по итогам конференции «Лингвистический форум 2020: Язык
и искусственный интеллект», Москва, ноябрь 2020 г. — М.:
Издательский Дом ЯСК, 2023. — 328 с. — (Studia philologica).

ISBN 978-5-907498-47-1

В сборнике представлены статьи участников конференции «Лингвистический форум 2020: Язык и искусственный интеллект» (Институт языкознания РАН, Москва, ноябрь 2020 г.), отражающие общую и частную проблематику научных исследований в области лингвистики и компьютерных технологий. Авторы публикуемых статей предлагают решения специальных вопросов на фоне более масштабных объектов научной эвристики, обязательных при разработке и применении искусственного интеллекта: мозг, сознание, мышление, дискретность/недискретность единиц речи и текста, машинная обработка вербальных данных и роль автора, знаковая деятельность и смыслообразование, механизмы коммуникации. Сборник адресован специалистам, а также всем, кто интересуется лингвистическими аспектами развития компьютерных технологий.

УДК 81
ББК 32.813

В дизайне обложки использованы работы художника Д. А. Федорова

ISBN 978-5-907498-47-1



© Коллектив авторов, текст, 2023
© Издательский Дом ЯСК, оригинал-макет, 2023

СОДЕРЖАНИЕ

<i>А. А. Кибрик. О лингвистическом форуме</i>	7
<i>А. В. Вдовиченко. Предисловие редактора</i>	9
 <i>Н. Ш. Александрова</i>	
О ПОНИМАНИИ РЕЧИ ЧЕЛОВЕКОМ. В ПОИСКАХ «КОНСТРУКТОРСКИХ ИДЕЙ» ПРИРОДЫ	13
 <i>В. П. Белянин</i>	
КАТЕГОРИАЛЬНЫЙ АНАЛИЗ ТОНАЛЬНОСТИ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ	34
 <i>Е. Г. Борисова</i>	
ФОРМАЛИЗУЕМЫЕ СПОСОБЫ ОТРАЖЕНИЯ НЕДИСКРЕТНЫХ ОСОБЕННОСТЕЙ ЯЗЫКА: МОДЕЛИРОВАНИЕ ВЫБОРА ЕДИНИЦ И ИМПЛИКАТУР	53
 <i>В. Л. Введенский, К. Г. Гуртовой</i>	
СТРУКТУРЫ В МОЗГЕ ЧЕЛОВЕКА, РАСПОЗНАЮЩИЕ СМЫСЛ СЛОВ	67
 <i>А. В. Вдовиченко</i>	
ЗНАКОВЫЙ ПРОЦЕСС В ЕСТЕСТВЕННОМ И ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ: СЕМИОТИЧЕСКОЕ ВОЗДЕЙСТВИЕ И АЛГОРИТМ	91
 <i>Д. А. Девяткин, Л. А. Каджая, Н. В. Чудова, В. А. Мишланов, В. А. Салимовский</i>	
ИССЛЕДОВАНИЕ РЕЧЕВЫХ ЖАНРОВ НАУЧНОГО АКАДЕМИЧЕСКОГО ТЕКСТА МЕТОДАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	107
 <i>Д. А. Залманов, А. А. Кибрик</i>	
УРОВЕНЬ АКТИВАЦИИ И СПЕЦИАЛЬНЫЕ ОГРАНИЧЕНИЯ ПРИ МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ РЕФЕРЕНЦИАЛЬНОГО ВЫБОРА	142

<i>А. В. Колмогорова, А. А. Калинин</i> ВИЗУАЛИЗАЦИЯ ДАННЫХ В ЭМОЦИОНАЛЬНОМ АНАЛИЗЕ РУССКОЯЗЫЧНЫХ ИНТЕРНЕТ-ТЕКСТОВ НА ОСНОВЕ МОДЕЛИ «КУБ ЛЁВХЕЙМА»	167
<i>Н. В. Лавров, В. В. Лавров</i> ОСОБЕННОСТИ ВЗАИМОПОНИМАНИЯ В КОНТЕКСТЕ ДИАЛОГА: СПЕЦИФИЧЕСКИЕ И НЕСПЕЦИФИЧЕСКИЕ ФАКТОРЫ КОММУНИКАЦИИ	182
<i>И. Г. Подгорбунская</i> К ВОПРОСУ ОБ ИЗМЕРЕНИИ КОЛИЧЕСТВА ИНФОРМАЦИИ: ВЕРБАЛЬНЫЙ КОД МИНИМАЛЬНОЙ ИНФОРМЕМЫ	226
<i>Н. К. Рябцева</i> ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПРОЦЕССАХ ПЕРЕВОДА	237
<i>У. В. Смирнова</i> ТРАНСЛИНГВАЛЬНЫЕ ОСОБЕННОСТИ ЗНАКОВ (НА ПРИМЕРЕ РЕКЛАМНЫХ ВЫВЕСОК ЦЕНТРАЛЬНЫХ УЛИЦ ГОРОДА МОСКВЫ)	262
<i>О. Л. Филипеня, В. В. Ткаченко</i> ЛИНГВОКОГНИТИВНЫЙ МЕТОД АЛГОРИТМИЗАЦИИ ПРОЦЕССОВ ЛОГИЧЕСКОГО МЫШЛЕНИЯ	273
<i>Е. В. Чистова</i> ДИЗАЙН КОГНИТИВНОЙ СРЕДЫ VR-ТРЕНАЖЕРА ДЛЯ СИНХРОННОГО ПЕРЕВОДА	287
<i>С. А. Шаповал</i> ДИСКУРСИВНЫЙ АНАЛИЗ ТЕКСТОВ НЕЙРОСЕТИ ПОРФИРЬЕВИЧ	308

О ЛИНГВИСТИЧЕСКОМ ФОРУМЕ

Начиная с 2019 года в Институте языкознания РАН проводится ежегодная международная конференция «Лингвистический форум». Лингвистический форум задуман как центральная годовая конференция Института языкознания. Каждый из форумов посвящен конкретной теме, причем темы выбираются общественно значимые, то есть представляющие интерес не только для профессиональных лингвистов, но и для более широких кругов.

В 2020 году было решено посвятить второй Лингвистический форум теме «Язык и искусственный интеллект». В информационном письме (https://iling-ran.ru/web/ru/conferences/2020_lingforum/call_for_papers) было сказано:

Цель форума 2020 — диалог специалистов, занимающихся исследованиями на стыке лингвистики и искусственного интеллекта, в том числе в области компьютерной лингвистики / Natural Language Processing. Развитие методов искусственного интеллекта обеспечило последние успехи технологий, связанных с порождением и пониманием языка, и расширило границы их применения. Нейросетевые подходы и векторные репрезентации замещают инженерию признаков, основанных на традиционных дискретных категориях лингвистического описания. Граница между фундаментальными и прикладными лингвистическими исследованиями стирается. Эмпирическая лингвистика берет на вооружение новые технологии, в частности, для нужд моделирования языка и документации данных. Искусственный интеллект все активнее входит в повседневную жизнь носителей языка. Может ли фундаментальная лингвистика на современном этапе предложить технологически реализуемые идеи или методы? Эти и подобные концептуальные и методологические проблемы будут в центре обсуждения на форуме.

К числу актуальных вопросов для приоритетного обсуждения Программный комитет форума отнес следующие:

- Нейронные сети в ряду других методов машинного обучения и проблемы их интерпретируемости;
- Порождение vs понимание языка — общее и различное;

- Языковое разнообразие и мультилингвизм как проблемы для искусственного интеллекта;
- Типы дискурса: модусы, жанры, стили;
- Сложность и связность текста;
- Язык в контексте мультимодальной коммуникации;
- Моделирование индивидуальных различий в языке;
- Недискретные явления в языке;
- Возникающие и будущие лингвистические технологии.

Форум получился по-настоящему международным и масштабным. В Программном комитете были представлены ученые из пяти стран. Было сделано более 70 докладов, в том числе три пленарных доклада:

- С. О. Кузнецов (НИУ ВШЭ). Искусственный интеллект. Основные направления и вызовы.
- А. С. Бердичевский (Университет Гетеборга). Корпуса, векторы, модели: компьютерные подходы к лингвистической типологии.
- Д. Талбот (Яндекс). Получение знаний о языке на основе данных.

На форуме встретились ученые, которые в «обычной жизни» практически не пересекаются, в том числе представители компьютерных наук, лингвистики, других дисциплин. В ряде докладов был представлен общегуманитарный и философский контекст искусственного интеллекта. Я полагаю, что эти встречи были полезными. Надеюсь также, что форум был определенным шагом в сторону будущего, в котором идеи и предложения лингвистов будут востребованы разработчиками компьютерных систем.

По материалам докладов, сделанных на форуме, ранее уже вышел один сборник:

Proceedings of the Linguistic Forum 2020: Language and Artificial Intelligence. Moscow, Russia, November 12–14, 2020. Edited by Valery Solovyev, Natalia Loukachevitch, Olga Lyashevskaya. CEUR Workshop Proceedings, Vol-2852, 2021. <http://ceur-ws.org/Vol-2852>.

Настоящий сборник дополняет и расширяет интеллектуальную картину, представленную на форуме 2020 года. Я благодарен Алексею Дмитриевичу Кошелеву за инициативу по созданию сборника, а Андрею Викторовичу Вдовиченко — за деятельную работу по подготовке издания.

А. А. Кибрик,
председатель Программного комитета конференции
«Лингвистический форум 2020: Язык и искусственный интеллект»

ПРЕДИСЛОВИЕ РЕДАКТОРА

Никто из авторов, представленных в сборнике, не согласовывал свою статью с общим замыслом, не вписывал свой труд в тему коллективной монографии, которой настоящий сборник не является и не замышлялся как таковой. Однако обнаружить нечто объединяющее — значит, оправдать книгу в данном качестве, дать ей возможность существовать в виде идейной целостности, которую заинтересованный читатель надеется обрести, держа в руках материальный носитель.

Тематику и название сборника определила конференция «Лингвистический форум 2020: Язык и искусственный интеллект» (Москва, Институт языкознания РАН, ноябрь 2020 г.). Выступления некоторых участников прямо или косвенно связаны с содержанием их статей, представленных здесь. Вместе составлять один из итогов форума — это уже фактор единства, скорее, правда, формальный.

Содержательная общность обнаруживается в наборе важных лингвистических (и в целом гуманитарных) вопросов, на которые каждый автор так или иначе дает ответы и которые встают с новой силой при попытках приспособить «человеческий язык» к действиям искусственного интеллекта «на входе и выходе»:

- В чем состоят сходства и различия в пользовании знаками, в том числе вербальными, со стороны машины и обладателя естественного интеллекта (те сходства и различия, которые позволяют или, наоборот, препятствуют довериться обманчивым свидетельствам «мыслительного процесса» машины)?
- Работает ли естественный интеллект (мозг, сознание, мышление и пр.) на основе вербальных — или каких-то еще — знаков (тех самых, правильное пользование которыми считается доказательством корректной работы искусственного интеллекта)?
- Как возникает связь между телом знака и содержательным контентом, семантикой (та связь, которая известна по классической диаде «означающее — означаемое» в языке, где она почему-то считается прочной, автономной, ничем не опосредованной, в то время как опосредование

семантики алгоритмом машины или сознанием человека составляет ключевую проблему формирования конкретного «означаемого»)?

- Каковы единицы когнитивно-коммуникативного процесса, и существуют ли они в дискретной форме, пригодной для применения затем в искусственном интеллекте (те единицы, которые зачастую известны лингвисту как «слова со своими значениями», которые почему-то сами не могут иметь определенную семантику без недискретного «контекста»)?
- Считать ли естественный слово-содержащий семиотический процесс «работой языка» (тот самый многофакторный полимодальный процесс, который явно не исчерпывается вербальными данными, но который все равно в виде «языка» зачастую пытаются привить искусственному интеллекту)?
- Стоит ли отождествлять когнитивный процесс, «мысль», и семиотический процесс, так называемую «передачу мысли» (те самые процессы, которые полностью совпадают в «интеллекте» машины, а в человеческом сознании принципиально разделены)?
- Сравнимы ли действия, производимые естественным и искусственным интеллектами (одни, которые совершаются как попытки целенаправленного лично осознанного и персонально ответственного изменения внешних психо-когнитивных состояний, и другие, возникающие как строгие исполнения заложенных команд)?
- Можно ли научить искусственный интеллект правдоподобно имитировать не только «говорение» и «понимание», но и «воплощенность» (то есть быть столь же воплощенным, каким без сомнения является, а не кажется естественный интеллект, черпающий свою естественность как раз из воплощенности)?
- Не слишком ли амбициозна претензия создателей быстродействующего устройства называть его не просто эффективным инструментом, удобным помощником, полезным имитатором, но Самим Обладателем Интеллекта?

Последовательно отвечая на эти и другие вопросы, можно прийти к заключению, что аппаратно-программный комплекс, называемый искусственным интеллектом, никогда не сможет стать в подлинном смысле интеллектом, подобно, скажем, птице, наученной произносить по утрам «Вперед, Петр Петрович», которая не мыслит и не говорит, как человек, несмотря на несомненную уместность и правильность произносимых слов.

Можно, впрочем, прийти и к прямо противоположному заключению и подобрать совершенно другую развернутую метафору, в которой

естественный интеллект перестанет, наконец, быть «мерилом всего» и сам на соответствующем фоне окажется лишь говорящей птицей. Местами такое уже происходит и в кого-то вселяет надежду.

Как бы то ни было, совершенствовать и оптимизировать помощника-имитатора, коим искусственный интеллект с определенностью уже зарекомендовал себя на настоящий момент, — несомненно нужная и перспективная задача. Вопросы-сомнения остаются в виде небесполезного *caveat*. Атрибут «искусственный» ограждает любой заново получаемый результат (очередную версию интеллекта и/или его применения) от возможных упреков перфекционистов и скептиков, подчеркивает постепенность и рукотворность создаваемого продукта.

Авторы публикуемых статей предлагают решения специальных вопросов на фоне (или внутри) более масштабных объектов научной эвристики, обязательных при разработке и применении искусственного интеллекта: мозг, сознание, мышление (Н. Ш. Александрова; В. Л. Введенский, К. Г. Гуртовой; О. Л. Филипья, В. В. Ткаченко), дискретность и недискретность единиц речи и текста (Е. Г. Борисова; Д. А. Залманов, А. А. Кибрик; И. Г. Подгорбунская), машинная обработка вербальных данных и роль автора (В. П. Белянин; Д. А. Девяткин, Л. А. Каджая, Н. В. Чудова, В. А. Мишланов, В. А. Салимовский; А. В. Колмогорова, А. А. Калинин), знаковая деятельность и смыслообразование (А. В. Вдовиченко; У. В. Смирнова; С. А. Шаповал), механизмы коммуникации и языка (Н. В. Лавров, В. В. Лавров; Н. К. Рябцева; Е. В. Чистова).

Кроме самих участников сборника, внесших наиболее весомый содержательный вклад в его создание, слова благодарности и признательности нужно непременно адресовать А. А. Кибрику и А. Д. Кошелеву, взявшим на себя труд задумать, организовать и координировать издание, а также всем, кто занимался технической стороной публикации столь непростого текста, с его иллюстрациями, таблицами, формулами, схемами. Кроме того, в подготовке издания на раннем этапе деятельно участвовала О. С. Орлова.

А. В. Вдовиченко

Н. Ш. Александрова
Sprachbrücke e.V. Берлин, Германия

О ПОНИМАНИИ РЕЧИ ЧЕЛОВЕКОМ. В ПОИСКАХ «КОНСТРУКТОРСКИХ ИДЕЙ» ПРИРОДЫ

Понимание речи, как, вероятно, и подавляющее большинство других когнитивных функций человека, обеспечивается совместной работой двух разнонаправленных мозговых механизмов. При нормальном функционировании механизмы сосуществуют, дополняют друг друга и разглядеть каждый из них «по одному» чрезвычайно сложно, но в случае патологии, когда один из механизмов слабеет, а второй продолжает работать, это становится возможным. Также при освоении взрослыми нового языка в языковой среде можно наблюдать процессы, которые человек переживает в раннем детстве при освоении родного языка. Опираясь на данные языковой патологии и билингвизма, один из механизмов понимания речи можно охарактеризовать как произвольное понимание общего смысла обращенной речи без опоры на отдельные слова, изолированно его можно наблюдать и исследовать при освоении новых языков в языковой среде, т. е. при становлении естественного билингвизма. По своей сути этот процесс — извлечение общего смысла высказываний из конкретной ситуации, опираясь на весь комплекс невербальных средств коммуникации. Второй механизм человеческого понимания речи — понимание произвольное, осознанное, опирающееся на известные слова и требующее включения произвольного внимания. Изолированно этот механизм можно наблюдать при обучении языкам детей с импрессивной (сенсорной) алалией, у которых стойко нарушена способность освоить язык как родной. Описанные механизмы

понимания речи — проявление в языковой сфере двух глобальных путей человеческого познания.

Ключевые слова: родной язык, иностранный язык, пластичность мозга, видовое (спонтанное) знание *Homo sapiens*, понимание речи, импрессивная (сенсорная) алалия, билингвизм.

Изучение человеческого мозга во многом напоминает изучение сложной системы без понимания ее основных «конструкторских идей». Мы знаем многое о его внутренней структуре, глубоко изучены процессы, протекающие в нейронах, но... обилие разнообразных знаний заслоняет понимание достаточно простых принципов, лежащих в самой основе.

А. Редозубов

Эпиграф взят с сайта, посвященного искусственному интеллекту. Позиция автора этих строк мне близка и понятна, т. к. многолетний опыт помощи пациентам в восстановлении когнитивных функций привел меня к тому же выводу: существует вполне логичный, доступный человеческому познанию уровень организации функций мозга.

«Как же все это устроено?» — этот вопрос, вероятно, задают себе многие специалисты, пытающиеся помочь пациентам после мозговых поражений. Работа эта во многом интуитивная, т. к. нет достаточных теоретических опор. А клиническая картина бывает не просто сложная, но и парадоксальная: например, пациент нормально видит и пишет, но не может прочесть даже то, что сам написал, или без труда узнает цифры, но не может узнать буквы и 0 для него — только ноль, но не буква и т. п. Специалист может обследовать больного и «рассмотреть» нарушенную деталь. Но при этом организация здоровой функциональной системы остается неизвестной. То есть стараемся восстановить то, что непонятно как устроено. И нередко получается. И одновременно раскрываются сложные, но в то же время познаваемые закономерности мозговой организации.

Известно, что при разработке искусственного интеллекта большинство задач решаются без опоры на знание человеческого мозга. И тут

причиной является не только объективная сложность мозговой организации, но и недостаточное описание именно основных природных «конструкторских идей». Предлагаемое описание и анализ некоторых таких идей может либо помочь разработчикам искусственного интеллекта, либо укрепит их решение идти своим путем.

1. Два полушария и две стратегии обработки информации

Давно известно о различных стратегиях переработки информации и регуляции функций, которые свойственны правому (наглядно-образная, конкретная, непосредственная, синтетическая, симультанная, бессознательная, произвольная регуляция) и левому (вербально-логическая, абстрактно-схематическая, аналитическая, сукцессивная, с участием сознания, произвольная регуляция) полушариям мозга. Осмысление асимметрии полушарий имеет долгую историю, и оно не закончено, оно продолжается. Интерпретация поистине огромного количества клинических наблюдений и всевозможных исследований может быть разной и меняется, если появляются новые данные. Но очевидно, что одна из главных конструкторских идей природы — двойственность организации. И те закономерности, которых мы далее коснемся, как правило, проявляются как дихотомии, т. е. как два противоположно направленных механизма, совместное функционирование которых создает эффект континуума. При нормальном функционировании разглядеть каждый из механизмов «по одному» чрезвычайно сложно, но в случае патологии, когда один из механизмов слабеет, а второй продолжает работать, это становится возможным. Также при освоении взрослыми нового языка в условиях живой языковой среды можно наблюдать процессы, которые человек переживает в раннем детстве при освоении родного языка. Поэтому билингвизм, а также его становление предоставляет важную информацию о механизмах понимания речи.

Почему мы говорим о двух стратегиях обработки информации полушариями мозга, а не о том, что полушария обрабатывают разные стимулы? Потому что это больше соответствует известным фактам. Предположение, что дело не в стимулах, которые обрабатывает мозг, а в способе обработки, оказалось весьма продуктивным. Также важной является догадка о значении задачи обработки стимула. Задачи

обработки одного и того же стимула могут быть разными (например, задача прочитать рукописный текст и задача определить по почерку автора текста) и могут активировать разные стратегии обработки информации. Кто первым предложил эти гипотезы — сказать трудно. Изучение разных мозговых функций и разными методами приводило исследователей к сходным выводам. В то же время и сейчас немало работ, авторы которых исходят из посылки, что полушария обрабатывают те или иные стимулы. Особенно много таких работ на тему нарушения узнавания лиц (лицевая агнозия).

Как известно, у взрослых людей право- и левосторонние поражения мозга стабильно приводят к разным синдромам, которые в нейропсихологии известны как право- и левополушарные синдромы. В докомпьютерную эпоху результаты нейропсихологических тестов использовались для определения места очага поражения (правое или левое полушарие, горизонтальная и вертикальная локализация внутри полушария) при операциях на мозге. Эти факты свидетельствуют, что, во-первых, организация мозговых функций достаточно устойчива, т. е. повторяется от человека к человеку и от поколения к поколению, и, во-вторых, доказывают, что у взрослых полушария специализированы и выполняют разные функции. Но в небольшом проценте случаев наблюдаются отклонения от обычной организации (к примеру, зеркальная организация у левшей). Это говорит о возможной изменчивости. Известны также случаи практически полного восстановления мозговых функций после раннего удаления одного из полушарий, т. е. здоровое полушарие начинает обеспечивать обе стратегии обработки информации. То есть чтобы быть здоровым не обязательно иметь два полушария, но обязательно иметь две стратегии обработки информации. А вот слабость или нарушение одной из стратегий обработки информации обязательно приведут к какому-либо нейропсихологическому синдрому. В настоящее время есть доказательства изменения организации функций мозга, которые связаны с освоением грамоты [Dehaene, Cohen 2007]. Письменная речь — продукт культуры, навык, который может быть сформирован лишь при определенных условиях в результате обучения. Нейрофизиологические исследования, использующие фМРТ, свидетельствуют, что обучение чтению влияет на церебральный субстрат для восприятия