

<https://doi.org/10.30853/filnauki.2019.8.56>

Левина Виктория Александровна

ЗНАЧЕНИЕ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ В СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ

В данной статье автор проводит анализ практического эффекта от использования достижений лингвистической науки в глобальном процессе цифровизации и информатизации общества. Особое внимание уделяется проблеме синхронизации скоростей распознавания входящей информации человеком и машиной. Автор предлагает более пристально изучить свойства символа как мощного инструмента лингвистического сжатия, позволяющего человеку производить распознавание почти мгновенно и в полном объеме без потери качества информации.

Адрес статьи: www.gramota.net/materials/2/2019/8/56.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2019. Том 12. Выпуск 8. С. 283-289. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2019/8/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net
Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

Прикладная и математическая лингвистика

Applied and Mathematical Linguistics

УДК 811

Дата поступления рукописи: 17.05.2019

<https://doi.org/10.30853/filnauki.2019.8.56>

В данной статье автор проводит анализ практического эффекта от использования достижений лингвистической науки в глобальном процессе цифровизации и информатизации общества. Особое внимание уделяется проблеме синхронизации скоростей распознавания входящей информации человеком и машиной. Автор предлагает более пристально изучить свойства символа как мощного инструмента лингвистического сжатия, позволяющего человеку производить распознавание почти мгновенно и в полном объёме без потери качества информации.

Ключевые слова и фразы: функции мозга; сложноорганизованные системы; системы распознавания образов; автоматическое лингвистическое индексирование; логические операции; естественные и искусственные нейроразподобные сети; символ и его значение в современных информационных системах; компьютерное распознавание речи.

Левина Виктория Александровна, к. филол. н.

Национальный исследовательский университет «Высшая школа экономики», г. Москва
vittoriauno@yandex.ru

ЗНАЧЕНИЕ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ В СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ

Введение

Успешное развитие современных информационных систем, безусловно, приводит к качественным изменениям в нашей повседневной жизни. С точки зрения социальной философии, политической экономии, экономической науки в целом человечество стоит на пороге новой технологической революции или смены экономического уклада.

Однако успех информатизации общества являлся во многом результатом многолетней последовательной работы в различных областях науки и техники, включая, безусловно, и лингвистическую науку.

Лингвистика, являясь по своей сути гуманитарной наукой, стала неотъемлемой частью построения логических систем и архитектуры сложнейших электронных программ, их семантического обеспечения самого широкого спектра, включая электронные поисковые машины, системы распознавания образов, электронные переводчики.

Актуальность работы заключается в комплексном анализе преимуществ символа и символьных систем как инструмента восприятия входящей информации мозгом человека в условиях высокой скорости либо повышенной сложности этой информации в современной информационной среде.

Научная новизна состоит в сопоставлении ранее известных науке скоростей реакции человека на раздражители с новейшими открытиями неврологов в части количества и способа взаимодействия нейронов в мозге человека, и как следствие, – решение главной задачи в анализе в том числе речи, звуков, изображений – задачи распознавания.

Практическая значимость подобных исследований заключается в том, что возникают новые возможности и технологии «сжатия» информации без потери ее смысла для пользователя информационной системы. Инженеры-программисты получают дополнительные инструкции о том, какие качественно-количественные, графические, текстовые данные можно презентовать, например, в виде символа, мнемосхемы, иконки; достаточно ли уже существующих символов либо графическим дизайнерам необходимо разрабатывать новые.

Необходимость использования достижений лингвистики в технических исследованиях

Наука о языке, его законах, эволюции развития устной и письменной речи является в значительной степени частью науки о функциях мозга человека. В этом смысле очень интересными представляются высказывания

математика, исследователя сложных динамических систем, лауреата Нобелевской премии И. Пригожина и Г. Николаса. «Из всех природных объектов живые существа, несомненно, и функционально, и морфологически являются наиболее сложными и высокоорганизованными... Более того, живые системы совершенно определенно функционируют вдали от равновесия... Эти различия, подразумевающие сильную неравновесность состояния, лежат в основе процессов типа распространения нервного возбуждения, играющего важную роль в жизни» [12, с. 43].

Исследования в области электронных вычислительных систем начались в США и несколько позже в СССР в период 50-60-х годов XX века.

Работы по систематизации терминологии в области науки и техники начаты в нашей стране еще в 1930-1940-х гг. в рамках Комитета по терминологии АН СССР под руководством Д. С. Лотте. Именно благодаря этому комитету в нашей стране заложены основы научной индексации больших массивов научной информации с целью упрощения поиска необходимых данных.

В большинстве современных информационных систем в процессе индексирования и поиска документов используются самые различные лингвистические методы, такие как специализированные словари, тезаурусы, тематические базы данных, графическая информация, видео- и аудиозаписи.

Работа с терминологическими структурами является наиболее простой, так как эти структуры принято считать линейными.

Существенно более сложной проблемой является неоднозначность поставленной пользователем задачи, когда приходится использовать расширение поисковых образов с обработкой всех возможных в данном случае словоформ. Существует метод автоматического кодирования слов по аффиксам и окончаниям. Достаточно часто программисты применяют и индекс частоты встречаемости (особенно термина) в информационном массиве.

Современные компьютеры имеют очень большую вычислительную мощность и объем памяти, а сети позволяют передавать даже качественные видеоизображения почти мгновенно. Следовательно, и методы лингвистической обработки должны соответствовать возможностям оборудования.

«Мегаинформационный» компонент обеспечивает семантику словоупотреблений. Этот шаг достаточно эффективно автоматизируется. В основе этого лежат исследования в области теоретической и прикладной лингвистики, создание «машинных» словарей, средств и алгоритмов синтаксического анализа [4, с. 219].

Наиболее удобными для автоматической системы поиска являются иерархические системы. В основе таких систем лежит принцип, когда в классификации имеется только один исчерпывающий класс для каждой группы объектов. С точки зрения логики для построения подобной классификации по правилам формально-логического построения деление должно на каждом уровне иерархии проводиться только по одному основанию. Эффективность этого метода зависит как от количества условий в системе, так и от заданной пользователем «глубины проникновения» поиска.

В современных научных исследованиях все больше используются данные сложных областей науки. Более того, эти данные могут находиться на другом континенте, быть записаны на другом языке, либо диалекте, либо представлять собой исключительно специальную область знаний. Кроме того, подобная информация может быть лишь гипотезой по интересующему вопросу, и по этому же поводу могут существовать и другие гипотезы или данные экспериментов. В подобных случаях более эффективными для поиска являются «Булевы классификации».

Математическая логика (алгебра высказываний) достаточно просто стала основой в математическом обеспечении компьютеров. Она не сложно встраивается в «битовую» логику, где истинность высказывания обозначается: 0 – ложь, 1 – истина. Кроме того, для управления алгоритмом поиска или вычислений используются классические в логике понятия «да – или – нет».

Врожденные логические операции свойственны и мозгу человека. Очевидно, что эффективность информационной системы напрямую зависит от синхронности действий человека и машины. Иначе говоря, технические возможности современных компьютеров и сетей связи уже вступают в конфликт с реальными возможностями скорости работы организма человека. Однако исследования последних десятилетий в области психологии, психолингвистики, биологии и медицины показывают, что существует достаточно обширная область для существенной синхронизации возможностей компьютера и человека.

Именно вопрос «механизма» работы биологического «логического вентиля», состоящего из нейронов или их групп, является сегодня основным не только для нейрофизиологов, но и для ученых во всех смежных дисциплинах, в том числе и лингвистов. Каким образом или под влиянием каких факторов мозг человека переключает свое внимание в том или ином направлении, остается пока загадкой.

Развитие современных биоподобных информационных технологий

Существует интересная гипотеза известного американского нейропсихолога К. Прибрама, которую он предложил как результат многолетних исследований мозга человека и животных, в том числе совместно с выдающимся отечественным ученым А. Р. Лурия.

К сожалению, Вторая мировая война привела к большому количеству тяжелых ранений головы как у советских солдат, так и у солдат США, которые участвовали в борьбе с фашизмом. Это позволило ученым не только совершенствовать технологии лечения, но и проводить исследования того ущерба, который нанесли умственной деятельности повреждения мозга при ранениях.

К. Прибрам выдвинул гипотезу, что мозг человека в неповрежденном состоянии воспринимает картину мира (в доступном ему диапазоне) как голограмму. «Все голограммы в целом обладают рядом интересных свойств, которые делают их потенциально важными для понимания функционирования мозга. Первое и наиболее важное для нас состоит в том, что информация о каждой точке объекта распределяется по всей голограмме и тем самым делает регистрацию ее устойчивой к разрушению. Любая малая часть голограммы содержит информацию обо всем объекте-оригинале и, следовательно, может восстановить ее. Голограмма обладает фантастической способностью к эффективному (то есть восстанавливаемому) хранению информации. Плотность хранения информации ограничена не только длиной волны когерентного света» [13, с. 171].

Существенный интерес представляют исследования Сюзаны Херкулано-Хузел (Suzana Herculano-Houzel) из университета Рио-де-Жанейро. Ее эксперименты показали, что мозг человека содержит около 86 миллиардов нейронов [16], часть которых имеет связь с органами чувств, часть отвечает за связь с исполнительными органами организма (сердце, руки, ноги, другие органы) и т.д. Очевидно, с деятельностью нейронов связаны и память, и способность человека мыслить. Каждый из нейронов имеет связь с 10-15000 других нейронов, образуя таким образом многомиллиардную нейронную сеть, имеющую как вертикальные, так и горизонтальные связи. На этой базе существует и естественный язык. По функциональной классификации нейроны делят на:

- афферентные;
- эфферентные;
- ассоциативные (осуществляют связь между эфферентными и афферентными).

По форме клетки нейроны могут быть:

- сферическими;
- зернистыми;
- звездчатыми;
- пирамидными – 80-90% всех клеток коры;
- грушевидными;
- веретеновидными [17].

Размеры нейронов могут составлять от 5 до 150 мкм.

Таким образом, мы видим, что части мозга человека достаточно хорошо изучены. Главной задачей сегодняшнего дня является выявление основных принципов функционирования нейронов, их реакции на входящий сигнал, типов связей между нейронами и того сложнейшего механизма взаимодействия, которое компьютерщики называют термином «нейроподобная сеть».

В данном высказывании ясно видны намерение и необходимость создавать электронные системы, как можно близкие по своей архитектуре к принципам работы мозга человека. Исследования нейронов показывают, что естественная биологическая сеть, в которую они встроены, имеет определенную структуру. Каждая из групп нейронов отвечает за восприятие определенного входного сигнала и при его получении изменяет свое состояние (электрическим или химическим путем), передавая обработанный сигнал на более высокий уровень. Передача, очевидно, возможна только в одном направлении. На более высоких уровнях нейронной сети уже необходимо производить более «тонкую» обработку сигналов с использованием логических сравнений с уже существующими в памяти образами.

В значительной степени подобие естественным процессам в технике возникло уже на первых этапах развития вычислительной индустрии, оказалось, что принципы классической логики одинаковы для функционирования и технических устройств, и живых организмов. Например, устройство, пропускающее сигнал только в одну сторону, называется «диод», впоследствии он получил возможность делать пропуск сигнала постепенно, и появился тиристор. Переключение движения сигнала в том или ином направлении можно осуществлять с помощью транзистора. Это устройство фактически осуществляет функцию «логического вентиля».

Транзисторы соединяют таким образом, чтобы выход используемого вентиля был подключен к одному или нескольким входам других вентилях. Собственно, логический уровень самого сигнала представлен в виде уровня напряжения или в виде значения тока. Подобным образом работают и ячейки памяти компьютера. В различное время они представляли собой ферритовую катушку, которая или намагничивалась, переходя в фазу «1», или, напротив, разряжалась, переходя в фазу «0», аналогичную функцию может выполнить и конденсатор, либо принимая значение своей емкости, либо разряжаясь. На этом примере мы видим, что в основе своей работы и живые нейроны, и полупроводниковые приборы используют одинаковые принципы – «изменение состояний».

Со второй половины XX века получает распространение представление о мозге как о компьютере – биологическом, сложном и сверхмощном, но все же компьютере. Сущность идеи заключается в том, что мозг, так же как и компьютер, принимает информацию из окружающей среды, анализирует, объединяет и сохраняет в памяти, а затем использует эту информацию для управления действием. «Нервные импульсы перемещаются по крошечному отростку (аксону), исходящему из тела клетки. Целые пучки аксонов (подобно проводам) протянуты из одной части мозга в другую. Наши мышцы сокращаются и заставляют двигаться тело лишь благодаря тому, что они получают соответствующие указания (импульсы) от нейронов» [4, с. 62].

Количество связей одного нейрона может достигать 10000.

Существенный вклад в развитие идеи нейроноподобных сетей внесли ученые У. МакКалок и В. Питтс еще в 1943 г. [9]. Они впервые предложили методы нейроматематики. В наши дни существует более 200 различных

моделей нейроподобных сетей. Эту задачу решает и чрезвычайно быстрое развитие электронного машиностроения, кристаллографии, пленочных и нанотехнологий.

Объективные предпосылки создания искусственного интеллекта

К основным задачам в технологиях создания искусственного интеллекта следует отнести дальнейшее сближение возможностей машины и естественных возможностей человека, о необходимости которого мы говорили выше. В этом существенную помощь могут оказать исследования особенностей психики человека, в том числе связанные со свойствами мозга формировать образы. Отчасти понятие «образ» соответствует понятию «голограмма», и то и другое также вполне соответствует теории формирования понятий у человека. Этот вопрос в науке (философии и психологии) ставился еще Платоном и Аристотелем, а также, позже, такими учеными, как Дж. Локк, У. Джеймс, Ф. Гальтон и др. Существенный вклад в изучение психологической части этой проблемы внес К. Г. Юнг. Он ввел в психологическую науку понятие «архетип». «Образы архетипов являются своего рода управлением подсознания. Они довольно самостоятельны, что позволяет им самим перерабатывать ситуацию, согласно их символике, и вторгнуться в нее, принимая свои импульсы и мысли», «архетипы представляют собой динамичные явления, которые проявляются так же неожиданно, как и инстинкты, а подсознание, содержащее в себе архетипы, мыслит, руководствуясь инстинктами. Архетипы заключают в себе союз образов с эмоциями. Именно через этот союз они воздействуют на человека и его чувства» [15, с. 126]. Большой вклад в развитие теории искусственного интеллекта внес Марвин Л. Мински. Еще в 1954 г. он опубликовал свою диссертацию в Принстоне на тему «Нейросети и проблема модели мозга» [10]. Значительное влияние на современную вычислительную лингвистику оказала его работа «Семантическая обработка информации», вышедшая в 1968 г. В этой связи существенно развились такие приложения к логике, как

- структурное программирование;
- доказательное программирование;
- логическое программирование;
- комбинаторная логика;
- семантическая паутина.

М. Мински неоднократно консультировал Стэнли Кубрика и создателей фантастических фильмов о будущем, которые оказались поистине пророческими. Например, в фильме «Космическая Одиссея» еще в 1968 году можно было наблюдать, как герои фильма возникают в виде голограмм, разговаривают с компьютером на естественном языке, получают ответы машины в виде объемных изображений, то есть в виде, максимально похожем на естественное восприятие человека через органы зрения, слуха, осязания [Там же].

В случае, когда задача заключается в точном переводе текста или точном ответе на запрос пользователя, существенным фактором является семантика, то есть распознавание «значения» и «смысла». Различие этих понятий выявил выдающийся отечественный ученый Л. С. Выготский ещё в 1930-х годах [2].

Эффективность работы поисковых электронных машин в значительной степени зависит и от способов представления информации в базах данных. Если речь идет о масштабном исследовании проблемы, на которую имеется много точек зрения, спорных материалов, целесообразно использовать такие методы, как лексическая или синтаксическая компрессия, свертывание информации. В составе синтаксической компрессии существуют такие инструменты, как эллиптирование, грамматическая неполнота, бессоюзия, синтаксическая асимметрия.

В новом качестве, особенно в компьютерной лингвистике, выступает знак – символ. Очевидно, что все знаковые системы, существующие сегодня, начались именно с символов – сначала с наскальной живописи, затем – с пиктограмм и сакральных знаков. Эти символы были образительно прочно связаны с объектом восприятия и позволяли мгновенно и безошибочно классифицировать, например, оленя среди других животных. «Можно догадываться, что формирование образов – это удобный или экономичный метод передачи информации через некоторый вид интерфейса. Заметим, что, когда человек должен действовать в контексте человека – машинного взаимодействия, удобно, чтобы машины передавали ему свою информацию в форме образов» [3, с. 51]. Очевидно, что подобный способ выдачи запросов может оставить «в тени» другие свойства объекта, например, его взаимосвязь с другими объектами, но скорость и удобство выдачи, конечно, многократно увеличивают и скорость реакции пользователя. «Еще одна форма примитивного кодирования, использующего наглядность, – это кодирование вида “часть – вместо – целого”». «В нашей жизни объекты восприятия – это почти всегда восприятие частей, и наши догадки о целом постоянно подтверждаются или опровергаются последующей демонстрацией других частей. Возможно, целое вообще нельзя предъявить, поскольку это потребовало бы непосредственной коммуникации» [4, с. 51].

В современных компьютерных системах широко используются символьные обозначения основных настроек устройств, дополнительных подпрограмм и сервисов, управления центрами сетевых операторов в сети Интернет или в ее мобильных версиях.

Все символы – идеограммы. «Символ – научно постижимая форма знаковых явлений, отличающаяся от знаков-индексов, знаков-знаков, знаков-жестов, знаков-рисунков. Сущностно расходясь с пиктограммой, логограммой, в точном смысле слова символ есть разновидность идеограммы.

Как емкий знак идеограмма имеет:

- план выражения: графемы, фонемы, слоги, слова, синтагмы;
- план содержания: семы, семемы, морфемы, лексемы, пропозиции, развертывания.

Синтез обоих планов дает символ, скрепляющий все уровни гуманитарной значимостью» [5, с. 370].

Выдающийся отечественный философ А. Ф. Лосев ввел понятие «номинативное мышление». Он относил его к методу научного мышления в целом, однако это понятие в полной мере отражает и подходы к современному машинному автоматическому поиску документов. «Номинативное мышление впервые оказывается способным мыслить предмет именно как таковой, узнавать его в смутном и смешанном потоке вещей, отождествлять его с ним же самим вопреки текучести и спутанности всего ощущаемого и воспринимать его. <...> Действительно, развитая номинативная морфология есть языковое выражение той центральной особенности номинативного мышления, которая стоит в разыскании закономерностей, в определении закономерных переходов между сущностью и явлением или, попросту говоря, в разыскании и установлении законов природы и общества» [8, с. 370].

Не последнюю роль в обработке естественного языка, извлечении необходимых данных, понимании смысла предложений играют модель и виды ее семантических представлений информации. Наиболее часто используемыми являются сетевые и графовые модели, фреймы и сценарии.

Семантическая сеть (semantic network) представляет собой ориентированный граф, где понятия находятся в узлах, а дуги передают отношения между понятиями.

В сети реализуются следующие отношения:

- род – вид, класс – подкласс, часть – целое, элемент – множество;
- функциональные и логические отношения;
- пространственные, временные, количественные отношения;
- атрибутивные отношения.

Разновидностью сетей являются концептуальные графы, которые позволяют совместить логику с выразительностью естественного языка. Для таких графов создан стандарт представления и язык описания: от исчисления предикатов, KIF (Knowledge Interchange Format) до CGIF (Conceptual Graph Interchange Format) [11].

Семантический модуль состоит из двух типов данных – это словарь значений языковых выражений, включающий связи между этими значениями и правила интерпретации синтаксических структур, комбинации и преобразования значений.

Представление данных в форме фреймов также было разработано М. Мински в связи с исследованиями многомерных ситуаций. Стандартный фрейм содержит имя ситуации, поля значений слогов и поля процедур.

Особую роль в обработке текстов играют компьютерные тезаурусы. Это лексическая база значений с установлением в ней не только связей между понятиями, но и между значениями и выражающими их словами, а также и между самими значениями.

Информационно-поисковые тезаурусы используются, например, в специальных корпусах и словосочетаниях специальной лексики, терминов, а также для поиска нужного значения в составе терминологической системы. В основном такие системы бывают иерархическими, и составление их стандартизировано. Эти стандарты предусматривают следующие типы отношений:

- род – вид;
- часть – целое;
- причина – следствие;
- сырье – продукт;
- административные иерархии;
- процесс – объект;
- функциональное сходство;
- процесс – субъект.

Наиболее интересным для нас в общественно-политической тематике является тезаурус RuТез [19], который разрабатывается с 1992 г. в МГУ. Он включает в себя термины экономики, политики, законодательства, международных отношений, военной сферы, социальной сферы. RuТез содержит более 26000 понятий, 62000 терминов, 100000 прямых и 700000 наследуемых отношений между понятиями.

Тезаурусная поддержка информационного поиска заключается в процессах уточнения и расширения запросов.

Чрезвычайно интересных результатов достигли технологии распознавания речи, работа над которыми была начата ещё 20-25 лет назад. В этих технологиях одновременно используются методы и понятия нескольких прикладных и научных дисциплин – общей и прикладной лингвистики, ее составных частей (семантики, прагматики, грамматики, фонетики/филологии), технических наук – электротехники, акустики, теории информационных систем, теории электромагнетизма, электронного машиностроения. Единицами словаря речевой основы выступают целые фразы, отдельные слова, словосочетания, слоги, звуки.

Задача распознавания речи может решаться различными методами, например, распознавания отдельных ключевых слов или коротких фраз, изолированных фраз и слов. Наиболее сложными системами являются системы распознавания слитной речи. Анализ свойств речевого сигнала учитывает закономерности построения речевых цепочек (фонематический уровень). Последовательность фонем и грамматика формируют лексический и синтаксический уровни. Использование статистических методов позволяет увеличить скорость распознавания. Статистические методы сегодня считаются наиболее эффективными (СММ или НММ – Hidden Markov Models). При работе с высокофиктивным языком (в том числе и с русским), в котором существует несколько значений одного и того же слова, используют более сложные «гибридные» модели.

При использовании статистической модели распознавания применяют, как правило, 5 модулей (см. Рис. 1).

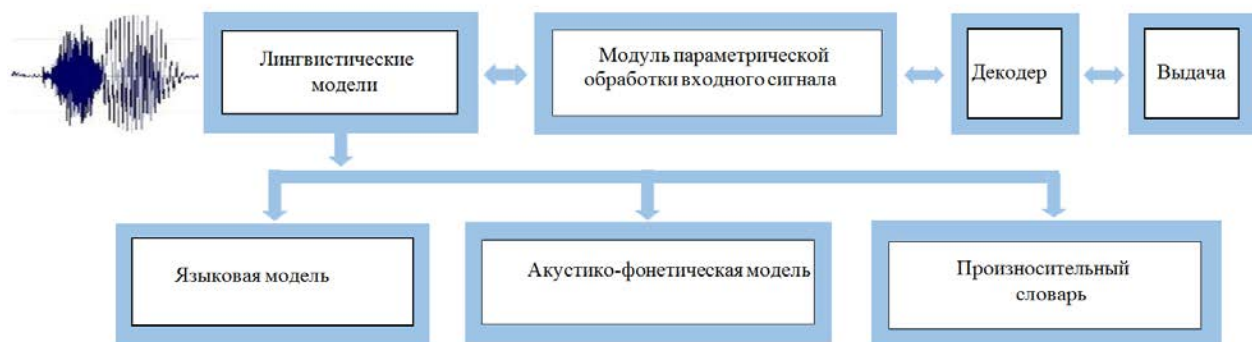


Рисунок 1. Модель распознавания речи компьютером

Без результатов, которых более чем за 50 лет достигла научная лингвистика, работа современных информационных систем была бы невозможна.

Лингвистический корпус (corpus linguistics) характеризуется прежде всего разметкой, или аннотацией (tagging, annotation).

Национальный корпус русского языка включает [18]:

- диалектный корпус;
- синтаксический корпус;
- обучающий корпус;
- поэтический корпус;
- мультимедийный корпус;
- исторический корпус;
- газетный корпус.

Специальные корпуса по различным тематикам насчитывают более 200 тысяч наименований.

Разметка в корпусе осуществляется методом приписывания текстам или их компонентам специальных тегов: собственно лингвистических, которые описывают лексические и грамматические и прочие характеристики элементов текста, и внешние в том числе; экстралингвистические признаки или элементы стандартизации (издательство, год издания, тематика, автор и т.д.).

К способам лингвистической разметки можно отнести частеречную разметку (part-of-speech tagging, pos-tagging). Кроме того, существуют семантическая и синтаксическая разметки.

Заключение

В качестве заключения очень точным определением самого процесса информатизации является высказывание известного лингвиста – философа А. Соломоника. «Каждый раз, когда происходит качественный скачок в распространении языковых систем, как, например, при изобретении книгопечатания, человечество выходило на новую информационную зрелость, напрямую связанную с умственной зрелостью человека. Наше время характеризуется небывалым информационным взрывом, базирующимся на новых средствах общения; нет никакого сомнения, что это обстоятельство ведет к кардинальным изменениям не только в материальных основах существования людей, но также и в способах и приемах нашего мышления» [14, с. 104].

Лингвистика в ее современном понимании как академической науки стала именно той основой, которая в симбиозе с логикой философии и математической логикой позволила создать фактически новое направление в науке – компьютерное программирование. В свою очередь, успехи в электронном машиностроении, кристаллографии, оптике и других областях позволили создать по сути новую отрасль промышленности – информационные системы. Однако увеличение технических возможностей оборудования, начало эпохи квантовых компьютеров делают эти достижения почти бессмысленными без более глубокого понимания внутренних механизмов языка как неотъемлемой части нашего собственного биологического компьютера – мозга. Очевидно, что язык – это не только то, что мы непосредственно произносим. Это может быть и существенно большая его часть, хранящаяся в памяти в виде образов, совокупности ощущений, может быть, и наследственной информации. Исследования в этой области ведутся во всех развитых странах с привлечением также и смежных дисциплин. Но уже сегодня можно утверждать, что именно лингвистика способна задать верное направление исследований, чтобы на основе законов функционирования языка найти физиологические закономерности и только потом повторить их в технике.

Повседневная жизнь современного человека становится все сложнее. Это понятно и в органах государственного управления. Символ является тем универсальным средством, которое в огромном потоке окружающей вас информации в виде речи, звуков, текста, речевых конструкций также позволяет кратчайшим путем донести нужный фрагмент информации до человека. Разработка новых символов и символьных систем, безусловно, должна стать следующим шагом в развитии информационных технологий и совместимых с ними новейших направлений лингвистических исследований.

Список источников

1. Бейтсон Г. Разум и природа: неизбежное единство. М.: URSS; Либроком, 2016. 256 с.
2. Выготский Л. С. Мышление и речь // Выготский Л. С. Собрание сочинений: в 6-ти т. М.: Педагогика, 1982. Т. 2. Проблемы общей психологии. С. 5-361.
3. Голицина О. А., Максимов Н. В., Попов И. И. Информационные системы. М.: Форум; Инфра-М, 2014. 448 с.
4. Данкан Дж. Где рождается интеллект. М.: Карьера Пресс, 2015. 256 с.
5. Ильин В. В. Теория познания. Символика. Теория символических форм. М.: Изд-во МГУ, 2013. 384 с.
6. Кларк А. 2001: Космическая Одиссея (2001: A Space Odyssey). Нью-Йорк, 1968. 222 с.
7. Ковальчук М. Курчатовский институт готовит новый вызов цивилизации // Аргументы и факты. 2018. 29 марта.
8. Лосев А. Ф. Знак. Символ. Миф. М.: Изд-во МГУ, 1982. 480 с.
9. МакКалок У. С., Питтс В. Логические исчисления идей, относящихся к нервной деятельности // Автоматы. М.: Изд-во иностр. лит., 1956. С. 363-384.
10. Мински М. Нейронные сети и проблема модели мозга. Принстон, 1954. 199 с.
11. Митрофанова О. А., Азарова И. В. Компьютерная и прикладная лингвистика. М.: Ленанд, 2017. 320 с.
12. Николис Г., Пригожин И. Познание сложного. М.: Ленанд, 2017. 360 с.
13. Прибрам К. Языки мозга. Экспериментальные парадоксы и принципы нейропсихологии. М.: Либроком, 2009. 464 с.
14. Соломоник А. Философия знаковых систем и язык. М.: ЛКИ, 2010. 408 с.
15. Юнг К. Г. Архетип и символ. М.: Ренессанс, 1991. 300 с.
16. Herculano-Houzel S. Brains matter, bodies maybe not: The case for examining neuron numbers irrespective of body size [Электронный ресурс]. URL: <https://www.pubfacts.com/detail/21535005/Brains-matter-bodies-maybe-not-the-case-for-examining-neuron-numbers-irrespective-of-body-size> (дата обращения: 07.07.2019).
17. <http://humbio.ru/humbio/physiology/001c0ace.htm> (дата обращения: 15.05.2019).
18. <http://www.ruscorpora.ru> (дата обращения: 15.05.2019).
19. <https://nlpub.ru/PyTez> (дата обращения: 15.05.2019).

LINGUISTIC RESEARCH IMPORTANCE IN MODERN INFORMATION TECHNOLOGIES

Levina Viktoriya Aleksandrovna, Ph. D. in Philology
National Research University "Higher School of Economics", Moscow
vittoriauno@yandex.ru

The article analyses the practical effect from the use of the linguistic science achievements in the global process of digitalization and informatization of the society. Special attention is paid to the problem of synchronizing the rates of incoming information recognition by a man and a machine. The author suggests examining more closely the properties of a symbol as a powerful tool of linguistic compression allowing a person to make recognition almost instantly and in full without information quality loss.

Key words and phrases: brain functions; complex systems; pattern recognition systems; automatic linguistic indexing; logical operations; natural and artificial neural networks; symbol and its significance in modern information systems; computer speech recognition.